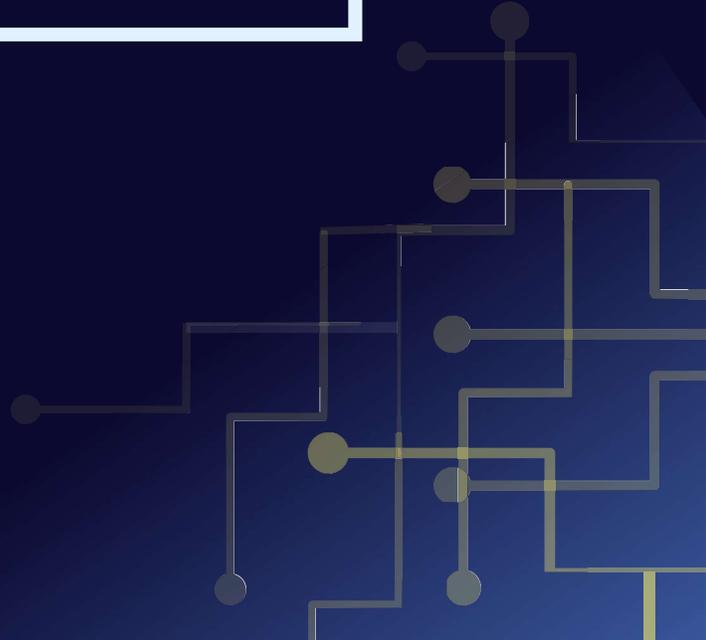


**AI PRINCIPLES
IN CONTEXT**

**JESSICA CUSSINS
NEWMAN**

8 · 20 · 2020



AI PRINCIPLES IN CONTEXT:

Tensions and Opportunities for the United States and China

INTRODUCTION

Over the past five years, awareness of the transformative impacts of artificial intelligence (AI) on societies, policies, and economies around the world has been increasing. Unique challenges posed by AI include the rise of synthetic media (text, audio, images, and video content);¹ insufficient transparency in the automation of decision making in complex environments;² the amplification

GAINING A BETTER UNDERSTANDING OF THE CULTURAL AND POLITICAL CONTEXT SURROUNDING KEY AI PRINCIPLES CAN HELP CLARIFY WHICH PRINCIPLES PROVIDE PRODUCTIVE OPPORTUNITIES FOR POTENTIAL COOPERATION.

of social biases from training data and design choices, causing disparate outcomes for different races, genders, and other groups;³ and the vulnerability of AI models to cyberattacks such as data poisoning.⁴ Awareness of these and numerous other challenges has led to serious deliberation about the principles needed to guide AI development and use. By some counts, there

are now more than 160 AI principles and guidelines globally.⁵

Because these principles have emerged from different stakeholders with unique histories and motivations, their style and substance vary. Some documents include recommendations or voluntary commitments, while a smaller subset includes binding agreements or enforcement mechanisms. Despite the differences, arguably a consensus has also been growing around key thematic trends, including privacy, accountability, safety and security, transparency and explainability (meaning 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.⁶ However, in the face of complex and growing geopolitical tensions, it is important to investigate the feasibility of any such “consensus” between the United States and China.

The proliferation and partial consolidation of AI principles globally are an important phenomenon that has helped define aspirational technological and political trajectories, but it is only the beginning of a much longer road of AI governance. High-level principles on their own remain vague and hard to implement. The benefits to this flexibility include that principles may be more resilient to changes in the future. However, such principles risk remaining superficial or, more insidiously, providing a cover of responsibility without offering real accountability. To prevent this outcome, numerous efforts are underway to operationalize AI principles via mechanisms

ranging from technical tools to oversight boards and committees, frameworks and best practices, standards, certifications, and regulations.⁷

Gaining a better understanding of the cultural and political context surrounding key AI principles can help clarify which principles provide productive opportunities for potential cooperation between the United States and China and provide insight into the kinds of accountability mechanisms most likely to gain traction from diverse stakeholders.

