• ASIA SOCIETY NORTHERN CALIFORNIA •

EXPLORING AI ISSUES ACROSS THE UNITED STATES & CHINA

CONCLUSION

HEATHER EVANS AND BENJAMIN CEDRIC LARSEN

EXPLORING AI ISSUES ACROSS THE UNITED STATES AND CHINA

CONCLUSION & KEY TAKEAWAYS

The United States and China are superpowers when it comes to AI development and adoption. Given each county's different political and socioeconomic regimes, however, it is increasingly clear that the United States and China are taking varied approaches to the adoption and governance of artificial intelligence technologies. **By looking at and learning from the lessons of the United States and China, we may be better situated to understand where AI development and forms of governance are likely going.**

"Exploring AI Issues Across the United States & China" has covered the areas of AI Principles, Deepfakes, Affect Recognition, AI in Healthcare through a focus on Deep Learning in Drug Discovery, and Facial Recognition. In doing so, we have compared the approaches of the United States and China and have outlined important differences as well as areas of alignment in terms of AI governance.

We currently face the prospect of increasing balkanization of technological development and application, during a time of rising international tension. Such tension increases the risk of a race-to-the-bottom in terms of deploying unsafe automated systems that hold ample space for premature application as well as potentials for misuse. At the same time, rhetoric and practices supporting an "AI race" for strategic advantage have proliferated despite the risks of cutting corners on safety and governance, which only increases the likelihood of further conflict.[1]

Rapid advances in research and development position AI technologies to have transformative socioeconomic impacts in the years to come. Challenges posed by AI include the rise of synthetic media (text, audio, images, and video content);[2] insufficient transparency in the automation of decision making in complex environments;[3] the amplification of social biases from training data and design choices, which causes disparate outcomes for different races, genders, and other groups;[4] as well as the vulnerability of AI models, prone to cyberattacks and data poisoning.[5]

At the same time, AI technologies are not easily confined by national borders but are frequently deployed across online platforms accessible from around the world. Julia Chen notes in her piece on deepfakes (synthetic media produced by machine learning) that such media has already been used to commit fraud and to influence politics.[6] In the future, it is possible that deepfakes could cause a government to take military action on the basis of false information. This is an example of where collaboration at the international level is needed to avoid AI being used to sow mistrust and conflict between countries.

As a result of the above considerations, it is imperative that norms and best practices for the responsible development and use of AI technologies are fostered at the international level.

Alignment across international stakeholders would provide enhanced guidance for AI development and deployment while securing cooperation in terms of limiting potential negative externalities. There have already been a number of efforts driven by international forums such as the G20, industry organizations such as the IEEE, and multilateral institutions such as the Organization for Economic Cooperation and Development (OECD) and the United Nations to achieve consensus on best practices for aspects of AI governance.

Reaching international consensus on specific issues related to AI development and adoption is difficult, however. Obstacles include the rapid development and deployment of AI technologies, as well as AI being a general-purpose technology that has a broad impact on many industries. In terms of regulation, this means that a sector-specific and case-specific focus often is required.[7] These characteristics make AI technologies hard to regulate at the national level, which makes it even more difficult to secure alignment at the international level.

As bilateral tensions between the United States and China continue to rise, collaboration in the development, adoption, and governance of AI technologies only becomes more difficult. The US government has started to decrease the number of visas issued to students holding Chinese passports and those who have been accepted to study in STEM disciplines at American universities.[8] Concerns of intellectual property theft and espionage have impacted inter-university research collaborations, and supply chains are being severed. China continues to control and censor information infrastructure, which includes banning a number of American companies from operating in the country. The US government has also taken steps to restrict Chinese software companies' operations in recent months. **A range of tit-for-tat policies such as the creation of Entity Lists and Export Bans on AI technologies are currently raising the barriers to cooperation across the Pacific and elsewhere.**

> **This deteriorating relationship has become a part of our motivation for exploring potential pathways to cooperation, as well as charting some of the obstacles and barriers that inhibit it. International governance on AI only makes sense as long as the world's two leading developers and adopters of AI technologies are on board and are able to convene on important areas and issues.**

In terms of cooperation, bilateral trust has become an essential issue. In recent years, trust at the institutional level has deteriorated between the United States and China, complicated by different values and world views.[9] Where China promotes the virtues of cyber sovereignty as the primary organizing principle of internet governance, the United States has long supported a global, open internet.

The current US-China decoupling is not only rooted in a bifurcation of the internet and rising socio-technological incompatibilities. It is also witnessed in the institutional and ideological bifurcations in terms of rule of law vs. rule of ruler and the role of the Party-state in China.

Even though trust has been deteriorating, opportunities for two of the world's superpowers to lead the way in reducing accidents and limiting dangerous uses of AI exist. Without such measures, the early deployment of unsafe and unreliable AI systems in consequential sectors is likely to lead to unpredictable, harmful, and destabilizing outcomes.[10] Neither country will benefit from the dangerous deployment of AI technologies that increase the likelihood of catastrophes.

