I-AIIG
清华大学人工智能国际治理研究院
INSTITUTE FOR AI INTERNATIONAL GOVERNANCE
TSINGHUA UNIVERSITY

国际支持机构
Supporting International Organization

Tsinghua University International AI Cooperation and Governance Forum 2020 White Paper

March 2021

CONTENT

I. About the Forum

During the COVID-19 pandemic, various AI technologies have been used for critical functions such as virus detection and diagnosis, contact tracing,as well as recovery monitoring. However, the widespread application of these technologies has also raised broader concerns about the balance between benefits and potential risks, both economically and socially. Many questions remain regarding the effect of AI technologies on individual privacy and inequality, and whether such technologies will lead to new discrimination, prejudice, and social divide.

In recent years, governments, non-governmental organizations (NGOs), scientific research groups, and private institutions around the world have successively published AI principles documents, the balance between AI development and risk control has been the focus of much discussion. Against this backdrop, hosted by Institute for AI International Governance of Tsinghua University (I-AIIG), with the United Nations Development Programme (UNDP) as international supporting organization, the Tsinghua University International AI Cooperation and Governance Forum 2020 was held from December 18th to 19th in Beijing, China. Themed “International AI Cooperation and Governance in the Post-COVID World”, the forum focused on both the opportunities and challenges posed by AI technologies. The representatives shared understandings on AI cooperation and governance, exchanged the wisdom of international governance, and proposed academic agendas for future AI governance research.

II. Outcomes

1. Consensuses on strengthening international AI governance were reached among representatives

Firstly, responsible AI technologies should be developed, including data analysis models, prediction-based epidemic prevention, and applications that help improve health. In this process, in addition to addressing the challenges surrounding data quality, algorithms and computing power, we need to pay attention to AI-associated problems that call for prompt solutions, such as the impact of AI on public security, ethics, employment, market competition and privacy protection, as well as governance issues in the development of digital governments and smart cities.

Secondly, international cooperation needs to be enhanced with in-depth dialogues and exchanges at the international level. No single country or company can design comprehensive and anticipatory guidelines to manage the rise of AI and its ripple effects around the globe. Countries must come together to create an international cooperation and governance framework for AI, and mobilize all parties concerned to actively engage in and jointly push forward the cause. We have seen how multilateralism and international cooperation have been essential to the pandemic response around the world and the conversation surrounding AI must likewise be a global one. To this end, an inclusive community of international cooperation or an international governance committee for AI can be established and make concerted efforts to study, discuss and follow the constructive recommendations and advice offered by different partners to form an international framework for AI cooperation and governance. We can also learn from existing international mechanisms to prompt international cooperation on AI.

Third, governance principles should be formulated to ensure the security of relevant technologies. AI has the potential for social progress, yet at the same time, there are risks of data abuse and infringement on individual privacy. On this account, great care is needed to prevent AI from expanding inequalities and creating new challenges, which would consequently undermine global progress towards achieving the Sustainable Development Goals (SDGs). In response to a situation where AI technologies develop too fast while governance principles are introduced slowly and gradually, we should resort to "agile governance." In the meantime, the commonalities of all stakeholders should be identified and continuous discussions and exchanges should be engaged in to find universally applicable principles and rules of AI, to ensure that AI development benefits mankind.

2. Summarized the advanced experience of AI applications in COVID-19 response

The forum focused on the important role of AI technologies in facilitating precise identification, adopting targeted prevention and control measures, and supporting the resumption of work and production, especially best practices across the globe.

However, as the security systems and digitalization degree vary by country, we are now confronted with the urgent task of digital transformation and governance. It is necessary to encourage clear approaches to global cooperation and coordination. In the development of AI technologies, we also need to pay close attention to ethical issues. The boundary and core of AI ethical risks is to develop responsible AI and regulating the application based on an effective combination of technological advancement, governments and laws.

Besides seeking the instant improvement of AI technologies, we should find ways to make the AI systems more robust. Considering healthcare issues during our fight against COVID-19, we should turn our attention to the balance between individual and collective benefits and the sharing of health data, so as to deliver benefits to everyone.

3. Emphasized the role of AI in sustainable development

At present, the role of AI governance in facilitating sustainable development is not balanced, with significant research conducted in the field of education and healthcare but few in climate action, zero hunger, life on land, and many other areas. As the experts noted, the 17 SDGs are not an independent list of goals. When AI technologies impact any of the SDGs, the others are likely to be affected as well. Therefore, we should view the SDGs as an integrated framework, and pursue long-term impact with AI technologies. Moreover, countries should be encouraged to discuss AI strategies openly for the realization of our shared positive vision across different cultural backgrounds.

The digital divide still exists within and among countries. Largely coming from rich countries and communities, the current datasets used to train AI can have inherent biases and result in machines with insufficient knowledge of low-income populations. On this account, more attention should be paid to AI applications for medium- and low-income groups. Possible measures include establishing an effective check-and-balance system between governments and companies, introducing a third party for independent assessment, drawing a “red line” of unacceptable AI applications with the focus on privacy and biases, and designing tools that enable the execution of socio-legal accountability and technical accountability, such as auditing and record-keeping mechanisms.