INTERNATIONAL CONTEXT

Although the majority of AI strategies and principles stem from particular countries, companies, or organizations, several noteworthy international and intergovernmental efforts have also appeared in this space. Because the United States and China are leaders of many such initiatives, the international context is a relevant part of understanding the goals of each country. For example, both 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).⁸ International standard setting is an important element of AI governance that can enable certain technology companies and countries to have an outsized impact on the global stage, for example, by improving the interoperability of Chinese AI systems. Meanwhile, the United States has focused on the development of principles and governance frameworks through bodies including the G7 and the Organization for Economic Cooperation and Development (OECD). These efforts have helped establish widespread norms for AI development but may fall short of influencing standards or regulation.

For example, in June 2018 the G7 (consisting of Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States) committed to the Charlevoix Common Vision for the Future of Artificial Intelligence.⁹ This vision includes 12 commitments that promote human-centric AI that fosters economic growth, societal trust, and equality and inclusion. More recently, discussions originating at the G7 led to the development of the Global Partnership on Artificial Intelligence (GPAI), which officially launched June 2020 with member countries including the G7 as well as Australia, India, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, and the European Union.¹⁰ GPAI aims to guide the responsible development and use of AI, grounded in human rights, inclusion, diversity, innovation, and economic growth. The political significance of this is discussed further in the Analysis section.

The OECD AI Principles, adopted in May 2019 by 42 countries, are of particular importance in understanding the international context of AI governance because they highlight a growing intergovernmental consensus on a set of principles.¹¹

The five values-based principles follow:

- **Inclusive growth, sustainable development, and well-being**
- **Human-centered values and fairness**
- **Transparency and explainability**
- **Robustness, security, and safety**
- **Accountability**

One month after their launch, the OECD AI Principles were endorsed by the G20, which notably includes China. That endorsement further expanded the reach of the principles to much of the world.¹² The practical value of this growing consensus is not yet clear, but the OECD launched the AI Policy Observatory in early 2020 to provide a platform to facilitate efforts by governments and stakeholders around the world to implement these principles. The Observatory is collecting examples of AI policies and tools and helping make them more accessible globally.

THE UNITED STATES

The landscape of AI principles development in the United States has involved numerous and varied efforts from across government, civil society, and the private sector, resulting in a relatively diverse selection of AI principles overall. There have also been examples of multi-stakeholder collaboration since the early days of AI governance. For example, a U.S. government report from October 2016, “Preparing for the Future of Artificial Intelligence,” was informed by multiple public workshops co-hosted with universities and other organizations.¹³ This report highlighted several themes that have come to be standard in articulations of AI principles: justice, fairness, accountability, safety and control, security, and international cooperation. The subsequent launch of the American AI Initiative in February 2019 implemented a whole-of-government strategy, including explicit collaboration and engagement with civil society and the private sector.¹⁴ Though designed by a new President’s administration and highlighting different priorities, the initiative similarly calls for safe and trustworthy AI development created with “American values” including freedom, human rights, and privacy.

Another important example of AI principles led by the U.S. government was the January 2020 White House Guidance for Regulation of Artificial Intelligence Applications, which defines 10 principles for the stewardship of AI applications.¹⁵ The first two principles are “public trust” and “public participation.” Also of note is the National AI R&D Strategic Plan, initially published in

2016 and updated in 2019.¹⁶ The updated version reiterates commitments to addressing the ethical, legal, and societal implications of AI as well as ensuring the safety and security of AI systems, among other priorities. It also adds an additional strategic priority to expand public-private partnerships to accelerate AI advances, again highlighting the role of multi-stakeholder initiatives in the United States.

Lastly, the U.S. Department of Defense (DoD) adopted ethical principles for AI in February 2020, which state that DoD's use of AI must be responsible, equitable, traceable, reliable, and governable.¹⁷ These principles stemmed from recommendations made by the Defense Innovation Board following more than a year of consultation with AI experts in commercial industry, government, and academia, as well as the American public.

