The International Governance of Artificial Intelligence: Security Preference and Practical Impacts

Li Yan *

Abstract: A new wave of artificial intelligence (AI) development is sweeping the world. Considering its technology maturity and application scale, AI is in an infant stage of development. However, as the core technology and application driving the fourth scientific and technological revolution, AI has attracted worldwide attention for its potential transformative nature. The international community has expressed great concern over AI security, and the call for strengthening governance is growing. The international governance of AI has exhibited a strong preference for security. The underlying logic relates to AI's history, technological characteristics, and geopolitical changes. Such a security preference has determined the cognition, vision, and practical priorities of AI governance and will impact the future of AI and even the international balance of power.

Keywords: artificial intelligence, international governance, security preference

A new wave of AI technology and applications is emerging, prompting the international community to take action in addressing potential risks. In the international governance of AI, security concerns mainly come from the particularity of the development of artificial intelligence technology and the impact of geopolitical competition. This security preference will significantly impact the future of AI governance on the international stage and provide valuable insights for enhancing participation in related processes.

* Li Yan is a research professor and Director of the Institute of Sci-Tech and Cyber Security Studies, China Institutes of Contemporary International Relations, focusing on science and cyber security strategy and related global governance.

Security Preference of International Governance of Artificial Intelligence

The latest developments in artificial intelligence have brought the issue of international governance to the forefront of global politics. From the perspective of promoting development, it is crucial to recognize the global nature of the AI industry, with its networks and computing resources spread across multiple nations. This underscores the necessity for international cooperation to ensure a smooth and secure operation of these networks.[1] From the perspective of ensuring security, the risks posed by AI cannot be overlooked. These risks have the potential to cause various social, economic, and moral issues. Ignoring these risks would be detrimental to global stability and could hinder the progress of technology as a whole.[2] Therefore, international governance could play a vital role in formulating a framework to guide the development of artificial intelligence and benefit all mankind equally.[3] This round of AI technology application development and international governance is in its early stages. It is distinct from previous international governance experiences and rules. In contrast with the development preference in the early stage of technology application development, the current AI international governance process demonstrates a pronounced security preference.

International governance preference refers to the phased focus of governance concepts and practices that leans toward a certain direction. The primary objective of international governance is to achieve a balance between development and security. However, it should be noted that achieving absolute balance is an ideal state. In practice, the so-called balance point is always undergoing dynamic change, and the focus of phased governance may be more inclined towards development or more prominent in security. There are two key points that may be helpful to consider when understanding international governance preferences. First, based on different historical stages, such as the preferences in the early and mature stages of technology and application

[1] Robert Trager et al., "International Governance of Civilian AI: A Jurisdictional Certification Approach," August 31, 2023, https://arxiv.org/pdf/2308.15514.

[2] Esmat Zaidan and Imad Antonie Lbrahim, "AI Governance in a Complex and Rapidly Changing Regularory Landscape: A Global Perspective," *Humanities and Social Sciences Communications*, 2024, 1–18.

[3] David Leslie et al., "'Frontier AI,' Power, and the Public Interest: Who Benefits, Who Decides?," *Harvard Data Science Review*, September 9, 2024, 1–20.

development, they are bound to be different. Second, development and security preferences are always relative.

There are specific guidelines that govern the international governance of emerging technology applications. In the initial stages, the emphasis is on development preference, which means that the focus is on promoting development through the concepts, cognition, mechanisms, and policy measures. However, as the technology matures, particularly in the promotion and popularization of social applications, the security preference becomes increasingly evident. In the early stages of technology and applications, the promotion of development often takes precedence over security concerns, and all parties are willing to assume more risks. The social impact and consequences of security issues arising from technology and applications will be addressed in the governance agenda as they become a sufficiently significant concern. The international governance of cyberspace regarding Internet technology and applications is a typical example. When the Internet first began to develop in the 1990s, the international community focused on promoting Internet technology and applications as soon as possible to share the benefits of the Internet and promote human society to move towards an information society. At this stage, the priority is to ensure the security, stability, and development of technology. As the social application of the Internet continues to evolve, the global security issues it brings have begun to receive attention. A significant development was the launch of the World Summit on the Information Society Process in 2003 under the United Nations framework. It promotes the transformation of cybersecurity governance from a technology-centered to an integrated governance model to address a broader range of social issues. In the wake of the Snowden Incident in 2013, there has been a notable rise in national security concerns in cyberspace governance, and the influence of geopolitics on cybersecurity has become more evident. In response, countries have updated their national cybersecurity strategies and accelerated the international governance process for additional security issues.