4. Discussed international security challenges and opportunities brought by AI

The potential challenges AI poses to international security are mainly manifested in the uncontrollable AI arms races and the abuse of AI weapons around the world. The abuse or deviation of AI technologies may undermine strategic stability, change the rules of engagement, and add to military risks. In the field of international security, the security of training data for AI technologies should be addressed following at least three principles, upholding multilateralism, balancing security and development, and ensuring fairness and justice. When it comes to AI-empowered platforms, countries should, on the basis of multi-stakeholder engagement and within the framework of the United Nations, refrain from unexpected conflict escalations and the abuse of intelligent weapons through mutual monitoring according to the principles of proportionality and distinction.

As global leaders in AI technology, China and the US should play an active role in the development of international AI security norms. Through dialogues at the government level as well as unofficial forums, both sides should further discuss the inherent relationship between AI and international security, especially the national security risks posed by AI, so as to mitigate the potential risks. The two sides should also grasp the current opportunities to arrive at a consensus on the actual steps to be taken against potential national security risks resulting from AI, and to regulate and lead the development and application of AI technologies.

5. Proposed a cooperation initiative on international AI governance

At the current stage, what’s most lacking in the existing AI ethical principles are voices from the civil society, especially from the underprivileged population. Their concerns about AI and resistance to AI

deserve full consideration by policymakers. They should have the chance to determine what kind of world they would like to construct. The panel experts held that the general public needs to be empowered through “the democratization of knowledge”, so that they can participate in developing solutions to AI governance issues and designing AI development models.

To cope with the challenges AI poses to international governance, there is an urgent need for auditors familiar with AI governance to ensure the effective operation of governance frameworks. Meanwhile, companies should reflect on their fundamental business models to adapt to the changes.

From a global perspective, it is also urgent to establish a resource network composed of scientists and relevant experts worldwide, which will support the development of AI governance systems in each country. Countries should not only care about their own digital sovereignty and economic interests but consider the accessibility of responsible AI in a more comprehensive manner. International cooperation should also be leveraged to maximize the role of AI in generating social benefits.

6. Elaborated the important connotations of security and safety of AI and data

AI and data security and safety consist of two parts, safety, from an internal perspective, and security, from an external perspective. Specifically, safety, being inward-looking, requires preventive measures against unintentional harms, thus protecting the environment within the system. In contrast, security, skewing towards guarding against attacks from the outside, focuses on preventing behaviors that intentionally harm individuals, organizations, and properties, thus protecting the system within the environment.

Data and AI security and safety face both short-term challenges and long-term uncertainties. Short-term challenges refer to internal risks and external defense issues concerning AI safety and security. AI uses large amounts of private data that often need to be uploaded onto the cloud for analysis and computing, which risks more serious data leakage and other security problems, as well as cross-border derivatives of data breaches. Long-term challenges derive from the great uncertainties intrinsic to generalized AI and superintelligence, encompassing those posed by a superintelligence, technological singularity, and the survival risks of the human society as a whole.

Countries across the globe should improve dialogue and promote cooperation to reduce vulnerability, advocate and encourage data security governance worldwide, and integrate the data security strategies worldwide for the benefit of mankind. Countries and organizations should join hands to develop clearer and more effective laws and regulations, construct unified security assessment principles and criteria for AI models and systems, engage in scientific research and technological exchanges in a more open and collaborative manner on a global basis, facilitate the faster application of AI technologies without compromising security, and build a more secure and better world. Ethical rules for responsible AI development and data use should be put in place at the corporate, national and even international levels. AI and data products should be developed following the idea of “ethics by design,” so as to integrate fairness, security and other ethical values into the whole process of product design. Also, guaranteed measures guarding against abuse need to be instituted.

III. Summaries

1. Opening Remarks

Wang Qinmin, Vice Chairman of the 12th National Committee of the Chinese People's Political Consultative Conference (CPPCC) and former Chairperson of the All-China Federation of Industry and Commerce (ACFIC), gave the opening remarks at the forum. He noted that, in the development and application of AI technologies, in addition to addressing the challenges surrounding data quality, algorithms and computing power, we need to pay attention to AI-associated prominent problems that call for prompt solutions, such as the impacts of AI applications on public security, ethics, employment, market competition and privacy protection, as well as governance issues in the development of digital governments and smart cities. As a world-renowned, high-level research-led university, Tsinghua University should leverage its global academic and policy clout in the field of AI governance, assume its deserved role in exploring the academic frontier and serving the major needs of China, constantly enhance its capacity to innovate and strive to advance the academic frontier, thus actively offering advice on national development and contributing wisdom to human civilization.

Qiu Yong, president of Tsinghua University held that, universities—where the AI ideas originate, where new trails blazed for social progress and where the future of mankind is nurtured—need to open up new prospects and make new accomplishments in AI cooperation and governance: they should lead the development of AI, constantly innovate AI theories and technologies, and vigorously advance global AI cooperation. In June this year, Tsinghua University set up the I-AIIG. Dedicated to the research on major theoretical issues and policy demands concerning international AI governance, the I-AIIG strives to construct a fair and reasonable international AI governance system that provides intellectual support for coping with global major challenges.