Prominent AI principles in the United States have also emerged from civil society. The Asilomar AI Principles are an early and influential set of AI principles that was led by the nonprofit organization the Future of Life Institute and developed through a consultative multi-stakeholder process that concluded in January 2017. These principles are especially notable because they have been endorsed by thousands of AI and robotics researchers and explicitly address research issues as well as ethics and values and longer-term issues. The California State Legislature also formally supported the Asilomar AI Principles in August 2018, providing a symbolic endorsement of safe and beneficial AI development, which was subsequently referenced by the state oversight agency in its AI roadmap for California.¹⁸

Finally, many examples of high-profile AI principles have come from the private sector in the United States. Google published a set of principles in June 2018 that includes issues such as avoiding creating or reinforcing bias and incorporating privacy into design.¹⁹ Google's principles also name several AI applications the company will not pursue, including weapons and surveillance technologies that violate international norms, laws, or human rights.

In May 2018, IBM published a set of principles to guide data responsibility, which includes ensuring transparency and explainability of AI systems.²⁰ The company has also identified five areas of ethical focus for AI.²¹ Another noteworthy private sector effort comes from Microsoft, which published a set of AI principles and also established an Office of Responsible AI as well as the AI, Ethics, and Effects in Engineering and Research (Aether) Committee to ensure the principles inform company behavior from teams' research decisions to corporate policies.²²

AI stakeholders in the United States have been called out for not adhering to their principles. For example, Microsoft was criticized for supporting a facial recognition company in Israel, AnyVision, which was reportedly used to identify and track Palestinians in the West Bank.²³ And Google was criticized for developing a censored search engine that internet users in China could

access.²⁴ Companies have also been driven to action in efforts to stand up for their principles. For example, in response to nationwide protests against police brutality and racial profiling, Microsoft, Amazon, and IBM all made announcements and pledges to stop selling facial recognition technology to police departments, either for a designated period of time or until a federal regulation is in place to help protect people's privacy and rights.²⁵

CHINA

The idea that China is not interested in AI ethics or principles is a largely unwarranted misconception.²⁶ In fact, the exploration of AI principles has been robust for several years. However, the development of AI principles in China has generally been characterized by greater government involvement, which has led to a relatively more consistent overarching policy. The State Council of China released the first major AI national strategy, called the New Generation Artificial Intelligence Development Plan, in July 2017. The Development Plan provides a comprehensive strategy that includes specific benchmarks for AI industries including the intention to become the “world's premier artificial intelligence innovation center” by 2030.²⁷

The Development Plan highlights many goals for AI including calls for risk prevention and threat mitigation to ensure safe, reliable, and controllable AI development and adherence to “the principles of security, availability, interoperability, and traceability.” A further emphasis of the Development Plan is ensuring social stability and public security, which includes monitoring and early warning systems such as sensor technologies, video analysis and identification technology, biometric identification technology, and police products.

The Ministry of Science and Technology (MOST) helped lead the Development Plan and is also one of the bodies designated to help carry out its implementation. MOST established a New Generation AI Governance Expert Committee in March 2019 as one component of this work, and the committee subsequently released a set of governance principles for responsible AI.²⁸ These principles include harmony and friendliness, fairness and justice, inclusivity and sharing, privacy, security/safety and control, shared responsibility, open collaboration, and agile governance. Discussions of these principles specify further criteria of responsibility including “respecting human rights, avoid[ing] misuse, and prohibit[ing] abuse and malicious application.”²⁹ The principles also call for supporting environmental conservation, avoiding data and platform monopolies, anticipating potential future risks from increasingly advanced AI, and supporting international dialogue and cooperation with the goal of developing “broad consensus on an international AI governance framework, standards, and norms.”³⁰

Another notable effort is the Beijing AI Principles, released in May 2019 by the Beijing Academy of Artificial Intelligence (BAAI) and a coalition including Peking University; Tsinghua University; the Institute of Automation and the Institute of Computing Technology in the Chinese Academy

of Sciences; and an AI industrial league including the firms Baidu, Alibaba, and Tencent.³¹ The Beijing AI Principles cover AI R&D including taking responsibility for broader impacts and controlling risks, while benefiting “all mankind and the environment.” Descriptions of the principles highlight lofty terms including privacy, dignity, freedom, autonomy, and rights. The principles also provide guidance for the use and governance of AI including avoiding a malicious AI race, continuous research, and the monitoring of risks.³²

AI principles have also been released by the private sector in China. For example, in the face of increasing international scrutiny about its role in aiding repression and surveillance,³³ Beijing-based facial recognition company Megvii Technology released Core Principles of its AI Practice in 2019, which include commitments to not weaponize its technology; to prevent discrimination; and to ensure human oversight, robustness, accountability, and data privacy.³⁴