The current international governance of artificial intelligence, still in its early stages, exhibits a distinct security preference that differs from the aforementioned general rules. First, scholars explore the various security risks and threats that artificial intelligence poses in political, economic, social, and

ideological security. They also examine how artificial intelligence affects future national strength and the international power structure. In April 2024, the Stanford Institute for Human-Centered Artificial Intelligence published the Artificial Intelligence Index Report 2024. This report counted the number of security papers submitted to academic conferences in the field of artificial intelligence from 2019 to 2023. The results showed that the number of submissions in 2023 increased by 70.4% compared with 2019.[1] This indicates that with this round of breakthroughs in artificial intelligence technology and its application, the academic community has increased its focus on artificial intelligence security issues. Its potential security risks have gradually become a research hotspot in academic research and an important challenge in technological applications.

Second, the policy community focuses on the principles and framework design of security governance. In addition to the security, international organizations and relevant countries also use high-frequency words such as reliable, trustworthy, AI for Good, and responsible in various international agendas and policy documents to promote AI governance. This fully demonstrates the expectations for the development of AI technology and applications, which are all based on security considerations at different levels. António Guterres, Secretary-General of the United Nations, announced the establishment of a High Level Advisory Body on AI in 2023 and released the final report of Governing AI for Humanity in 2024, proposing an action plan to strengthen global cooperation. In the same year, the United Nations General Assembly adopted Resolution on Enhancing International Cooperation for AI Capacity Building and Resolution on Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development. Key international organizations and mechanisms also recognize AI as a significant subject. For instance, in June 2023, the World Economic Forum established the AI Governance Alliance, which proposed recommendations such as responsible development, open innovation, and international cooperation. In September 2023, the G20 New Delhi Summit reaffirmed the commitment to a human-centered approach to AI, emphasizing the importance

[1] "Artificial Intelligence Index Report 2024," Stanford Institute for Human-Centered Artificial Intelligence, 2024, 187.

of ensuring that AI benefits all, while also outlining the need for effective governance frameworks and global supervision. In addition, the international community is also building new AI governance platforms and mechanisms. For example, the United Kingdom convened the inaugural AI Safety Summit in November 2023 at Bletchley Park. The Bletchley Declaration was issued at the meeting, emphasizing the importance of scientific and technological development to promote the common well-being of mankind and environmental sustainability and calling for global cooperation to ensure that the positive impact of technology is maximized while reducing potential risks and negative impacts.[1] On October 18, 2023, China released the Global AI Governance Initiative, which systematically expounded China's solution to AI governance in the three aspects of AI development, security, and governance.

Third, the industry is exploring how to ensure the security of artificial intelligence. In the process of promoting the implementation of applications, greater attention is being paid to security in the design and launch of artificial intelligence products and services. For example, the Global System for Mobile Communications Association launched the Responsible AI Maturity Roadmap in September 2023 to provide telecom operators with adaptation tools and guidance to help them assess the maturity level of responsible AI. In February 2024, at the Munich Security Conference in Germany, 20 technology companies, including Amazon, Google, IBM, Meta, Microsoft, OpenAI, TikTok, and X, signed the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. This accord aims to resist deceptive AI-generated content, reduce the risks it poses, and propose solutions to address these issues on their respective platforms or products. It also pledges to collaborate with global organizations and academia to raise awareness among the public and media about the potential dangers of AI-generated deceptive content. In May 2024, the AI Seoul Summit drafted the Frontier AI Safety Commitments, urging signatories to manage risks responsibly when developing and deploying frontier AI models and systems. 16 AI companies or organizations from China, the United States, Europe, the Middle East, and other Asian countries signed the commitment. Signatories voluntarily commit to implementing a series of practices related to cutting-edge AI security. These

[1] "AI Safety Summit 2023: The Bletchley Declaration," November 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration.

practices include evaluations of internal, external, and independent third-parties; information sharing; investment in cyber security; a vulnerability reporting mechanism of third parties; and public and transparent reporting of internal security framework and crisis management measures.