Beate Trankmann, Resident Representative at UNDP China started her welcoming remarks by emphasizing the importance of international cooperation and multilateralism in the pandemic response. Similarly, the discussions on AI need to be global. She pointed out that the effectiveness of AI systems relies on data, which can lead to data abuse and infringement on individual privacy, a problem that has only been exacerbated by the proliferation of data at an exponential rate in the digital age. We must be cautious, to ensure that AI supports human development, rather than expanding inequalities and creating new challenges, which would consequently undermine global progress towards achieving the Sustainable Development Goals. She urged investments in both digital infrastructure and education to ensure that people can benefit from technological innovations and have the enhanced capabilities necessary to engage in the new world of work being shaped by AI and automation.

Fabrizio Hochschild, UN Under-Secretary-General, Special Adviser for the UN Secretary-General on digital cooperation believes that the onset of the COVID-19 pandemic has only accelerated and accentuated the use of AI, and the most pronounced proliferation will probably be in the healthcare sector. Despite the massive potentials, he laid out several challenges including our increasing reliance on AI's predictive capabilities which makes us vulnerable to systemic errors, cyber-attacks, lethal autonomous weapons, and the widening digital divide. Of all the emerging technologies, artificial intelligence stands alone as the one with the greatest potential to empower, but also to disrupt. He urged international collaboration to ensure that AI is used in a transparent and trustworthy manner that upholds human rights and human dignity, that promotes our safety and security, and fosters

inclusive peace.

Zhao Houlin, Secretary-General of the International Telecommunication Union (ITU), presented via video link, said that as the international governance environment and framework building is critical to the making of AI market rules and social ethical norms, all parties concerned should be mobilized to actively engage in the cause. He pledged that the ITU will work together with Tsinghua University to advance technological innovation and collaborative governance of AI globally, with an aim of delivering greater benefits out of AI for mankind.

Xu Jing, Director-General of the Department of Strategy and Planning, the Ministry of Science and Technology (MOST) noted in his keynote speech that the Chinese government has attached great importance to AI governance issues. In handling AI governance, we need to stay committed to scientific and technological advances, work towards the goal of improving people's wellbeing, hold the bottom line of ensuring security and controllability, and uphold the philosophy of global co-governance. He hoped that Tsinghua University will, on the basis of the I-AIIG, conduct thorough research and advance exchanges and dialogues, so as to pool global wisdom for the building of a consensus on cooperation.

Wang Xiaolong, Director-General of the Department of International Economic Affairs, the Ministry of Foreign Affairs (MOFA), said in the keynote speech that, effective AI governance entails the involvement and concerted efforts of governments, the scientific research community, industries, think tanks and other stakeholders. Since its founding, the I-AIIG has worked to build a bridge for international collaboration and played an important role in consolidating AI cooperation between China and other countries. In the future, when working on AI governance, we need to ensure tech

for social good, create a beneficial environment of open cooperation, set inclusive and fair rules, uphold the principle of extensive consultation, joint contribution, and shared benefits, refrain from "decoupling and technology blockade", respect the governance sovereignty and legislations of different countries, and create space for the iteration of AI technologies.

Zhu Xiumei, Deputy Director-General of the Science and Technology Department, the Ministry of Industry and Information Technology (MIIT), stressed in the keynote speech that we need to pay heightened attention to AI governance, step up international cooperation on AI and pool the wisdom of the international community together, so as to generate positive impact from AI development and governance. We should build the broadest possible consensus on AI governance issues to build a sound momentum for common development and governance consultation. The Tsinghua University International AI Cooperation and Governance Forum 2020, which seeks to initiate in-depth dialogue and exchanges on international AI governance, has a positive and practical influence. She hoped that the forum can help contribute China's wisdom to the building of international AI governance system, and deliver pragmatic outcomes for international scientific and technological exchanges and cooperation.

2. Main Plenary

Andrew Chi-Chih Yao, Turing Award Winner and Dean of the Tsinghua Institute for Interdisciplinary Information Sciences (IIIS), took the Secure Multi-Party Computation (MPC) technology as an example to pinpoint that encryption can play a bigger part in data governance, and recommended that we combine a series of credible algorithms with relevant institutions, and join hands to lay the foundation of credible data governance.

Zhang Bo, Academician of the Chinese Academy of Sciences (CAS)

and Honorary Director of the Tsinghua Institute for Artificial Intelligence (IAI), started by talking about the current vulnerabilities of AI. In addition to introducing the latest work of the IAI on third-generation AI, he argued that AI development and governance should grow in tandem and reinforce each other, and called on countries to unite in developing the secure, reliable, credible and expandable AI technologies of the third generation.

Yolanda Gil, former President of the Association for the Advancement of Artificial Intelligence (AAAI), believed that the diversified and in-depth development of AI technologies has facilitated scientific research and set the stage for tackling more social problems. She pointed out the need to enhance international cooperation and data sharing in the field of AI, so as to further boost scientific research, especially interdisciplinary research.

Chaesub Lee, Director of the ITU Telecommunication Standardization Bureau, presented via video link the work of ITU in the field of AI, and held that the stakeholders should increase dialogue and cooperation, discuss ways to enable AI to better serve the human society and facilitate the realization of the SDGs.