In terms of assessing the relationship and current climate between the United States and China, we have deployed a Green-Yellow-Red Light Framework to illustrate the perceived ease or difficulty of AI-induced cooperation across the two countries. Out of five case studies, we have found one "Green possibly turning Yellow Light" (Deep Learning in Drug Discovery); two "Yellow Light" issues (AI Principles, Deepfakes), signaling that some barriers need to be overcome before cooperation can be initiated; and two "Red Light" issues (Affect Recognition, Facial Recognition), signaling that substantial barriers must be overcome before bilateral cooperation can be initiated.

In this conclusion of the series, we offer a roadmap to understand convergence in approaches to AI governance across the United States and China. Based on the case studies of the series, we highlight where substantial divergence in governing mechanisms may create systemic barriers or irreconcilable differences to AI governance that could be hard to overcome in the long run, unless stakeholders on both sides become more willing to act and engage now.

BARRIERS TO COOPERATION

From the lessons learned in the case studies, barriers to cooperation can be grouped into three overarching categories referring to (1) cultural barriers (i.e., values and principles), (2) political barriers, and (3) economic barriers. Each of these barriers to cooperation is discussed in greater detail below.

(1) Cultural Barriers

When AI principles are being framed, they most often incorporate existing values that are reflected in society, which express a shared set of attitudes toward how AI should be governed, and on which ethics should guide or shape technological implementation and adoption.

As culture permeates different socioeconomic and political realities, it is also important for how AI technologies are applied in varying contexts. The clearest example of differing systemic attitudes to AI applications is found in the deployment of facial recognition systems across the United States and China.

While the United States is currently reevaluating the role of facial recognition technologies, spurred by strong civil rights movements and self-regulation by industry, China has been pushing ahead with more rapid deployment of surveillance systems. In his piece on Facial Recognition, Samuel Curtis argues that in the United States, civil rights movements advocate for retrenchment from automated decision making that infringes on individual liberties, while in China, great merit is placed on surveillance mechanisms that deter criminal behavior.

Jessica Cussins Newman argues in her piece on AI Principles that some Chinese scholars focus on "harmonious design," as opposed to "human-centered design," which has been a more common focus of researchers in the United States. While there are several counterexamples, such as the focus of China's Artificial Intelligence Industry Alliance (AIIA) on "people-oriented design" (以人为本), a comparison is often drawn between the prevailing historical and cultural values of individual liberty in the United States and that of a harmonious society (*hexie shehui* 和谐社会) in China. Chinese characteristics stand apart from Western ethical engineering guidelines, where responsibility in China arguably precedes freedom, obligation precedes rights, the group precedes the individual, and harmony precedes conflict. [11]

Of course, cultural differences do not have to impede international cooperation, but when underlying values and different cultural dispositions are used as political arguments for the incompatibility of socio-technological systems, new barriers to cooperation are automatically erected. This is currently the case as the United States seeks to enable safe and trustworthy use of AI technologies, embedded with "American values."

AI principles formulated in the United States tend to focus on democratic values, including freedom, diversity, privacy, and human rights. The politicization of AI principles is witnessed in America's adherence to the Global Partnership on Artificial Intelligence (GPAI), which Michael Kratsios (White House chief technology officer) has stated as being a "check on China's approach to AI," while acting as a way to counter "China's record of 'twisting technology' in ways that

threaten civil liberties."[12] The US government is generally trying to distinguish itself internationally for upholding more stringent human rights principles in its deployment of AI systems.

The politicization of AI principles comes with the risk of failing to recognize that not all societies are based on, and governed by, the same cultural values and set of principles. Samuel Curtis argues that the United States and China are unlikely to reach any semblance of high-level goal alignment in the near term, as that requires a degree of multilevel value alignment, which at present does not exist.

Stakeholders from civil society also play an important role in shaping cultural attitudes, which influence public policy making, though they function differently across the United States and China. Generally speaking, civil society and media organizations play a convening role while promoting greater civic awareness.

In the United States, civil society organizations play a critical role in affecting AI governance, especially in terms of flagging potentially discriminating uses. In his piece on Affect Recognition, Sébastien Krier notes that nonprofits such as AI Now, the Partnership on AI, and the American Civil Liberties Union (ACLU) already have expressed opposition to the widespread use of affect recognition, and researchers in the field have frequently highlighted the risk of misuse.

In China, stakeholders from civil society, including academics, think tanks, and NGOs, generally retain a degree of operational autonomy while also operating at the risk of positioning themselves in abject opposition to the state, which would label their organizations and members as national security risks.[13] Censorship thus affects the convening role of civil society organizations in China.