Chinese technology company Tencent released AI principles in 2018, which state that AI should be available, reliable, comprehensible, and controllable (referred to as ARCC).³⁵ Tencent’s ethical framework further clarifies that “ethics is just the beginning” and that trust in AI will require a spectrum of rules including regulation and international law. In 2017, Tencent’s Research Institute also published a book that includes a chapter discussing AI principles and the need for better regulations.³⁶ Nonetheless, a 2020 Tencent AI White Paper highlights the benefits of controversial AI deepfake technology and calls on regulators to avoid being overly restrictive.³⁷

ANALYSIS



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. There are also tensions and divergences, both in the principles themselves and in the broader political and institutional environment in which they are employed. Overall, the issue of AI principles represents a “Yellow Light” issue in the Green-Yellow-Red (GYR) Framework put forward by the Asia Society,³⁸ meaning that both opportunities and substantive challenges stand in the way of greater collaboration between the United States and China when it comes to artificial intelligence.

Differences

Several notable and unique features of AI principles have emerged from China. At a high level, the focus is more on concepts of social stability and harmony. To some extent, this is reflective of standard rhetoric from the Chinese government, but it also takes a specific form in relation to AI. For example, there is frequent discussion of the idea of harmony between humans and machines. Prominent Chinese AI scholars have written about how AI principles should move from a focus on “human-centered design” to “harmonious design,” which is not merely about serving human needs but also considers the coexistence of humans and AI as cognitive living systems.³⁹

In a similar vein, principles of sustainability, moderation, and protection of the environment are also more commonly cited in Chinese AI principles. Sustainable development has been linked to the value of harmony in China, and some scholars have argued that it is central to traditional Chinese philosophies such as Confucianism and Taoism.⁴⁰ Other principles that emerge more frequently are related to shared benefits, equal development, and the avoidance of monopolies. China has experienced monopolies in numerous sectors but has also fought against them.⁴¹ For example, the Anti-Monopoly Law, which went into effect in 2008, seeks to restrain monopolies and to “promote the healthy development of the socialist market economy.”⁴² Moreover, according to an engineering ethics textbook taught across the country, unique Chinese characteristics stand apart from Western engineering ethical guidelines, including “responsibility precedes freedom, obligation precedes rights, the group precedes the individual, and harmony precedes conflict.”⁴³

Another difference in Chinese-led AI principles is that there is greater attention to long-term considerations and planning. China has tended to prioritize medium- and long-term industrial planning, and its approach to AI is no different. While some AI principles from the United States (such as the Asilomar AI Principles) encourage thoughtful foresight about potential changes in the future, most U.S.-led AI principles are more focused on the near term.

Lastly, there is somewhat more attention in China to the idea of “controllability” of AI systems. For example, China’s New Generation of AI Development Plan discusses the need for “reliable and controllable development of artificial intelligence,” as do the Beijing AI Principles. U.S.-led AI principles have tended to focus instead on the idea of human intervention and the ability to challenge AI-based outcomes. Such contestability can be understood as a means of control, but it may also come into tension with the ideal of human-machine harmonization alluded to in many Chinese-based AI principles.

U.S.-led AI principles have tended to focus on democratic values, including diversity and human rights. This was emphasized in the recent decision by the United States to support the GPAI, which is grounded in democratic values, human rights, inclusion and diversity, and support for innovation and economic growth. The insistence on these values for AI can in part be understood as a direct challenge to China's role in the global AI ecosystem. The White House's chief technology officer, Michael Kratsios, has argued that GPAI will be a "check on China's approach to AI," to counter "China's record of "twisting technology" in ways that threaten civil liberties."⁴⁴

To take one example, AI has become entwined in the accounts of human rights violations against the Uighurs, a Muslim minority in China's Xinjiang province. AI-enabled surveillance technologies including facial recognition technology have reportedly been used for racial profiling and "minority identification."⁴⁵ Law enforcement in the United States has also come under fire for potential abuses of facial recognition technologies,⁴⁶ but the U.S. government is generally trying to distinguish itself internationally for upholding more stringent human rights principles in its deployment of AI systems.