Fu Ying, Honorary Dean of I-AIIG and former Vice Foreign Minister of China, called for establishing an inclusive international governance committee for AI, and making concerted efforts to study, discuss and follow the sound recommendations and advice offered by different parties concerned and put in place a common set of international norms. She said, “China has taken practical steps in the governance and legislation of artificial intelligence. Of course, from a broader perspective, this is a challenge for all of humanity, not an issue that one or two countries can solve alone. It is vital that international agencies and countries work together on this topic, and we believe that artificial intelligence should

ultimately benefit all of humanity.”

Zhang Yaqin, Dean of the Institute for AI Industry Research at Tsinghua University, highlighted three dimensions for AI governance. The first is to develop responsible AI through technological innovation, including data analysis models, prediction-based epidemic prevention, and applications which help improve health. The second is to enhance international cooperation and carry out in-depth dialogues and exchanges at the international level, in a bid to cope with the most urgent issues. The third is to promote the use of AI at a large scale while ensuring the security of relevant technologies at the same time.

Xue Lan, Dean of I-AIIG, believed that the biggest challenge for AI development at present is that the technology is growing too fast, while governance principles, as a social system, develop in a slow and gradual manner. Therefore, we need to resort to “agile governance”. “In handling AI, we can learn from the existing international regimes, and constantly advance international cooperation on AI in the future. What’s more, we need to identify commonalities of all stakeholders and engage in continuous discussion and exchanges to find universally applicable principles and rules of AI, thus ensuring that AI develops soundly and benefits mankind,” he said.

Wang Lei, Coordinator for Cyber Affairs of the MOFA, made it clear that data security is crucial for AI governance. Based on the discussions on multilateral platforms like the UN for years as well as data security governance practices of different countries, the Chinese government proposed the Global Initiative on Data Security to provide constructive solutions to major data security issues concerning various parties. Countries should uphold multilateralism and work jointly to advance data

security and AI governance.

Kay Firth-Butterfield, Head of Artificial Intelligence and Machine Learning at the World Economic Forum (WEF), held that AI should be in the service of mankind. While benefiting from AI, we need to maintain control over the technologies. Thus, it is of vital importance to hold discussions on AI ethics. Relevant multi-stakeholders should strengthen international cooperation on setting ethical rules for AI.

3. Thematic Sessions

Thematic session #1: Role of AI in combating Covid-19: what are the lessons?

The thematic session - "Role of AI in combating Covid-19: what are the lessons?" was led by I-AIIG. Co-organizers of the session include Tsinghua University's Institute for Artificial Intelligence, the China Institute for Science and Technology Policy, the Center for Science & Technology Development and Governance, and the Center for Industrial Development and Environmental Governance. The event was moderated by Professor Liang Zheng, Vice Dean of I-AIIG. Professor Xue Lan, Dean of I-AIIG, delivered welcome remarks.

Yu Yang, Director for International Academic Exchange at I-AIIG and Assistant Professor of the Institute for Interdisciplinary Information Sciences at Tsinghua University, believes that information and data technologies used for the pandemic responses in China, as represented by AI, feature two characteristics—comprehensiveness and rapidity. China's pandemic response is achieved through an integrated system of co-governance by an AI-adaptive government and proactive and capable enterprises. AI technologies' comprehensive and rapid penetration into China's pandemic response relied on three key factors: proactive corporate

engagement, an algorithmically-oriented government, and the pivotal units based in both government and corporate.

Effy Vayena, Co-chair of the WHO's expert advisory group on Artificial Intelligence health ethics and governance and Professor at the Swiss Federal Institute of Technology (ETHZ) for Big Data and Artificial Intelligence Ethics, noted the urgency for digital transformation and governance, and called for clear collaboration and coordination across countries due to different security systems and varying levels of digitization in individual states. As AI technologies advance, however, AI ethics also requires our attention. For example, European governments have sought to protect people from the pandemic without violating the rule of law, democracy and fundamental rights.

Xia Huaxia, Vice President and Chief Scientist at Meituan, applauded the role of AI technologies in pandemic response, which have been used for infrared thermal imagers, unmanned delivery robots, remote working, online education, live-stream shopping and many other functions, stating that ethical concerns of AI can be addressed with improved technology. For example, end-user data processing can be realized through AI technologies to better ensure user privacy. There are nevertheless certain AI issues that cannot be easily resolved with technology, and therefore depend on better government regulation and legislation on the use of AI technologies.

Shi Jun, President of the Asia Pacific Business Group and Vice President for Strategic Planning at SenseTime, believes that AI will continue to emerge as the center of gravity in post-pandemic global competition. Given the accelerated development of AI technologies and data governance, as we have seen in countries, there is a need to further

refine regulation and governance rules. Pointing out that the boundary and core of the ethical risks of AI lie in the development of responsible AI, he suggested that AI enterprises think and act one step ahead on relevant governance and compliance issues by taking regulatory initiatives to establish dynamic management mechanisms.

The first panel of the session was themed "The innovative technology solutions for fighting COVID-19: what have been done globally?" Panelists shared practical experience of how high-tech companies used AI to help fight the pandemic. The panel was moderated by Professor Liang Zheng.

Hsiao-Wuen Hon, Managing Director of Microsoft Research Asia, resorted to three "R"s to encapsulate the pandemic—response, recovery, and reimagining, outlining the important impact of AI technologies in these three phases. For instance, in the response phase, AI technologies enabled telemedicine and remote working via online communication; in the recovery phase, they allowed rapid tracking with health QR codes and facilitated vaccine development; and lastly, the future of work in the post-pandemic era is reimagined with the emergence of hybrid online-offline working style.