China's state media also plays a significant role in forming public opinion and attitudes, such as mounting support for intelligence-led policing efforts. As the Chinese government has rolled out its SkyNet and Sharp Eyes initiatives, state-sponsored media has incorporated the technology into its harmonious society narrative by highlighting its capability to assist local-level anti-corruption efforts, as well as locating missing criminals via facial recognition technologies.[14]

> **The convening role of civil society stakeholders, either in questioning or in approving the utilization of specific AI technologies, plays varying parts across the United States and China. While this does not necessarily impede any forms of cooperation, it nonetheless holds important consequences for how public opinion is shared or shaped during the processes of implementing novel AI systems.**

(2) Political Barriers

Technological decoupling has been especially pronounced in terms of erecting new political and economic barriers to cooperation. The US Bureau of Industry and Security's Entity List, which bars foreign entities from doing business with US partners, has been a pronounced tool in this regard. The US Commerce Department has used the Entity List to sanction several Chinese surveillance technology companies for being complicit in human rights violations in terms of delivering high-tech surveillance technologies that have been used to surveil ethnic minorities in China's Xinjiang Autonomous Region.[15] As supply chains across the United States and China are severed, Beijing has responded by establishing its own "Unreliable Entity List,"[16] as well as an Export Ban on Strategic Technologies, including AI,[17] which mirrors policy initiatives first implemented in the United States.[18]

Political barriers to cooperation are especially strong in relation to deepfake technologies, where the United States and China are taking very different approaches in terms of governance.

In the United States, where policymakers are concerned about the security risks presented by deepfakes, US legislation mandates reporting on the foreign weaponization of deepfakes with specific reference to the activities and capabilities of China and Russia.[19]

In China, the response to deepfakes has been driven through new regulations issued by the Cyberspace Administration of China (CAC), which requires publishers to label false audiovisual material made using new technologies, including deep learning and virtual reality. Further legislative efforts against deepfakes are embodied in the CAC's "Provisions on the Governance of the Online Information Content Ecosystem," which aims at addressing illegal and negative online information.[20] The term "negative" is, of course, arbitrary and subject to definition by China's online censors.

In both the United States and China, a strong pressure exists to prevent the spread of unlabeled deepfakes on online platforms, though the sources of the pressure differ. In China, companies are responding to top-down legislation applicable across audiovisual service providers. In the United States, platforms are mostly self-regulating and are reacting to pressure from the media, civil society, and policymakers, but without clear federal regulations to which they need to conform.

China's more closed Internet may help explain why the Chinese government has not singled out foreign-created deepfakes in the same way that the United States has. Julia Chen argues that the relatively open US Internet infrastructure could make it more vulnerable to outside influences, which likely demarcates a shift in policy to focus more on adversarial actors and aspects of online interference. The requirement for the US intelligence services to report on Chinese developments in deepfake technology is symptomatic of deep mistrust, which discourages cooperation.

The backdrop of inconsistent rules for new areas, such as deepfakes and the quest for controlling online information, has generally called for democratic governments to define a set of minimum standards for platforms to apply to avoid the exploitation of loose and inconsistent rules and legislation.[21]

While a model of self-regulation by industry continues to stay the dominant model in the United States, it is feasible that this voluntary model could lead US platforms to converge toward labeling deepfakes, just as Chinese platforms are required to do through national legislation. [22]

The vulnerabilities that are linked to a liberal model of internet governance have, in turn, caused the current US administration to embrace entirely new mechanisms, such as the proposed establishment of a "Clean Network." According to the US State Department, the Clean Network initiative is a new approach "to safeguarding the nation's assets including citizens' privacy and companies' most sensitive information from aggressive intrusions by malign actors, such as the Chinese Communist Party." [23] The executive orders targeting the Chinese-owned social media platform TikTok as well as the messaging app WeChat [24] appear to be the most prominent examples of the United States pursuing its nascent ideas of enforcing a Clean Network.

On the one hand, proposals such as the Clean Network in the United States are indicative of a new approach to governing networks and information infrastructure, which ultimately gravitates around some of the same measures that China has long deployed, including the exclusion of specific actors perceived as potential threats to national security as well as domestic cohesion.

While it is clear that the varying approaches to the control of online information, including the notion of AI-generated deepfakes, erects some political barriers to cooperation, it is also clear that there is space to learn from the different approaches that are currently being pursued by the United States and China.

(3) Economic Barriers

Economic barriers to cooperation are primarily linked to differing notions of industrial policy, market control, and economic planning. China tends to prioritize medium- and long-term industrial planning, while policies in the United States tend to provide a more laissez-faire approach to liberal market governance. The key difference between the liberal market approach of the United States and the state-capitalist approach of China is the role of the state in China in terms of subsidizing and protecting industry, as well as creating demand. These varying systemic approaches to market and industry governance also affect how AI technologies are implemented and publicly supported, which erects economic barriers to cooperation.