The Underlying Logic of Security Preference

The reasons why this round of international AI governance exhibited a pronounced security preference in the early stage are multifaceted and intricate, primarily due to factors such as development, technology, and politics. In terms of development, this round of development wave has opened up an important link for AI technology to be widely used, and technological breakthroughs and application landing have accelerated the consideration of the security agenda. From a technical perspective, AI technology has introduced many new security challenges due to its high uncertainty and even uncontrollability. These challenges have also become the focus of international governance. From a political perspective, against the backdrop of intensifying geopolitical competition, political factors are highly embedded in the entire process of the development of artificial intelligence technology and applications. Security has become a tool of bargaining and a pretext for attacking opponents, which continues to influence and shape the international policy environment.

I. The application of AI accelerates the advancement of the security agenda

Historically, security issues will be considered after the advancement of technology and applications. However, the development of artificial intelligence is different from the past. The development of artificial intelligence relies on a multifaceted and interdisciplinary approach, requiring a diverse array of technical support, including computing power, algorithms, and research in brain science. Therefore, artificial intelligence is often restricted by technical conditions and remains in laboratories or limited specific fields. It has not achieved technological breakthroughs and widespread social applications, nor has it had a broad and transformative impact on society.

The origins of artificial intelligence can be traced back to the mid-20th century. In 1943, Warren McCulloch and Walter Pitts designed the first

artificial neural network model, and in 1950, Alan Turing proposed the Turing test method to determine whether a machine had human intelligence. The term "artificial intelligence" was first proposed at the Dartmouth Conference in 1956. Artificial intelligence has experienced several rounds of development, such as the introduction of early chatbots and expert systems from 1960 to 1970. With the emergence and rapid development of computer technology, artificial intelligence technology has been further advanced. The landmark event was IBM's DeepBlue defeating the world chess champion Garry Kasparov in 1997, demonstrating the potential and powerful thinking and analytical capabilities of artificial intelligence. The concept of Generative Adversarial Network emerged within the scientific community in 2004. In 2006, Geoffrey Hinton introduced the concept of deep learning, leading to significant advancements in artificial intelligence technology. In 2011, Watson, an IBM-developed computer, achieved a landmark victory by defeating two champions of the popular American television program "Jeopardy," showcasing remarkable advancements in artificial intelligence's capacity to comprehend and process natural language. In 2014, Ian Goodfellow and his team formally introduced the concept of generative adversarial networks, a revolutionary tool that significantly boosted creativity and innovation in the field of artificial intelligence.

In recent years, artificial intelligence has undergone a significant qualitative leap forward, largely due to the advent of technologies and applications such as big data, algorithms, and advanced computing capabilities. In 2022, the emergence of general-purpose large models, exemplified by ChatGPT, signaled a paradigm shift in human–computer interaction. This transition marks the culmination of a pivotal and challenging step: the transformation of technological innovation into practical social applications. OpenAI's ChatGPT-4, Google's Bard, Microsoft's Bing AI, and DeepSeek-R1, developed by China's DeepSeek, have all been released in a short period of time. Concurrently, a variety of application scenarios based on general models have emerged, signifying the emergence of the transformative potential of artificial intelligence. For example, AlphaFold 3 predicted the structures and interactions of all biomolecules overnight with unprecedented atomic accuracy. Another example is that GNoME successfully predicted 2.2 million crystal structures, of which 380,000 of the most stable crystal structures have the potential to become materials

for future transformative technologies, providing power for fields such as superconductors, electric vehicle batteries, and supercomputing power supply. Currently, scientists have begun to further synthesize new materials with the assistance of GNoME.

The implementation of a new round of artificial intelligence technology has given rise to a range of security risks and problems, including concerns regarding individual privacy, ethical considerations, social equity, employment substitution, military security, and political security. These issues have become important practical considerations that need to be addressed and resolved within the framework of international governance. Consequently, this underscores the necessity for the security agenda of international governance of artificial intelligence to be prioritized proactively.

II. Special technological forms bring new security challenges

Artificial intelligence is typically regarded as an emerging technology, and emerging technologies generally exhibit five characteristics: novelty, rapid growth, coherence, significant impact, and uncertainty and ambiguity. An increasing number of experts and scholars contend that a narrow focus on emerging technologies in the study of artificial intelligence is insufficient for a comprehensive understanding of the subject. In the context of emerging technologies and applications, the potential challenges associated with artificial intelligence may be unprecedented, which has led to heightened security concerns. Consequently, governance has become an area of focus to address these issues.