Song Jiqiang, Managing Director of Intel Lab China, introduced how AI technologies have been used for testing, treatment and prevention during the pandemic: For testing, Intel developed AI technologies to help realize preliminary lung CT scanning as a screening tool for COVID-19, which significantly improved testing efficiency during the early stage of the outbreak; for treatment, medical robots have reduced infectious risks for healthcare workers; and for prevention, AI technologies have facilitated quality control of face mask production, thereby boosting production efficiency. AI technologies, therefore, have made multi-faceted contribution

to controlling and preventing COVID-19 outbreaks.

Tang Jian, Chief Scientist of Intelligent Control at DiDi, presented on DiDi's AI efforts to empower the fight against COVID-19 overseas, including offering free GPU computing power to research institutions and businesses combating COVID-19 and using AI algorithms to detect whether drivers are wearing face masks, which have greatly empowered overseas communities.

Wang Jingjing, Vice Director of the Institute of Public and Environmental Affairs, shared how the organization supported pandemic response and preparedness by upgrading its Blue Map shortly after the outset of the pandemic, making it the first map application in China that shows the spread of COVID-19 down to the district level. The app has played a crucial role in alleviating public panic since the start.

Kong Qiushi, Deputy Director of Xiaoshan District Statistics Resource Management Bureau, Hangzhou, touched on the lessons learned from the district's AI-empowered pandemic response: the local government set up a data team immediately after the outbreak to develop and deploy smart applications for disease prevention and control, including smart conversational robots, the City Brain digital cockpit, the Fighting COVID-19 app (for information verification and reporting), and the Xiaoshan Landlord Union app, etc. By integrating AI technologies and big data for the front-line pandemic response, the district government ensured scientific decision-making, efficient epidemiological investigation, and targeted response.

The second panel, titled "The outlook of post-pandemic governance of AI in the globe," was moderated by Dr. Zhang Min, Vice Director of the State Key Lab of Intelligent Technology and Systems under the Department

of Computer Science and Technology at Tsinghua University.

Jiang Yan, Vice President at Megvii, briefed the panel on the company's key efforts in combating COVID-19, including an emergency R&D team set up in the early stage of the outbreak that promptly developed an AI-assisted temperature measurement system and its expeditious rollout in busy areas such as hospitals, transportation hubs and supermarkets, which has deterred the risks of virus transmission caused by heavy pedestrian traffic and served as an effective buffer against the COVID-19.

Jenny (Yu) Zeng, the Managing Partner at MSA Capital, weighed in from the perspective of a venture capitalist on the changes from the pandemic—the rapid growth in remote working, online education, telemedicine, and remote medical consultation, all of which was spurred by AI breakthroughs. Not only that, AI and other technologies have helped create effective synergies across businesses, NGOs and governments in infection prevention and control.

Tokuchi Tastsuhito, Executive Director at the Center for Industrial Development and Environmental Governance at Tsinghua University and former Managing Director at CITIC Securities Co., Ltd., looked to the future of AI international governance in the post-pandemic world by reflecting on the lessons from Japan's pandemic response. He believes that regarding AI digital governance, at least the following issues need to be addressed: firstly, the ownership of data, the way data should be used and shared, and data misuse accountability; secondly, personal privacy and security protection; thirdly, the balancing between the state, platform companies and society; and fourth, maintaining the integrity of the Internet as current international politics drives it toward fragmentation.

Adrian Weller, Principal Research Fellow in the Department of Engineering at the University of Cambridge and Programme Director for AI at The Alan Turing Institute, believes that instantaneous optimization should not be our sole pursuit for AI technologies; we also need to think about how AI systems can be made more robust. Living through a pandemic, we need to start finding a balance between individual and collective wellbeing. For example, how should health data be shared so that everyone benefits from it? What are the appropriate ways to use digital tracking technologies as a tool to combat a pandemic? These questions go beyond any efforts of individual countries; they press for the establishment of pertinent international principles and norms.

Rohinton Medhora, President of the Center for International Governance Innovation, offered his thoughts on the COVID-19 pandemic. Firstly, the pandemic has underlined the severity of the digital divide today: We might have grown accustomed to online communication, but half of the world still have no Internet access. Such divide has posed challenges to cybersecurity and supply chain resilience. Secondly, the pandemic has amplified the urgency and complexity to address concerns over the extensive application of AI technologies, including data collection, storage and use in contact tracking apps for the pandemic response, as well as the slippery slope from data to other governance issues such as privacy, security, and human rights.

Thematic Session #2：AI governance for Sustainable Development Goals

The second thematic session was hosted by UNDP China, with three keynote speeches and a panel discussion joined by leading international AI experts, sharing their perspectives from academic, private, and public

sectors. Thomas Davin, the Director of Innovation at UNICEF believes that there is so much AI can do if directed in the right way. He laid out two main challenges that AI brings. First is the capacity gap, that there is not enough expertise in the world to unleash AI's potential, and the AI talents tend to concentrate in the rich countries. Second is the technology gap, that half of the world is not connected digitally today. The data that current AI algorithms use is coming mostly from the richer world, which means that they are almost operating in a "bubble", embedding the assumptions and solving problems for the developed world. To solve these challenges, Thomas brought up two potential solutions. The first is to democratize the use of AI and invest in low- and middle-income countries. The second is to establish a governance system that allows external third parties to understand how the algorithms function. He introduced UNICEF's work in advancing the solutions by issuing the policy guidance on AI for children, as well as investing in tech ventures that sought to benefit the 1 billion most vulnerable people.