There is also a distinction in the actions Chinese technology companies can take to implement the principle of privacy in AI applications. Widespread data collection programs such as social credit scoring raise profound privacy concerns; despite privacy regulations like the Cybersecurity Law and Personal Information Security Specification, these frameworks have limits. As recently pointed out by technology policy scholar Danit Gal, "While privacy laws in China protect citizens' data from abuse by non-governmental actors, they do not limit the government's access to and use of private data."⁴⁷

Similarities

Despite the differences discussed, numerous similarities do exist between prominent AI principles articulated by the United States and China. At a high level, both countries commonly refer to the need for safety and security, fairness and justice, privacy, transparency and explainability, accountability, diversity, inclusion, access, education and training, and the reduction of discrimination and bias. The interpretation and implementation of these principles are hardly uniform, but the commonalities are notable nonetheless.

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 U.S. 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.⁴⁸ Concrete recommendations from AI scholars include pursuing joint projects and

enabling research transparency, developing common testing and evaluation methodologies, and facilitating Track 2 and Track 1.5 dialogues (diplomacy activities between a combination of state and non-state actors) on AI safety and security.⁴⁹

It is hardly surprising that AI researchers in particular would share many views about how to develop and use AI technologies. Research collaborations between AI experts in both countries are common and more than half of Chinese AI researchers study, work, and live in the United States.⁵⁰ There is also practical value from shared AI principles and standards given the interconnected global supply chain for AI components.⁵¹ Nonetheless, formal cooperation is complicated. Just one indication of this comes from the multi-stakeholder AI consortium the Partnership on AI (PAI). In 2018, PAI welcomed its first Chinese member, internet search company Baidu. At the time, PAI's executive director, Terah Lyons, explained, "We cannot have a comprehensive and global conversation on AI development unless China has a seat at the table."⁵² All PAI members are asked to uphold a set of shared principles.⁵³ However, in June 2020, Baidu suddenly left the partnership. The company cited the cost of membership and said it continues to "share the vision" and remains "committed to promoting the ethical development of AI technologies"; however, many have questioned if worsening U.S.-China relations are behind the change in affiliation.⁵⁴

CONCLUSION

As discussed throughout the paper, there are meaningful divergences in AI governance approaches between China and the United States. In particular, the translation of AI principles into practices surfaces complications as technical decisions collide with economic, social, and political realities. Many of these divergences are indeed deliberate, as national governments strive

NO COUNTRY WILL BENEFIT FROM DANGEROUS DEPLOYMENT OF AI TECHNOLOGIES THAT INCREASE THE LIKELIHOOD OF CATASTROPHES.

to differentiate themselves on the global stage, and the United States fights back against "high-tech illiberalism," the use of digital technologies to "polarize and pervert the politics of democratic societies."⁵⁵ In this environment, it has become more common to hear about competition and rivalry than cooperation.

Rhetoric and practices supporting an "AI race" for strategic advantage have proliferated, despite risks including cutting corners on safety and governance and increasing the likelihood of conflict.⁵⁶

We have already seen the difficulty of digital technology governance efforts that become mired in unsolvable debates about sovereignty and national priorities. For example, due to political and ideological differences between countries, the multi-year process to develop consensus on the international regulation of cyberspace at the UN Group of Governmental Experts has been unsuccessful and eventually resulted in two divergent processes.⁵⁷ Despite the current political

environment, the same fate is not inevitable for AI technologies. The partial international consensus of AI principles indicates significant common ground and potential opportunities including joint research initiatives, shared best practices, and alignment on international agreements. Moreover, existing differences do not mean that collaboration is out of reach; cooperation is possible without agreement on all principles.⁵⁸ This nuance may be more widely understood than often assumed; despite the increasing number of articles in the media that discuss AI competition, their proportion of all AI articles has been in decline since 2015.⁵⁹

Even in low-trust environments, opportunities arise for two of the world's AI superpowers to lead the charge in reducing accidents and limiting dangerous uses. 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.⁶⁰ Even as real and meaningful differences between AI principles continue, the risks of AI share similarities with climate change and global pandemics: they demand international cooperation.⁶¹ No country will benefit from the dangerous deployment of AI technologies that increase the likelihood of catastrophes.⁶² Shared AI principles represent a potential starting point for collaboration, which may reduce global security threats and advance the adoption of concrete accountability mechanisms for responsible AI.