The security risks caused by artificial intelligence technology can be roughly divided into the following three categories. First, all technologies have a double-edged sword effect, and artificial intelligence technology is no exception. As an enabling technology, it will exacerbate existing security risks. The jailbreak network attack of artificial intelligence technology results in more severe real threats. The rapid development of artificial intelligence technology has increased the complexity and inexplicability of its systems, making security vulnerabilities in the system more difficult to detect. Jailbreak attackers could bypass the security restrictions and rule constraints of artificial intelligence systems through carefully designed inputs to gain the ability to manipulate the system. This can result in the theft of user sensitive data, the creation of

malicious false information, the output of harmful content, the launching of cyber attacks, and the commission of real crimes. For example, in 2024, Microsoft launched an AI jailbreak tool called Skeleton Key, which can bypass the protection mechanisms in multiple AI systems and force them to generate content that violates moral ethics and social order.[1]

Second, AI hallucination is a problem. Artificial intelligence hallucination refers to the deviation of the output content of the artificial intelligence system from the actual situation. The output content seems reasonable but in fact lacks a reliable basis, is meaningless or even completely wrong. The primary reason for this phenomenon is that, during the construction of the model by the artificial intelligence system, the training data is inherently biased, making it challenging to encompass all potential scenarios. The continuous enhancement of the model's performance results in overfitting, and the ambiguity and complexity inherent in natural language hinder the system's ability to comprehensively grasp semantics. Hallucinations are more prevalent in text generation tasks. For instance, in the composition of news reports, fabricated details of non-existent events are frequently incorporated, significantly impacting the user experience. In 2023, Google's artificial intelligence system Bard answered in a public demonstration that the James Webb Space Telescope has captured for the first time photos of planets outside the solar system. However, the answer was wrong. Artificial intelligence has received widespread attention for providing inaccurate information due to hallucination problems.[2] The hallucination problem is not likely to be resolved with the development of artificial intelligence technology. It will require ongoing technological improvements to alleviate the issue. However, this will introduce significant safety hazards in high-risk scenarios, such as financial investment modeling, autonomous driving, and medical diagnosis.

Third, the emergence problem of artificial intelligence is uncontrollable. In recent years, there have been continuous practical cases showing that as the model scale increases, artificial intelligence systems will show emergence

---

[1] Chris Mckay, "Microsoft Reveals 'Skeleton Key': A Powerful New AI Jailbreak Technique," June 28, 2024, https://www.maginative.com/article/microsoft-reveals-skeleton-key-a-powerful-new-ai-jailbreak-technique/.

[2] Carrie Mihalcik, "Google ChatGPT Rival Bard Flubs Fact about NASA's Webb Space Telescope," February 9, 2023, https://www.cnet.com/science/space/googles-chatgpt-rival-bard-called-out-for-nasa-webb-space-telescope-error/.

phenomena. When the model reaches a certain critical scale, its ability will suddenly jump from a level close to random to a level far above random. Research indicates a correlation between this phenomenon and the size of the model, with its significance being particularly pronounced in the context of diverse tasks. This ability is essentially unexplainable, unpredictable, and even uncontrollable. From a positive perspective, this may represent the most creative manifestation of intelligence and could potentially serve as a significant driving force for future scientific progress. Conversely, from a negative perspective, this may also provide a realistic basis for intelligent machines to control human scenarios.

The international community has expressed concerns over the potential for a loss of control over AI technology, leading to the perception that humans lack the capacity to effectively govern the future of this technology. In the aftermath of the release of ChatGPT, the international community once again called for a halt to the development of AI technology that could potentially lead to significant security risks. Consequently, the development of AI has become increasingly focused on ensuring the safety and control of the technology. In May 2023, more than 350 industry executives, experts, and professors in the field of artificial intelligence signed a Statement on AI Risk, saying that mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war. The signatories of the Frontier Artificial Intelligence Security Commitment also pledged to set and assess intolerable risk thresholds, suspend the development or deployment of artificial intelligence models and systems when the risks are too great, and to ensure the safety and trustworthiness of the technology. On July 18, 2023, the UN Security Council convened a high-level briefing entitled Artificial Intelligence: Opportunities and Risks for International Peace and Security. This briefing underscored the global consideration of strategies to mitigate the potential hazards associated with emerging artificial intelligence technologies. These technologies are poised to profoundly influence the global economy and reshape the international security landscape. On January 15, 2025, at the plenary meeting of the United Nations General Assembly, Guterres reiterated his previous calls to action regarding the governance of artificial intelligence. He advocated for the prevention of uncontrolled and imbalanced development