MIT Professor Max Tegmark, co-founder of the Future of Life Institute elaborated two necessary prerequisites for successful international AI cooperation: a shared positive vision and clear red lines. He believes that AI is not a zero-sum game. Just as a shared positive vision has allowed us to eradicate smallpox even during geopolitical tension in the 1980s, the Sustainable Development Goals (SDGs) is another shared positive vision we need to strive for. A study by Prof. Tegmark and the collaborators found that AI may help to achieve 79.3%, 92.6% and 63.3% of the society, environment, and economy-related SDG targets while hindering 37.8%, 29.5% and 31.7% of them respectively. As a result, AI scientists can learn from experts such as biologists and the chemists to set clear red lines for AI applications and steer towards globally beneficial AI. He further discussed why China is uniquely positioned to do so, being a world-leading science and technology power with growing international influence, and one of the oldest surviving civilizations with a successful tradition of long-term planning.

Prof Zeng Yi is the co-director of China-UK Research Center for AI Ethics and Governance at the Institute of Automation, Chinese Academy of Sciences. He started by sharing that according to an analysis of all the research cases collected by the AI4SDGs Think Tank, most cases are associated with health and education, while few focus on environment-related SDGs. On a larger scale, only 0.1% of all the AI-related publications contribute directly to SDGs. He shared three cases where AI can be beneficial or harmful to social good: 1) Emerging AI monitoring in classrooms to regulate students' behavior is concerning but could prevent campus bullying. 2) AI can be used to monitor and analyze biodiversity data to better protect wild animals and improve farm animal welfare. 3) Facial recognition has helped contact-tracing to combat COVID-19 but has also raised privacy concerns. He brought up an interesting observation that while globally, AI is seen as a tool, Eastern cultures like Japan and China tend to view AI also as a partner, while western societies tend to see it as a competitor. He believes that in the future sustainable symbiotic society we will see humans, animals, other living beings and the environment in harmony.

The panel discussion was moderated by Wei Zhang, the Assistant Resident Representative of UNDP China and joined by the following panelists: Christian Guttmann, Executive Director of the Nordic Artificial Intelligence Institute; Zhu Xufeng, Executive Director of the Institute for Sustainable Development Goals at Tsinghua University; Rohan Samarajiva, Chairman of LIRNEasia; Mila Romanoff, Data Privacy Specialist at UN Global Pulse; Nannan Lundin, Innovation Counselor at the Swedish

Embassy and Zheng Zhitai, Chief Strategist at Horizon Robotics.

The panel first discussed whether AI is an enabler or an inhibitor of SDGs. Many examples of AI's positive application were given including predicting floods, monitoring food security, assessing gender gaps etc. However, in order to make them more mainstream, public trust in AI needs to be established. In addition, the data availability gap can be addressed as some AI algorithms do not need massive datasets. On the topic of an ideal AI governance framework, the panelists proposed that it needs to be globally inclusive of cultural and societal differences, and also has to be practical and relevant to business leaders to provide investment certainties. In terms of the balance between regulation and innovation, clear red lines should be drawn, and a sandbox approach can be used to test new applications. Innovation and policy should move in the same direction united by the SDGs instead of one struggling to keep pace with the other.

Thematic Session #3: AI and International Security: Challenges and Opportunities

The Center for International Security and Strategy (CISS), Tsinghua University, organized a thematic session titled "AI and International Security: Challenges and Opportunities". The session highlighted preliminary outcomes of AI and International Security, a joint research project between CISS and the Brookings Institution with the support of the Berggruen Institute and the Minderoo Foundation.

John Allen, President of the Brookings Institution, noted in his keynote speech that the US-China strategic competition in AI and other novel technologies has become increasingly visible, and narratives of an "AI arms race" have not been in short supply. He hopes that to avoid a runaway arms race, the security concerns discussed in the AI and International Security project will be further explored at both the governmental level and in non-governmental fora, particularly with regard to the challenges and risks posed by AI in four areas: the limits of AI technologies, the risks of AI-inflicted conflict escalation, risks created by AI technology proliferation, and humanitarian risks, in order to reduce the likelihood that those risks come to pass. As we are at a critical juncture in regulating the development and application of these technologies, he called on both governments and expert communities in the US and China to examine the opportunity that now exists to develop agreements around practical steps to reduce national security risks posed by AI.