China's AI Development Plan, for example, aims to position China as the world's leading AI innovation center by 2030, while ensuring social stability and public security through monitoring

and early warning systems based on sensors, video analysis, identification technology, biometric identification, and police products.[25]

The development and deployment of facial recognition systems, as well as AI in healthcare, have also been aided by mutually reinforcing policies issued by China's central government, such as Made in China 2025. China's industrial policies and plans provide massive capital flows toward innovation in domestic technology, through such measures as creating new subsidies and tax incentives for technology startups. [26] Another directive, introduced in 2018, requires all government offices and public institutions to rely entirely on domestic suppliers of technology by 2021. [27] As a whole, these policies are in line with China's long-term transition toward indigenous innovation, marked by spending in domestic research and development, while aiming for self-sufficiency in key sectors of the economy.

In the United States, the Trump administration has been using a variety of tools, including unilateral tariffs, WTO dispute mechanisms, and revisions to national guidelines on foreign investment, to put pressure on the Chinese government in areas of trade and market reform. [28] While some of these measures have exposed severe risks in Chinese technology companies' supply chains, in effect cutting them off from crucial American technology, the result only seems to reinforce Beijing's resolve to secure self-sufficiency in a number of strategic industries and supply chains that are vulnerable to disruption by adversarial actors.

At the same time, the United States has embarked on a new push toward increased government intervention in its domestic economy. New legislation has been proposed that would direct billions of federal dollars to American semiconductor manufacturers, with the hope of bringing production back home on American soil. [29] Similarly, US lawmakers have moved to propose a national strategy on AI to keep the United States from ceding power to other countries in the field.[30] These new measures that are taken by the United States signal that the country could be willing to shed some of its laissez-faire orthodoxy to counter the industrial policies that have long been embraced in China.

AVENUES TOWARD COOPERATION

Our case studies across AI Principles, Deepfakes, Affect Recognition, AI in Healthcare, and Facial Recognition have revealed that several important avenues toward cooperation exist. The most important of these have been highlighted in relation to cooperation based on (1) AI Principles; (2) Research; (3) Track 2 and Track 1.5 Diplomacy; and (4) Shared Guidance. In the following section, we provide some beginning insights into the kinds of accountability mechanisms that could be the most likely to gain traction across a diverse set of stakeholders from the United States and China.

(1) AI Principles

Shared AI principles between countries represent a potential starting point for collaboration, which may reduce global security threats and advance the adoption of concrete accountability mechanisms for responsible AI. Consensus has especially been growing around key thematic trends, including privacy, accountability, safety and security, transparency and explainability (i.e., that the results and decisions made by an AI system can be understood by humans), fairness and nondiscrimination, human control of technology, professional responsibility, and promotion of human values. [31]

Jessica Cussins Newman has argued that the emergence of AI principles from stakeholders in both the United States and China provides reasonable points of leverage for collaboration if there is sufficient political will from both countries to act upon them. As she notes, existing differences do not necessarily impede cooperation, as collaboration is possible without agreement on all principles, values, and matters of technological application.[32]

At a high level, both the United States and China commonly refer to the need for AI safety and security, fairness and justice, privacy, transparency and explainability, accountability, diversity, inclusion, access, education and training, and the reduction of discrimination and bias.

AI safety and security, in particular, have been highlighted for their potential role in AI-related diplomacy between the United States and China. For example, the US National Security Commission on Artificial Intelligence indicated that it is pursuing possible avenues for AI-related diplomatic discussions with China in the areas of AI safety and AI's implications for strategic stability. [33]

Both the United States and China are part of the G20, which has adopted a set of human-centered AI Principles that draw from the OECD AI Principles. The G20 could therefore be a constructive forum to build on the ideas of "human-centric AI," which calls on countries to use AI in a way that respects human rights and shares the benefits brought by it. This includes the building of

trust related to data flows and toward privacy and security based on domestic and international laws and agreements. [34]

(2) Research

In terms of research, concrete recommendations include pursuing joint research projects and enabling research transparency. **Research collaborations between AI experts in both countries are common, and there are many examples of collaboration across the United States and China.**

One concrete project, highlighted by Sébastien Krier, involves researchers from Harbin Engineering University and UCLA, which has resulted in the construction of multimodal datasets for affect recognition, in support of the international research community. [35] There have also been examples of researchers from Chinese and US institutions collaborating on efforts to support the detection of deepfakes. Researchers from the University of Chinese Academy of Sciences and the State University of New York, for example, have worked together to produce a deepfake video dataset improving on preexisting datasets suffering from low visual quality and lack of resemblance to deepfakes circulating on the Internet.[36] Research in healthcare has also been highlighted by Marshall Peihang Li as a potential area for cooperation, especially in relation to the drug discovery process.