of AI technology, ensuring equitable access to the latest knowledge and supporting developing countries in leveraging AI to achieve sustainable development. Guterres further proposed the establishment of an international, independent scientific group on AI and the initiation of a global dialogue on AI governance in 2025.

III. High politics narrative strengthens security policy environment

The world today is experiencing significant changes that have not been seen in a century, coinciding with the onset of a new era of scientific and technological innovation. Under the dual influences of technology and politics, the field of science and technology has emerged as a pivotal domain for competition among major powers. Technological issues are inherently politicized, underscoring the strategic importance of this sector. It is generally recognized that artificial intelligence technology is the core technology and an important application driving the fourth scientific and technological revolution. Major countries in the world are planning artificial intelligence development strategies and participating in international artificial intelligence governance from the perspective of maintaining national security, enhancing national competitiveness, and reshaping the international power structure. This underscores the intricate interplay of geopolitical dynamics within the overarching framework of international artificial intelligence governance. The high politics narrative, founded on national security and even international security, serves as a prominent theme in this landscape, alongside the continuous strengthening of security preferences within the international policy environment.

First, in the context of strategic competition between China and the United States, the competition between the two countries in the field of science and technology has intensified. The new scientific and technological revolution in the digital age has led to significant changes in the connotation, structure, and system of power. Emerging digital technologies have become an important factor in determining the international competitiveness and international status of China and the United States. In the field of artificial intelligence, the United States has consistently emphasized the importance of maintaining its competitive edge, perceiving China as "a significant competitor and a potential national security threat." In 2019, US President Donald Trump issued an

executive order on Maintaining American Leadership in Artificial Intelligence. This was intended to promote US artificial intelligence technology and innovation, thereby maintaining the United States' dominant position in this field. In the same year, the Congressional Research Service of the United States released a research report on Artificial Intelligence and National Security. In the past two years, with the advent of a new wave of AI proliferation, the political and academic spheres of the United States have persistently reinforced such security discourses, asserting that "China will challenge the United States' relative advantage and leadership in the domain of AI and on a global scale, thereby posing a grave threat to US national security." Since taking office, Biden has taken several actions to restrict China's development of artificial intelligence. These actions include tightening export control measures, interfering with international scientific research cooperation, and collaborating with allies to build a system that excludes China in the field of artificial intelligence. The administration has also attempted to create a parallel system in this field. In January 2025, Biden issued new AI regulations prior to his departure from office, further tightening export restrictions on AI chips and technologies to China through a tiered management approach. The objective of these measures is to ensure that advanced computing power remains within the United States and its allies while simultaneously restricting China's access to sophisticated AI chips and technologies.

Trump and Biden appear to share a similar perspective on preserving the United States' competitive edge in the realm of artificial intelligence. It is therefore anticipated that there will be no significant alterations to the current artificial intelligence policy during Trump's second term in office. Since January 2025, the US has demonstrated a shifting stance towards China's DeepSeek, as reflected in its actions and policies. On January 28, the newly appointed White House Press Secretary, Karoline Leavitt, claimed that Trump felt that the artificial intelligence model released by DeepSeek had sounded the alarm for the US artificial intelligence industry. She also said that the US National Security Council was investigating the potential impact of DeepSeek and that the White House would do its utmost to ensure the US's leadership in the field of artificial intelligence. According to reports from US Newsmax and other media outlets, Trump stated in an interview on February 7 that

DeepSeek would not pose a threat to US national security and that the US could eventually benefit from the startup's AI innovation. However, this statement merely indicates that although Trump has not yet employed the national security narrative, he has not altered his approach to maintaining the US's leading position.