Fu Ying, Chair of CISS and Honorary Dean of I-AIIG, focused her keynote speech on China's position on international cooperation regarding AI governance, recalling Chinese President Xi Jinping's message at the G20 Summit in November, 2020, which emphasized China's support for further dialogues and meetings on AI to push for the implementation of the G20 AI Principles and set the course for the constructive development of AI globally. In September 2020, Chinese State Councilor and Foreign Minister Wang Yi proposed a "global initiative on data security" in hopes that the international community would reach an agreement on AI security on the basis of universal participation, and expressed his support for the affirmation of the commitments in the initiative through bilateral or multilateral agreements. He laid out three principles for effectively addressing data security risks: upholding multilateralism, balancing security and development, and ensuring fairness and justice. Fu Ying pointed out that the joint China-US research project on AI security governance is a success story from which both sides have benefited greatly, and the findings are particularly relevant today. As we come to terms with the inevitable weaponization of AI technologies, right pathways to AI governance must be identified. We need to learn from history. Our consensus on the

governance of nuclear weapons, for instance, came too late, once posing an immense threat for all humankind. Another lesson learned is the early absence of Internet governance. This time, we hope that the governance of AI technologies, will stay ahead of the technological advances, and that we would fully understand their risks and reach an earlier consensus on their governance. Pugwash is a name that often came up in the discussions about AI international governance. The Pugwash Conferences on Science and World Affairs is a reputed institution for its expertise in the governance of nuclear weapons and has played a positive role in nuclear arms control. A similar approach could be taken for AI governance by establishing a Select Committee on AI International Governance that brings together researchers, governments and policy professionals.

In the panel session, participants took an in-depth look at how China and the US might cooperate in AI governance, and agreed that both of the two major AI powers need to guard against the risks from AI-enabled military technologies and improve their governance. The abuse or misuse of AI technologies could send shockwaves to strategic stability, change the rules of engagement, and intensify military risks. China and the US share a strong need for risk management as well as considerable interest convergence, creating tremendous potential for cooperation on AI international governance. As noted by some panelists, the use of AI must comply with the principles of proportionality and distinction in international law during data collection, algorithm training, and battlefield operations. AI-enabled weapons must not be used to attack or cause excessive harm to civilians or civilian facilities. In this connection, Chinese experts proposed a "traffic light" rule for data collection to prohibit AI-enabled weapons from attacking civilian targets and restrict their targets to fully identified military ones, thereby securing greater consonance with international law. The thematic session concluded with a shared agreement among the

participants that countries around the world, China and the US included, could build trust and clear misgivings through policy and document exchanges as well as academic dialogues in order to strive for international consensus on AI security governance.

Thematic Session #4: International cooperation on AI governance

UNDP China organized the fourth thematic session, to discuss the way forward of international cooperation on AI governance. Zia Khan, Senior Vice President of Rockefeller Foundation, started by stating that while many AI ethics principles have been published, few concrete governmental initiatives have been implemented. Furthermore, ethics protocols alone are insufficient in ensuring the development of ethical AI. We need to develop rules tailored for AI, otherwise it will be governed by outdated principles. He shared a few takeaways from his work: 1) each stakeholder speaks a different AI language, we need to be mindful of how these laws may be perceived in different socio-political environments; 2) AI technologies are being developed by multinational companies, yet most regulatory structures are at the national level; 3) we need urgent consensus for AI governance in life-threatening situations like pandemic and war; 4) independent agencies need to assess safety, efficacy and fairness of AI applications.

Yeong Zee Kin, the Assistant Chief Executive of the Infocomm Media Development Authority of Singapore (IMDA) believes that the international AI governance framework should be forward-thinking, open and interoperable, and at the same time commercially sensible, practical and realistic. He elaborated on Singapore's effort in AI governance with two pillars: laws that emphasize personal data protection and tools to encourage trust in AI development, such as "trusted data sharing framework", "data regulatory sandbox" etc. A suite of AI governance initiatives has been implemented to support industry adoption of responsible AI, including self-

assessment guidelines and case studies. In addition, Singapore has started training and certification programs for professionals in AI governance. He believes none of these efforts can be done in silos, but rather must have collaboration. Moving forward, these frameworks will be actively reviewed to ensure relevance, with "human-centricity" as the North Star.

The panel discussion was moderated by Devanand Ramiah, the Deputy Resident Representative of UNDP China and joined by the following panelists: Jae Moon, Director of the Institute for Future Government at Yonsei University; Cyrus Hodes, Vice President and Chair of the AI Initiative at the Future Society; Rudradeb Mitra, founder of Omdena; Surina Shukri, CEO of Malaysia Digital Economy Corporation (MDEC); Wilson Wong, Director of Data Science and Policy Studies Programme (DSPS) at The Chinese University of Hong Kong; Mark Findlay, Director of the Center for AI and Data Governance at Singapore Management University and Warren Hero, CIO of Webber Wentzel.

The panel discussed what international AI cooperation and governance would look like and two sets of opposite approaches were brought up: reactive/ paternalistic approaches and top-down/ bottom-up approaches. The panelists agreed that there needs to be a balance of reactive/ paternalistic approaches to encourage innovation in AI while mitigating the risks. Some advocated for a more bottom-up approach as most AI ethical principles lack the voice of civil society. There was also concern that while AI technologies have massive potential to address global challenges, most of the innovations are fueled by economic interest. On the topic of digital sovereignty, several panelists agreed that AI should not be promoted as a national interest, responsible access should be prioritized rather than regressive protectionism.

Thematic Session #5: Security and Safety of Data and AI

The thematic session - "Security and Safety of Data and AI" was organized by the Beijing Academy of Artificial Intelligence (BAAI). The session was moderated by Professor Zeng Yi, Director for the Research Center for AI Ethics and Safety at BAAI and Director of the China-UK Research Center for AI Ethics and Governance at the Institute of Automation, Chinese Academy of Sciences.