Furthermore, a large proportion of Chinese AI researchers already study, work, and live in the United States.[37] While some of these researchers may travel back to China, benefitting the Chinese AI ecosystem, several studies have shown that the core to US competitiveness in AI rests on international talent—a recent study by CSET has found that more than half of the AI workforce in the United States was born abroad, while some two-thirds of current graduate students in AI-related fields come from abroad.[38]

Harnessing AI talent in both countries while fostering academic cooperation focused on areas such as authentication or deepfake detection techniques could have positive-sum benefits for both the United States and China as well as for the broader international community. [39]

Even if security concerns mean that some of the most advanced AI detection models are developed separately in the United States and China and are supported and deployed separately by respective defense communities, there still exists a strong potential for collaboration between researchers on lower-stakes models to help companies and individuals identify the most common types of deepfakes, for example. In addition, the research community could cooperate on common standards for reporting progress in generative models, which would help researchers and developers keep track of the latest developments.

Generally speaking, AI research and development are based on a strong culture rooted in open-sourced development and research, which implies that the latest knowledge in many fields of AI, by and large, is open and accessible. Through embracing international research collaboration, a culture of trust and mutual partnership would be likely to benefit technological innovation in both the United States and China, while avoiding the pitfalls of engaging in a zero-sum AI race.

(3) Track 2 and Track 1.5 Diplomacy

Several authors of the series have highlighted Track 2 and Track 1.5 diplomacy as viable avenues to cooperation by assisting officials in managing and resolving conflicts while exploring possible solutions through public but unofficial channels. **Efforts at establishing Track 2 and Track 1.5 diplomacy have already been successfully implemented in areas such as AI safety and security**.[40]

Building shared glossaries of cyber terminologies, including terms relating to information operations and deepfakes, could also provide a foundation for bilateral dialogue.[41] Track 2 and Track 1.5 dialogues could, for example, serve to reach greater common recognition of different kinds of synthetic media and the risks associated with them. The US-Chinese Glossary of Nuclear Security Terms provides a possible model of cooperation in this regard.[42] Such projects could contribute to building trust and mutual understanding, as well as laying the foundations for future diplomatic engagement. Subsequent cooperation would likely need to address the difficult challenge of identifying and punishing perpetrators, for example, in the case of transnational deployment of deepfakes. This is an issue common to cyber offenses more broadly. No country is immune from the threat of deepfakes, collaboration in these areas could help accelerate technical and governance solutions and break away from the zero-sum framing of AI development. Track 2 and Track 1.5 engagements are recommended as pragmatic channels for communication and dispute resolution that should be maintained and thought of as viable avenues toward increased cooperation.

(4) Shared Guidance

The United States and China explicitly support international collaboration in their national AI strategies. However, in some cases, the two countries have opted to prioritize different international forums. China has played a prominent role in the development of international AI standards, primarily through the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) [43] as well as the UN's International Telecommunications Union (ITU). Meanwhile, the United States has focused on the development of principles and governance frameworks through bodies including the G7 and the OECD. These efforts have helped establish widespread norms for AI development but may fall short of influencing standards or regulations.

Stakeholders from the United States and China could engage in new avenues to provide shared guidance, which could have an impact on harmonizing certain standards and regulation methods. In terms of facial recognition technologies, Samuel Curtis has suggested that US and Chinese think tanks could provide guidance by developing a shared taxonomy of facial recognition use cases, and evaluating each use case with respect to some agreed-upon legal framework, such as the UN Declaration of Human Rights,[44] for instance. He also proposes that stakeholders could collaborate in drafting a document akin to the Electronic Frontier Foundation's "International Principles on the Application of Human Rights to Communications Surveillance," which has been praised by the UN Office of the High Commissioner for Human Rights (UNHCR) that focuses explicitly on the necessary and proportionate applications of biometric surveillance technologies. In terms of deepfakes, Julia Chen suggests that reporting on the progress in generative models could be a way of supporting better capabilities-tracking of deepfake technologies.[45] If this were to be done by a diverse group of Chinese and US researchers, potentially convened by IEEE or organizations such as the Association for the Advancement of Artificial Intelligence and the Beijing Academy of AI, this kind of guidance could help increase adoption and build trust between actors from different countries. In terms of affect recognition, Sébastien Krier has argued that more regulation, or "opt-in" requirements, as well as the potential use of model cards,[46] could be encouraged. The use of model cards could be a way to provide guidance about what a machine-learning system can and cannot do, as well as the types of errors to which a system might be prone. Engaging in new methods for providing shared guidance, would not only serve to build trust across the United States and China, but would also contribute to strengthen the foundations on which international standards and cyberspace-regulations are conceived. New forms of shared guidance could have a positive impact on creating more transparent and inclusive outcomes associated with a range of AI technologies.

Concluding Remarks – Where Are We Headed?