The competition between the United States and China in the field of artificial intelligence may completely change the balance of power and have a significant impact on global governance, which may have an impact on data governance, technical standards, moral ethics, and geopolitical situations.[1] When the interaction between the two is based on security as the main narrative logic, its impact will inevitably extend to the international governance process. Because the priority of such international governance must be the power struggle that is more conducive to maintaining its competitive advantage or national security, rather than the intention of cooperation based on promoting common development.

Second, the majority of developing countries have adopted high-political narratives on artificial intelligence from the perspective of national security and even international security based on geopolitical considerations. This has further created an international governance policy environment with a security preference. For instance, developing countries have raised the issue of the AI divide under the United Nations framework. The AI divide is defined as the disparity in an entity's capacity to master and apply intelligent technology during the process of intelligent development. This disparity is evidenced by uneven resource distribution, varying application capabilities, and unequal development opportunities. While this current phase of artificial intelligence development remains in its nascent stages, the AI divide it has engendered has already begun to manifest.

Presently, the companies that possess the capacity to develop and promote new intelligent systems are predominantly concentrated in China and the United States. American technology companies such as Microsoft, Google, and Facebook have significant advantages in data, computing power, and high-end chips, enabling them to be the first to develop technology and occupy the

[1] Asia Maqsood, Ahyousha Khan, and Muhammad Usama Siddiqi, "US–China Competition in Artificial Intelligence: Implications on Global Governance," *Journal of Asian Development Studies*, Vol. 12, December 30, 2023, 481–493.

market. The gap between developing countries and developed countries, which has been narrowed after years of efforts, is likely to be further widened by this technological change. Therefore, relevant countries have expressed significant concern regarding the challenges and threats posed by digital colonization. In response, some countries have proposed the concept of Sovereign AI, which encompasses the following principles. The technology development process exhibits a certain degree of autonomy and is not entirely dependent on external entities. Crucial infrastructure is controllable and does not adhere to external control. Product applications are tailored to align with the national conditions and cultural context of the country, thereby preventing the erosion of foreign values and the occurrence of digital colonization. The strong sense of crisis of non-AI powers about their security and common security will inevitably be reflected in the international governance process of AI.

The AI Action Summit was held in Paris, France, on February 10–11, 2025. The summit is a reflection of geopolitical dynamics. In comparison with the previous two summits held in the UK and South Korea, there has been a significant increase in the participation of developing countries. This reflects their strong and realistic desire to participate in the international process and to pursue development opportunities. The US Vice President Vance's speech emphasized the "America First" policy stance, asserting that "The US will continue to be the gold standard in AI." Given the focus on DeepSeek from China at this conference, the "gold standard" statement can be seen as a declaration of the United States' dominance in the field of artificial intelligence in the eyes of the global community. Vance criticized the AI regulatory policies of the European Union and other countries on the grounds of opposing overregulation, highlighting the differences between the United States and European countries on the issue of AI governance. After the meeting, the United States and the United Kingdom refused to sign the final statement document, conveying their tough policy stance to the outside world. This casts a shadow on the future international governance process.

Competition and gambling are not only a struggle for the right to influence governance but also a struggle for the future development of artificial intelligence. Therefore, geopolitics is a very important factor in the development and governance of artificial intelligence.

The Impact of Security Preferences on AI Governance

The security preference of AI governance is a key consideration that affects and runs through the entire governance process. With the release of AI technology and its potential for application and transformative impact, related governance presents some clear characteristics.

I. Security concerns become a priority for international governance

From the perspective of development logic, this round of development of artificial intelligence technology and applications will objectively bring about a more severe Collingridge's Dilemma. First, the social and developmental impacts of a technology cannot be accurately predicted in the early stages of its application. If control is imposed too early because of concerns about adverse effects, the technology may be difficult to develop. Second, in the event of issues or risk threats, technology applications have been thoroughly integrated into the social system structure. Altering or controlling this integration will be challenging and costly and may even prove impossible. Consequently, it is imperative to be able to predict in advance, follow up in real time, and respond swiftly to security concerns in the development process.

From a technical logic perspective, artificial intelligence technology is unpredictable and uncontrollable, thus necessitating international coordination for governance purposes. AI technology and applications exhibit significant increasing marginal characteristics. Due to the long-term accumulation of data and autonomous iteration of algorithms, the development speed and impact of AI may also demonstrate characteristics of increasing marginal returns.[1] The rapid pace of technological change and accelerated implementation has further reduced the time required for policy response and evaluation. If the pace is not kept up, it will inevitably lead to more significant control challenges in the near future. Consequently, security has become a paramount concern for many individuals.