KEY TAKEAWAYS

- **AI principles provide a starting point that can be leveraged to support further AI governance efforts.** Principles provide a common language and set of ideals against which stakeholders can be judged. They are generally insufficient on their own but have led to further work on tangible accountability mechanisms including policies and standards, which are gaining traction in both the United States and China.
- **Although both the United States and China have endorsed the OECD/G20 AI principles, substantive challenges remain in reaching shared understandings and opportunities for collaboration.** Analysis of U.S.-led and Chinese-led AI principles reveals different conceptualizations of some principles and some variation in priorities. Moreover, U.S. efforts to engage in international AI governance forums that exclude China may exacerbate the differences over the similarities.
- **Despite the political challenges, particular AI principles – including safety, security, explainability, sustainability, and fairness – show promise for U.S.-China research collaborations.** AI systems that are unreliable and vulnerable, that cannot be understood and managed, or that cause unintentional harm to people or the environment are dangerous for everyone. There is significant common ground and shared interest in reducing these risks, especially in research settings and multi-stakeholder forums. We

may expect that accountability mechanisms that address these challenges are more likely to gain traction in the near term.

Technology and policy related to this topic are constantly evolving. If you think we have missed something or have a comment, please contact Heather Evans (hevans@asiasociety.org).

The Asia Society Northern California Center and the Asia Society take no institutional position on matters of public policy and other issues addressed in the reports and publications they sponsor. All statements of fact and expressions of opinion contained in this paper are the sole responsibility of its author and may not reflect the views of the organization and its board, staff, and supporters.

ENDNOTES

- ¹ 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/>.
- ² 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>.
- ³ Joy Buolamwini and 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>.
- ⁴ 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>.
- ⁵ "AI Ethics Guidelines Global Inventory," AlgorithmWatch, April 2020, <https://inventory.algorithmwatch.org/>.
- ⁶ Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and 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.
- ⁷ Jessica Cussins Newman, "Decision Points in AI Governance," UC Berkeley Center for Long-Term Cybersecurity, May 2020, https://cltc.berkeley.edu/wp-content/uploads/2020/05/Decision_Points_AI_Governance.pdf.
- ⁸ Jeffrey Ding, Paul Triolo, and 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/>.
- ⁹ "Charlevoix Common Vision for the Future of Artificial Intelligence," G7, 2018, <https://www.mofa.go.jp/mofaj/files/000373837.pdf>.
- ¹⁰ "Launch of the Global Partnership on Artificial Intelligence," Government of France, June 17, 2020, <https://www.gouvernement.fr/en/launch-of-the-global-partnership-on-artificial-intelligence>.
- ¹¹ "OECD Principles on AI," OECD, Going Digital, May 2019, <https://www.oecd.org/going-digital/ai/principles/>.
- ¹² "G20 Ministerial Statement on Trade and Digital Economy," June 2019, <https://www.mofa.go.jp/files/000486596.pdf>.
- ¹³ "Preparing for the Future of Artificial Intelligence," Executive Office of the President National Science and Technology Council Committee on Technology, October 2016, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.
- ¹⁴ "Accelerating America's Leadership in Artificial Intelligence," Office of Science and Technology Policy, February 11, 2019, <https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/>.
- ¹⁵ "Guidance for Regulation of Artificial Intelligence Applications," Russell T. Vought, Executive Office of the President, Office of Management and Budget, January 2020,

<https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf>.

¹⁶ "The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update," Select Committee on Artificial Intelligence, National Science and Technology Council, June 2019, <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>.

¹⁷ "DOD Adopts Ethical Principles for Artificial Intelligence," U.S. Department of Defense, February 24, 2020, <https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/>.

¹⁸ "Artificial Intelligence: A Roadmap for California," The Little Hoover Commission, November 2018, <https://lhc.ca.gov/sites/lhc.ca.gov/files/Reports/245/Report245.pdf>.

¹⁹ "Artificial Intelligence at Google: Our Principles," Google AI, June 2018, <https://ai.google/principles/>.