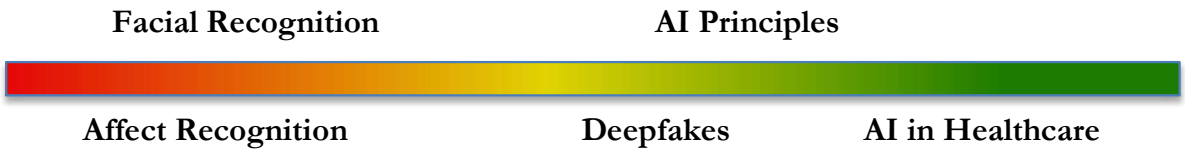
"*Exploring AI Issues Across the United States & China*" has highlighted that many areas of AI development, adoption, and governance rest on a continuum of apparent ease or difficulty in terms of cooperation. We have used the Green-Yellow-Red Light Framework as an indicator for this continuum and have found that most barriers to cooperation relate to cultural, political, and economic issues. These barriers mainly refer to structural imbalances, both in the United States and China domestically, as well as across the two countries.

Structural imbalances generally refer to long-run effects induced by AI technologies, such as labor displacement and shifts in the balance of power and autonomy, but they also refer to more immediate threats, such as deepfake technologies, for example. While it is clear that the United States and China respond differently to many AI-induced structural imbalances, it is also clear that

disparate forms of AI governance are emerging in the process. This is seen in the divergent application of facial recognition technologies across the two countries.

As the range of cyber-physical threats intensifies, questions of AI security have already become a matter of national security, including the need to protect critical information infrastructure. Without a stronger commitment to international coordination and responsibility, questions of AI security could cause further fragmentation of the international cyber regime. That is, without global coordination and joint interventions, the increasing demand for "digital sovereignty" could turn into a new kind of technological nationalism, which reinforces a low-trust environment.[47]

Due to these considerations, it is imperative that bilateral dialogue between the United States and China is strengthened. We find several tangible avenues to strengthen cooperation, including continued dialogue on AI principles, research-based collaboration, enhancing Track 2 and Track 1.5 exchanges, as well as providing shared guidance contributing to harmonizing technical standards and possibly some aspects of different regulations.



AI principles provide a starting point that can be and are already being leveraged to support international AI governance efforts. A guiding principle for all stakeholders should rest on the notion that not all societies are based on, and governed by, the same cultural values and set of principles. Fairness, privacy, and bias in one domain or society could have altogether different meanings in another context.[48]

The current lack of regional diversity in the formulation of AI principles is reflected in the concentration of AI research, where 86% of papers published at AI conferences in 2018 were attributed to authors in North America, Europe, and East Asia.[49] These regions also dominate the current discussions on AI ethics and technological development, more broadly,[50] while it is crucial that the United States and China refrain from using their relative dominance to eclipse the views of others. Instead, an embrace of multilateralism would allow for marginalized groups to be heard and included, which generally is a sought-after value when devising new and more equitable forms of international AI governance.

Although both the United States and China have endorsed the G20 AI Principles, challenges remain in reaching shared understandings and opportunities for collaboration. Nonetheless, this

endorsement provides a critical starting point for engaging in important dialogues and value-based exchanges.

Even though some variation in principles and priorities exists, the United States should be wary of engaging in international AI governance forums that exclude China, which could exacerbate differences and enhance tensions and instabilities. The recently launched Global Partnership on AI (GPAI), which was created by France and Canada along with the United States, the EU, and a host of other countries, most notably excludes China. While there can be benefits to fostering and consolidating efforts aimed at preserving a liberal democratic vision of technology development and use, this should not happen at the expense of collaboration. Although GPAI does not include China as a main member, the organization should still strive to encourage active dialogue and joint forms of collaboration.

Significant common ground and a shared interest in reducing certain risks of AI, across use cases such as deepfakes, affect recognition, or facial recognition technologies, provide a basis for engaging in joint efforts. These could include reporting standards for machine-learning models or the potential for creating shared glossaries covering cyber terminologies. Such subtle approaches to stakeholder engagement across the United States and China could contribute to building a greater sense of common and shared understandings while laying the foundation for deeper engagement and building more trust in the long run.

It is our sincere hope that this series may have contributed to an outline of some of the most pressing barriers to cooperation across the United States and China, as well as some of the potential opportunities to renew cooperation on a range of important fields.

ENDNOTES

[1] Stephen Cave & Seán S. ÓhÉigeartaigh, "An AI Race for Strategic Advantage: Rhetoric and Risks," AAAI/ACM Conference on AI, Ethics, and Society, February 5, 2018, https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_163.pdf.

[2] Tim Hwang, "Deepfakes: A Grounded Threat Assessment," Center for Security and Emerging Technology, July 2020, https://cset.georgetown.edu/research/deepfakes-a-grounded-threat-assessment/.

[3] Aleš Završnik, "Criminal Justice, Artificial Intelligence Systems, and Human Rights," ERA Forum, *Journal of the Academy of European Law*, February 20, 2020, https://link.springer.com/article/10.1007/s12027-020-00602-0

[4] Joy Buolamwini & Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," Proceedings of Machine Learning Research, Conference on Fairness, Accountability, and Transparency, 2018, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

[5] Marcus Comiter, "Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It," Belfer Center for Science and International Affairs, Harvard Kennedy School, August 2019, https://www.belfercenter.org/publication/AttackingAI.