From a political logic perspective, each technological revolution in history has resulted in significant changes to the global landscape and the rise and

[1] W. Brian Arthur, "Increasing Returns and the New World of Business," *Harvard Business Review*, Vol. 74, No. 4, July 1996, 100–109.

fall of major countries. The application and prospects of artificial intelligence have prompted relevant countries to prioritize the pursuit of their advantages. For instance, the United States has consistently emphasized the necessity of attaining a dominant advantage. Consequently, it is impractical to anticipate that coordination between major powers will achieve safe artificial intelligence. Currently, competition in the field of artificial intelligence has intensified and become unmanageable, which, in turn, exacerbates the deterioration of the security situation. For example, the United States proposed a project in the field of artificial intelligence that some have referred to as the Manhattan Project. This project involves a significant investment of 500 billion US dollars with the aim of opening a Stargate. In this context, it is important to note that many significant issues related to the security of artificial intelligence can be addressed through international mechanisms. The protection of security goals will undoubtedly be a primary concern for international governance. As an institutional communication platform, international mechanisms have the potential to facilitate consultation, dialogue, and conflict resolution among sovereign states on a broader scale and with greater efficiency. They can also leverage the institutional strength of international organizations to prevent the escalation of governance problems caused by the rapid evolution of technology.[1]

II. Technology communities will play a more important role in international governance

In the context of technological innovation, technology communities—comprising technology companies, among other entities—function as both proponents and practitioners of research and development and industrial implementation. Consequently, their role in governance is more pronounced in the domain of development than in security. Technology companies and technology communities have long pursued technological liberalism, frequently exhibiting resistance to the development constraints imposed by geopolitical factors. Despite their willingness to engage more actively in the formulation of international rules, these entities remain inherently constrained within the technical domain. However, the emergence and rapid development of artificial

[1] Robert O'Brien, *Contesting Global Governance: Multilateral Economic Institutions and Global Social Movements* (Cambridge University Press, 2000), 136.

intelligence, coupled with the heightened prominence of its security factors, have significantly augmented the rationality and necessity of the technical community's involvement in international governance.

As an unprecedented technological form, the most unpredictable or uncontrollable risks of artificial intelligence technology are precisely due to its technical characteristics. Whether it is a response at the technical level or a prevention at the policy level, it requires the participation of more professional technical communities. This provides a natural rationality for the technical community to participate in the international governance of artificial intelligence. At the same time, since AI data resources, computing power, algorithms, and large models are all in the hands of technology giants, they will inevitably participate in the international governance of AI to enhance their voice and influence. This will allow them to gain a better development space and policy environment so as to better transform technology and industrial advantages into governance effectiveness. The technical community's involvement in the international governance of artificial intelligence is essential. The technical community is well aware of the security concerns of all parties regarding artificial intelligence technology, as can be seen from the emphasis on security in its application products and services. In view of this, in the future, the technical community, including technology companies, will be more proactive in transforming the resources they control into greater voice and dominance.

III. Competition for governance platforms will intensify

International AI governance is still in its early stages. A systematic, mechanism-based governance platform with broad representation and recognition has yet to be established. On the one hand, existing governance mechanisms face challenges in incorporating all AI issues. For example, the mechanism of international cyberspace governance is mainly based on Internet technology and applications, while artificial intelligence technology and applications are quite different from it. On the other hand, the existing governance mechanism is relatively loose.

The fierce competition and game between China and the United States over artificial intelligence is mainly reflected in the conception of international governance mechanisms. China recognizes the significance of international

governance within the UN framework and values the UN's contributions to the global AI governance landscape. China believes that effective coordination of significant issues, such as the development, security, and governance of international AI, is crucial and can best be achieved within the UN framework. The United States is more inclined to establish an international governance mechanism that not only allows it to play a leading role but also excludes China's influence. The United States is hindering the establishment of a universal governance framework and instead focusing on cultivating a regional governance structure under its own leadership. It is endeavoring to enhance this regional initiative into a de facto international governance mechanism through a process of demonstration or diffusion. For instance, the 2024 Seoul Declaration aims to take the lead in establishing a global governance framework for artificial intelligence within the G7.