²⁰ "IBM's Principles for Trust and Transparency," IBM Data Responsibility, May 30, 2018, <https://www.ibm.com/blogs/policy/trust-principles/>.

²¹ Francesca Rossi et al., "Everyday Ethics for Artificial Intelligence," IBM, <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>.

²² "Microsoft AI principles," Microsoft AI, 2020, <https://www.microsoft.com/en-us/ai/responsible-ai>.

²³ Olivia Solon, "Microsoft Funded Firm Doing Secret Israeli Surveillance on West Bank," *NBC News*, October 28, 2019, <https://www.nbcnews.com/news/all/why-did-microsoft-fund-israeli-firm-surveils-west-bank-palestinians-n 1072116>.

²⁴ Alexia Fernández Campbell, "The employee backlash over Google's censored search engine for China, explained," *Vox*, August 17, 2018, <https://www.vox.com/2018/8/17/17704526/google-dragonfly-censored-search-engine-china>.

²⁵ Jay Greene, "Microsoft won't sell police its facial-recognition technology, following similar moves by Amazon and IBM," *The Washington Post*, June 11, 2020, <https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/>.

²⁶ Danit Gal, "China's Approach to AI Ethics," NESTA, May 18, 2020, <https://www.nesta.org.uk/report/chinas-approach-to-ai-ethics/>.

²⁷ "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/>.

²⁸ Xiang Bo, "China issues principles of next generation AI governance," *Xinhua Net*, June 18, 2019, http://www.xinhuanet.com/english/2019-06/18/c_138152819.htm.

²⁹ Lorand Laskai and Graham Webster, "Translation: Chinese Expert Group Offers 'Governance Principles' for 'Responsible AI'," New America Cybersecurity Initiative, June 17, 2019, <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/>.

³⁰ *Ibid.*

³¹ "Beijing AI Principles," Beijing Academy of Artificial Intelligence, May 2019, <https://www.baai.ac.cn/news/beijing-ai-principles-en.html>.

³² *Ibid.*

³³ Mercedes Ruehl, "US blacklisting fails to derail ambitions of Chinese AI start-ups," *Financial Times*, July 28, 2020, <https://www.ft.com/content/124824d6-3b13-4dbb-8b38-926797f9b695>.

³⁴ "Megvii is to promote responsible AI development," Megvii Company Announcement, July 8, 2019, https://www.megvii.com/en/news/ID?news_id=117.

-
- ³⁵ “ARCC”: An Ethical Framework for Artificial Intelligence," Tencent Research Institute, April 9, 2020, <https://www.tisi.org/13747>.
- ³⁶ Jeffrey Ding, "Deciphering China's AI Dream," Centre for the Governance of AI, Future of Humanity Institute, University of Oxford, March 2018, https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf.
- ³⁷ Karen Hao, "The owner of WeChat thinks deepfakes could actually be good," *MIT Technology Review*, July 28, 2020, <https://www.technologyreview.com/2020/07/28/1005692/china-tencent-wechat-ai-plan-says-deepfakes-good/>.
- ³⁸ Heather Evans and Benjamin Cedric Larsen, "Exploring AI Issues Across the United States & China," Asia Society Northern California, August 5, 2020, https://asiasociety.org/sites/default/files/2020-08/Introduction-AI%20Series-08%3A05%3A2020.pdf?mc_cid=cc541be210&mc_eid=97aef0067e.
- ³⁹ Yi Zeng, Enmeng Lu, and Cunqing Huangfu, "Linking Artificial Intelligence Principles," AAAI Workshop on Artificial Intelligence Safety, 2019, <https://arxiv.org/pdf/1812.04814.pdf>.
- ⁴⁰ Ying Li, "Sustainability from a Chinese cultural perspective: the implications of harmonious development in environmental management," *Environment Development and Sustainability*, May 2015, https://www.researchgate.net/publication/277598843_Sustainability_from_a_Chinese_cultural_perspective_the_implications_of_harmonious_development_in_environmental_management.
- ⁴¹ Peijun Duan and Tony Saich, "Reforming China's Monopolies," Harvard Kennedy School, Faculty Research Working Paper Series, May 2014, https://ash.harvard.edu/files/reforming_chinas_monopolies.pdf.
- ⁴² "Anti-Monopoly Law of the People's Republic of China," Order of the President of the People's Republic of China, Chairman Order No. 68, August 30, 2007, http://www.gov.cn/flfg/2007-08/30/content_732591.htm.
- ⁴³ Danit Gal, "Perspectives and Approaches in AI Ethics: East Asia," *The Oxford Handbook of Ethics of AI*, Oxford University Press, June 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3400816#.
- ⁴⁴ 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/>.
- ⁴⁵ Paul Mozur, "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority," *The New York Times*, April 14, 2019, <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>.
- ⁴⁶ Ahiza García-Hodges, Chiara Sottile and Jacob Ward, "Man wrongfully arrested due to facial recognition software talks about 'humiliating' experience," *NBC News*, June 26, 2020, <https://www.nbcnews.com/business/business-news/man-wrongfully-arrested-due-facial-recognition-software-talks-about-humiliating-n1232184>.
- ⁴⁷ Danit Gal, "China's Approach to AI Ethics," Nesta, May 18, 2020, <https://www.nesta.org.uk/report/chinas-approach-to-ai-ethics/>.
- ⁴⁸ "Interim Report," National Security Commission on Artificial Intelligence, November 2019, <https://drive.google.com/file/d/153OrxnuGEjsUvIxWsFYauslwNeCEkvUb/view>.
- ⁴⁹ Andrew Imbrie and 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>.
- ⁵⁰ "The Global AI Talent Tracker," Macro Polo, <https://macropolo.org/digital-projects/the-global-ai-talent-tracker/>.