[6] For an example of the use of deepfake audio for fraud, see Catherine Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case," *Wall Street Journal*, August 30, 2019, https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

[7] Yong Suk Lee, Benjamin Cedric Larsen, Michael Webb, Mariano Florentino Cuellar "How Would AI Regulation Change Firms' Behavior: Evidence from Thousands of Managers" (2019) SIEPR Working Paper 19-031 https://siepr.stanford.edu/research/publications/how-would-ai-regulation-change-firms-behavior-evidence-thousands-managers.

[8] Thomas Will, "Trump Clamps Down on Chinese Graduates and Researchers," American Institute of Physics, June 2, 2020, https://www.aip.org/fyi/2020/trump-clamps-down-chinese-grad-students-and-researchers.

[9] Allan Dupont, "New Cold War: De-Risking US-China Conflict" June 24, 2020. Hinrich Foundation, https://www.hinrichfoundation.com/research/wp/us-china/new-cold-war/.

[10] Paul Scharre, "A Million Mistakes a Second," *Foreign Policy*, September 12, 2018, https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/.

[11] Danit Gal, "Perspectives and Approaches in AI Ethics: East Asia," *The Oxford Handbook of Ethics of AI*. New York: Oxford University Press, 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3400816#.

[12] Mia Hunt, "US Abandons Boycott of Global AI Partnership," Global Government Forum, May 31, 2020, https://www.globalgovernmentforum.com/us-abandons-boycott-of-global-ai-partnership/.

[13] Emina Popović, "Lobbying Practices of Citizens' Groups in China," *SAGE Open*, n.d., 9.

[14] "中国纪检监察报 - '花式'逃亡终难逃法网," http://www.jjjcb.cn/content/2018-09/30/content_68514.htm; "'雪亮工程'保平安--法治--人民网," November 18, 2018, https://web.archive.org/web/20181118195610/http://legal.people.com.cn/n1/2017/0614/c42510-29337488.html.

[15] "Commerce Department to Add Nine Chinese Entities Related to Human Rights Abuses in the Xinjiang Uighur Autonomous Region to the Entity List," US Department of Commerce, https://www.commerce.gov/news/press-releases/2020/05/commerce-department-add-nine-chinese-entities-related-human-rights.

[16] Ministry of Commerce, People's Republic of China. "Provisions on the Unreliable Entity List," http://english.mofcom.gov.cn/article/policyrelease/questions/202009/20200903002580.shtml.

[17] Coco Liu, "China's New AI Export Curbs, Threaten TikTok's US Sale," Nikkei, August 29, 2020, https://asia.nikkei.com/Politics/International-relations/US-China-tensions/China-s-new-AI-export-curbs-threaten-TikTok-s-US-sale.

[18] James Vincent, "US Announces AI Software Export Restrictions," The Verge, January 5, 2020, https://www.theverge.com/2020/1/5/21050508/us-export-ban-ai-software-china-geospatial-analysis.

[19] Jason Chipmon & Matthew Ferraro, "First Federal Legislation on Deepfakes Signed into Law," JD Supra, December 24, 2019, https://www.jdsupra.com/legalnews/first-federal-legislation-on-deepfakes-42346/.

[20] "Provisions on the Governance of the Online Information Content Ecosystem," China Law Translate, December 21, 2019, https://www.chinalawtranslate.com/en/provisions-on-the-governance-of-the-online-information-content-ecosystem/.

[21] Hany Farid & Hans-Jakob Schindler, "Deep Fakes: On the Threat of Deep Fakes to Democracy and Society," Konrad-Adenauer-Stiftung and the Counter Extremism Project, June 2020, https://www.counterextremism.com/sites/default/files/CEP-KAS_Deep%20Fakes_062920.pdf.

[22] "Provisions on the Management of Online A/V Information Services," China Law Translate, November 29, 2019, https://www.chinalawtranslate.com/en/provisions-on-the-management-of-online-a-v-information-services/.

[23] US Department of State, "The Clean Network," 2020, https://www.state.gov/the-clean-network/

[24] White House, August 6, 2020, https://www.whitehouse.gov/presidential-actions/executive-order-addressing-threat-posed-tiktok/.

[25] "China's New Generation of Artificial Intelligence Development Plan," Foundation for Law & International Affairs, July 30, 2017, https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan/.

[26] Tristan Kenderdine, "China's Industrial Policy, Strategic Emerging Industries, and Space Law." *Asia & The Pacific Policy Studies*, May 12, 2017, https://onlinelibrary.wiley.com/doi/epdf/10.1002/app5.177.

[27] Yoko Kubota in Beijing & Liza Lin in Singapore, "Beijing Orders Agencies to Swap Out Foreign Tech for Chinese Gear," *Wall Street Journal*, December 9, 2019, sec. World, https://www.wsj.com/articles/beijing-orders-agencies-to-swap-out-foreign-tech-for-chinese-gear-11575921277.