Danit Gal, Associate Fellow at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge, drew a distinction between two important AI concepts——safety and security: AI safety is more internally oriented, referring to preventive measures against unintentional harm that protect the environment from within the system, whereas AI security deals more with defense against external attacks that would inflict intentional harm to individuals, organizations and property to protect the system from the environment. She believes that technology affects us all and so does its safety and security, yet our ability to address these impacts is utterly outpaced by the advances and application of the technologies. To complicate the dynamics further, AI and data flows have introduced and connected new opportunities and vulnerabilities. She suggested to reduce vulnerabilities and to promote and galvanize global governance of data security and safety through enhanced communication and collaboration.

Vincent Müller, Professor of Philosophy at Eindhoven University of Technology, gave a keynote presentation on the long-term risks of superintelligence and general AI in two moves: from superintelligence to technological singularity, and from that to existential risks. Superintelligence is defined as artificial intelligence that greatly exceeds the best cognitive performance of humans in virtually all domains of interests. According to Professor Müller, the shift from superintelligence to technological

singularity is based on three presuppositions: accelerated speed and more data, no need for cognitive science, and imperative exigency. The shift from technological singularity to existential risks depends on two presuppositions: the applicability of the rational choice theory to AI, and the orthogonality of intelligence and final goals.

Roman V. Yampolskiy, professor from the University of Louisville commented on the future of AI, security and defense in his keynote speech, noting that superintelligence is coming and with it comes SuperSmart, SuperComplex, SuperFast, SuperControlling, and SuperViruses. Touching upon concerns with AI, he highlighted some recent research including "Taxonomy of Pathways to Dangerous Artificial Intelligence," "Safe AI—is this possible?" "Mitigating Negative Impact," and "Limitations of AI," etc. He argued that the timeline of AI failures has an exponential trend, and AI failures will increase in frequency and severity proportionate to AIs' capability.

In the panel session, panelists shared their insights on the outlook for the opportunities, challenges and future development regarding the security and safety of data and AI.

Cao Jianfeng, Senior Fellow at Tencent Research Institute, shared his views on algorithmic bias and deep synthesis, and proposed suggestions to address relevant issues. Current AI systems require gigantic amounts of data for algorithm training, which calls for reasonable control over the quality of such data during AI development. However, addressing algorithmic bias takes more than that—ethical principles must be observed by AI practitioners. He then proposed four suggestions to address the inherent tension between AI development and privacy protection. Firstly, the principle of "ethical by design" must be followed when developing AI and

data products, so that ethical values such as fairness and security as well as safeguards against data abuse are incorporated into the whole cycle of product design. Secondly, ethics guidelines for responsible AI development and use of data are needed at the corporate, national and international levels. Thirdly, we could share our best practices across the international community. Fourth, the industry needs to explore privacy-preserving AI technologies, such as federated learning, to achieve a better balance between privacy protection and the use of data.

Wu Shenkuo, Executive Director of the International Center for Network Rule of Law at Beijing Normal University, called for an international community response to the emerging challenges and legal loopholes brought by data violations in the new technological landscape. Firstly, there is inadequate judicial regulation of novel types of data violations due to the absence of and differences in laws and regulations in countries and regions. Secondly, timely and efficient cross-border evidence collection and exchange cannot be absolutely guaranteed under the existing legal frameworks across countries. Finally, in the era of cyber anonymity, some countries find themselves increasingly disadvantaged in effectively responding to growing vulnerabilities with their current monitoring, early warning and contingency mechanisms. As globalization deepens, therefore, transnational data violations will also evolve, which requires stronger and more cooperative international governance to safeguard a world of trust.

Deng Yafeng, Director of Qihoo 360 Artificial Intelligence Institute, summarized potential AI threats in terms of data safety and security. Firstly, as a booming new technology, AI applications have bred security risks throughout the process of their development and deployment. Secondly, applications empowered by AI and big data generally involve large amounts of private data which is often sent to the "cloud" for analysis

and computation, leading to even more serious leakage and security risks than would otherwise be the case. Thirdly, although AI in human-like bodies will not, technically, have true self-awareness for some time, it is not impossible that they will be exploited by certain ill-intended persons, which would cause dreadful social and ethical crises. Finally, AI will inevitably take the jobs, in the short term, of those performing uncomplicated and repetitive tasks, which will trigger people's fear and rejection. On top of that, immature AI technologies can be perceived as discriminatory in their recognition results. He stated that the security and safety of data and AI is a global challenge that demands collective global endeavors by organizations and individuals around the world.

Ⅳ. Acknowledgments

This report is based on the Form and was prepared by a team consisting: LIU Chang (Tsinghua University) and Puna Zhao (UNDP China). We thank all invited speakers for their contribution to the Forum.

Forum Website: https://www.tsinghuaaiforum.org/

I-AIIG Website: https://aiig.tsinghua.edu.cn

Contact: i-aiig@tsinghua.edu.cn


-AIIG
清华大学人工智能国际治理研究院
INSTITUTE FOR AI INTERNATIONAL GOVERNANCE
TSINGHUA UNIVERSITY