-
- ⁵¹ Damien Ma, Houze Song and Neil Thomas, "Supply Chain Jigsaw: Piecing Together the Future Global Economy," *Macro Polo*, April 13, 2020, <https://macropolo.org/analysis/supply-chain-ai-semiconductor-lithium-oled-global-economy/>.
- ⁵² James Vincent, "US consortium for safe AI development welcomes Baidu as first Chinese member," *The Verge*, October 17, 2018, <https://www.theverge.com/2018/10/17/17988582/partnership-on-ai-pai-welcomes-first-china-baidu-member>.
- ⁵³ "Tenets," Partnership on AI, <https://www.partnershiponai.org/tenets/>.
- ⁵⁴ Will Knight, "Baidu Breaks Off an AI Alliance Amid Strained US-China Ties," *Wired*, June 18, 2020, <https://www.wired.com/story/baidu-breaks-ai-alliance-strained-us-china-ties/>.
- ⁵⁵ Daniel Kliman, Andrea Kendall-Taylor, Kristine Lee, Joshua Fitt and Carisa Nietzsche, "Dangerous Synergies: Countering Chinese and Russian Digital Influence Operations," Center for a New American Security, May 7, 2020, <https://www.cnas.org/publications/reports/dangerous-synergies>.
- ⁵⁶ Dr Stephen Cave and Dr 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.
- ⁵⁷ Anders Henriksen, "The end of the road for the UN GGE process: The future regulation of cyberspace," *Journal of Cybersecurity*, Volume 5, Issue 1, 2019, <https://academic.oup.com/cybersecurity/article/5/1/tyy009/5298865>.
- ⁵⁸ 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>.
- ⁵⁹ Andrew Imbrie, James Dunham, Rebecca Gelles, and Catherine Aiken, "Mainframes: A Provisional Analysis of Rhetorical Frames in AI," Center for Security and Emerging Technology, August 2020, <https://cset.georgetown.edu/research/mainframes-a-provisional-analysis-of-rhetorical-frames-in-ai/>.
- ⁶⁰ Richard Danzig, "Technology Roulette," Center for a New American Security, May 30, 2018, <https://www.cnas.org/publications/reports/technology-roulette>.
- ⁶¹ Katrina vanden Heuvel, "We need a broad, transpartisan debate on how to engage with China before it's too late," *The Washington Post*, August 11, 2020, <https://www.washingtonpost.com/opinions/2020/08/11/we-need-broad-transpartisan-debate-how-engage-with-china-before-its-too-late/>.
- ⁶² 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/>.

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.

© 2020 The Asia Society. All Rights Reserved.