[28] Yelling Tan, "Will US Trade Pressure Actually Change China's Industrial Policy," *Washington Post*, December 20, 2018, https://www.washingtonpost.com/news/monkey-cage/wp/2018/12/20/will-u-s-trade-pressure-actually-change-chinas-industrial-policy/.

[29] Ana Swanson & Duck Clark, "Lawmakers Push to Invest Billions in Semiconductor Industry to Counter China," *New York Times*, June 11, 2020, https://www.nytimes.com/2020/06/11/business/economy/semiconductors-chips-congress-china.html.

[30] James Rundle, "Lawmakers Issue Call for National AI Strategy," *Wall Street Journal*, September 16, 2020, https://www.wsj.com/articles/lawmakers-issue-call-for-national-ai-strategy-11600250400.

[31] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, & Madhulika Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," Berkman Klein Center Research Publication No. 2020-1, January 15, 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482.

[32] Seán S. ÓhÉigeartaigh et al., "Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance," *Philosophy & Technology*, May 15, 2020, https://link.springer.com/content/pdf/10.1007/s13347-020-00402-x.pdf.

[33] "Interim Report," National Security Commission on Artificial Intelligence, November 2019, https://drive.google.com/file/d/153OrxnuGEjsUvlxWsFYauslwNeCEkvUb/view.

[34] "G20 Ministerial Statement on Trade and Digital Economy," June 9, 2019, https://www.mofa.go.jp/files/000486596.pdf.

[35] Jing Chen et al. "HEU Emotion: A Large-Scale Database for Multi-Modal Emotion Recognition in the Wild." ArXiv:2007.12519 [Cs], July 2020. arXiv.org, http://arxiv.org/abs/2007.12519.

[36] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, & Siwei Lyu, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," March 2020, https://arxiv.org/pdf/1909.12962.pdf.

[37] "The Global AI Talent Tracker," Macro Polo, https://macropolo.org/digital-projects/the-global-ai-talent-tracker/.

[38] Remco Zwetsloot, James Dunham, Zachary Arnold, & Tina Huang, "Keeping Top AI Talent in the United States," CSET. December 2019, https://cset.georgetown.edu/wp-content/uploads/Keeping-Top-AI-Talent-in-the-United-States.pdf.

[39]For discussion of zero-sum and positive-sum framings in relation to US-China research collaboration, see Jenny J. Lee & John P. Haupt, "Winners and Losers in US-China Scientific Research Collaborations," July 6, 2020, https://link.springer.com/article/10.1007/s10734-019-00464-7.

[40] Andrew Imbrie & Elsa B. Kania, "AI Safety, Security, and Stability Among Great Powers," CSET Policy Brief, Center for Security and Emerging Technology, December 2019, https://cset.georgetown.edu/wp-content/uploads/AI-Safety-Security-and-Stability-Among-the-Great-Powers.pdf.

[41] The suggestion of a common glossary for cyber terminology is made in Ariel (Eli) Levite & Lyu Jinghua, "Chinese-American Relations in Cyberspace: Toward Collaboration or Confrontation?" Carnegie Endowment for International Peace, January 24, 2019, https://carnegieendowment.org/2019/01/24/chinese-american-relations-in-cyberspace-toward-collaboration-or-confrontation-pub-78213.

[42] National Research Council, "English-Chinese, Chinese-English Nuclear Security Glossary," 2008, https://doi.org/10.17226/12186.

[43] Jeffrey Ding, Paul Triolo, & Samm Sacks, "Chinese Interests Take a Big Seat at the AI Governance Table," New America, June 20, 2018. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/chinese-interests-take-big-seat-ai-governance-table/.

[44] "Universal Declaration of Human Rights," October 6, 2015, https://www.un.org/en/universal-declaration-human-rights/.

[45] Hwang, "Deepfakes."

[46] Margaret Mitchell et al. "Model Cards for Model Reporting," October 2018, https://arxiv.org/abs/1810.03993.

[47] Thorsten Jellenik, Wendell Wallach, & Danil Kerimi, "Coordinating Committee for the Governance of Artificial Intelligence," G20 Insights, June 25, 2020, https://www.g20-insights.org/policy_briefs/coordinating-committee-for-the-governance-of-artificial-intelligence/.

[48] Alexa Hagerty & Igor Rubinov, "Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence," 2019, https://arxiv.org/abs/1907.07892.

[49] HAI AI Index 2019 Annual Report, 2019, https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf.

[50] Abishek Gupta & Victoria Heath, "AI Ethics Groups Are Repeating One of Society's Classic Mistakes." *MIT Technology Review*, September 14, 2020, https://www.technologyreview.com/2020/09/14/1008323/ai-ethics-representation-artificial-intelligence-opinion/.

Series Edited and Overseen by Heather Evans and Benjamin Cedric Larsen
Cover Artwork Designed by Kisara Moore and Anya Lassila
Traffic Light Design by Anya Lassila

Thank you to the Asia Society Northern California Team for supporting this publication.