To bridge the intelligence gap and preserve sovereign AI, developing countries require a more global platform that can address the emerging North–South challenge in the intelligent era. Non-state actors, including technology giants and technology communities, also require a platform that can enhance their participation and voice. However, this platform cannot be merely technology-industry oriented; it must be able to engage in effective dialogue and coordination with the state. Regardless of the stance of the state, the development of AI is greatly changing the power distribution and interaction patterns between sovereign states and non-state actors.[1] Therefore, the improvement of the international governance mechanism of AI requires the creation of a more appropriate platform, and the competition around the platform will become more intense and complex. This involves not only the game between countries but also the coordination between countries and non-state actors.

Conclusion

As a responsible major country, China is committed to promoting the development and governance of artificial intelligence. China has proposed

[1] Juho Lindman, Jukka Makinen, and Eero Kasanen, "Big Tech's Power, Political Corporate Social Responsibility and Regulation," *Journal of Information Technology*, Vol. 38, No. 2 (June 2023): 144–159.

the Global AI Governance Initiative and the Shanghai Declaration on Global AI Governance, which outline forward-looking explorations and Chinese contributions to promoting AI technology for the benefit of all mankind. On February 11, 2025, the AI Action Summit was held in Paris, France. France, China, India, the European Union, and other countries and international organizations jointly signed the Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet, proposing the need for inclusive multi-stakeholder dialogue and cooperation on AI governance to further deepen trust and strengthen security cooperation. The sustained efforts of the international community have laid the foundation for the future international governance of artificial intelligence. However, to advance the governance process and achieve governance goals, all parties must take more action. The focus of future promotion will include the following aspects.

First, from the perspective of international agenda setting, given the security emphasis in the early stages of developing AI technology and applications, the most effective basis for cooperation is to address and respond to shared security concerns. These common security issues include, but are not limited to: cooperating with all relevant parties in the international community to explore best practices for addressing security risks in the application of technology; challenging the narrow national security narrative of the so-called "China–US artificial intelligence competition" led by the United States, and alleviating the concerns of all parties in the international community about conflicts; caring about the security dilemma faced by the vast developing country of China due to the intelligence divide, promoting relevant capacity building, striving for broader cooperation, and creating a favorable international policy environment.

Second, given the increasing importance of the voice and influence of technology and industry in the development of artificial intelligence, the international community should further mobilize the enthusiasm of technology companies and technology communities. It is essential to guide them to balance market interests and security concerns, avoiding technological liberalism and excessive politicization and securitization. The AI for Good Global Partnership Program, initiated by Chinese companies and research institutions in June 2024, is a public initiative with the objective of fostering global collaboration to advance the responsible development of artificial intelligence technology. This

initiative aims to ensure that technological progress benefits all humanity while achieving the goal of sustainable development.

Third, as geopolitical tensions rise, there is a noticeable decline in the willingness and investment of countries to cooperate. However, the international governance of artificial intelligence is the prevailing trend, and the establishment of an international governance platform for artificial intelligence should be prioritized. On the one hand, the primary function of the UN framework should be supported in an effective manner. The coordination role and influence of the UN framework on global issues are irreplaceable, especially in integrating the interests of all parties and providing voice channels and communication mechanisms for the Global South countries. On the other hand, the establishment of new governance platforms at the regional level should also be encouraged. The Middle East, Africa, Southeast Asia, and other regions are playing an active role in the global AI production and supply chain that is taking shape. As future technologies and applications continue to advance, these regions are poised for significant growth in the field of AI, which will, in turn, increase the demand for effective AI governance. The establishment of regional platforms will play a pivotal role in shaping this international governance framework.

The future of artificial intelligence is promising, yet it is replete with uncertainties. Concurrently, this period has given rise to novel security risks, underscoring the necessity for enhanced international governance, particularly in the domain of security. This period of AI proliferation coincides with the unprecedented changes in the world. The high politics narrative, influenced by the unique historical context, persists in shaping the international policy environment. The governance of AI on the global stage, which should be founded on international cooperation, is currently facing significant practical challenges. In light of these challenges, the future of AI governance will be shaped by the coordinated cognition, action, policy formulation, and value selection processes.

(edited by Zhang Yimeng)