CENTER FOR LONG-TERM CYBERSECURITY

EXECUTIVE SUMMARY

CLTC WHITE PAPER SERIES

Toward AI Security

GLOBAL ASPIRATIONS FOR A MORE RESILIENT FUTURE

JESSICA CUSSINS NEWMAN

CLTC WHITE PAPER SERIES

Toward AI Security

GLOBAL ASPIRATIONS FOR A MORE RESILIENT FUTURE

EXECUTIVE SUMMARY

JESSICA CUSSINS NEWMAN

FEBRUARY 2019

CLTC
Center for Long-Term Cybersecurity

CENTER FOR LONG-TERM CYBERSECURITY

Overview

Artificial intelligence (AI) may be the most important global issue of the 21st century, and how we navigate the security implications of AI could dramatically shape the future.[1] Although research in AI has been advancing since the 1950's, recent years have seen substantial growth in interest, investment dollars, and jobs in this field,[2] leading to important advances in real-world applications ranging from autonomous vehicles to cancer screening.[3] It has become clear that AI is a transformative general-purpose technology that will spread across geographies and sectors, resulting in massive potential benefits—and risks—that are difficult or impossible to foresee. This presents a set of coordination and cooperation challenges to firms, governments, and civil society organizations that are trying to understand and act, prospectively, to shape the evolution of AI for human benefit.

This report uses the lens of global AI security to investigate the robustness and resiliency of AI systems, as well as the social, political, and economic systems with which AI interacts. The report introduces a framework for navigating the complex landscape of AI security, visualized in the AI Security Map. This is followed by an analysis of AI strategies and policies from ten countries, using the AI Security Map to identify areas of convergence and divergence. This comparative exercise highlights significant policy gaps, but also opportunities for coordination and cooperation among all surveyed nations. Five recommendations are provided for policymakers around the world who are hoping to advance global AI security.

The steps nations take now will shape AI trajectories well into the future, and those governments working to develop thoughtful strategies that incorporate global and multistakeholder coordination will have an advantage in establishing the international AI agenda and creating a more resilient future.

1 Allan Dafoe, "AI Governance: A Research Agenda," Governance of AI Program, Future of Humanity Institute, University of Oxford, August 27, 2018, https://www.fhi.ox.ac.uk/wp-content/uploads/AI-Governance_-A-Research-Agenda.pdf.

2 Alex Gray, "These charts will change how you see the rise of artificial intelligence," World Economic Forum, December 18, 2017, https://www.weforum.org/agenda/2017/12/charts-artificial-intelligence-ai-index/.

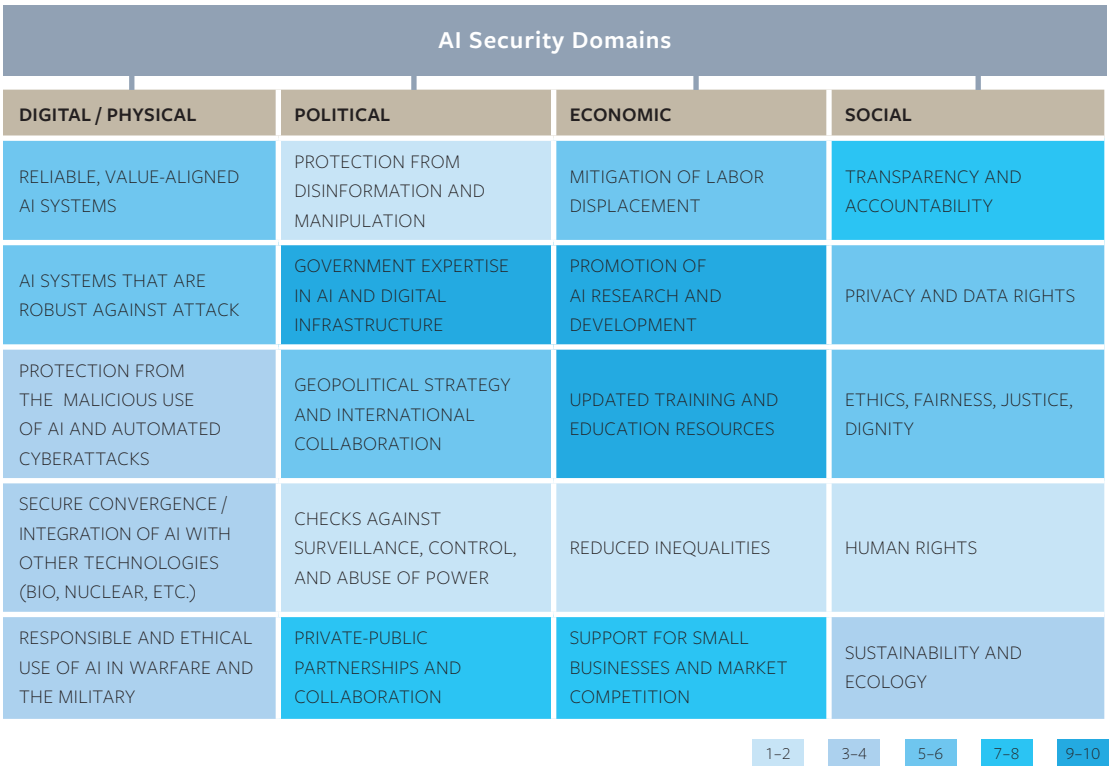
3 "Applying machine learning to mammography screening for breast cancer," DeepMind, November 24, 2017, https://deepmind.com/blog/applying-machine-learning-mammography/.

AI Security Map

The AI Security Map provides a simplified overview of the domains in which AI presents threats and opportunities, including: 1) Digital / Physical, 2) Political, 3) Economic, and 4) Social. The map is used first as a way to visually represent key domains and topics relevant to AI security, and later as a comparative tool to highlight which topics are addressed by different actors.

AI national strategies and policies from ten countries are assessed against the framework to provide visual and numerical comparisons between government approaches. While this analysis merely provides a snapshot of how AI challenges and threats have been framed in a sub-set of national strategies and policies, the comparison highlights interesting areas of convergence and divergence, and provides a lens into how AI security is being framed and addressed around the world.

Colors indicate the number of national strategies addressing each topic, based on analysis of ten countries: Canada, China, France, India, Japan, Singapore, South Korea, UAE, UK, and the US.



Global AI Security Priorities

Recent years have seen a significant increase in government attention to AI, as at least 27 national governments have articulated plans or initiatives for encouraging and managing the development of AI technologies.[4] Governments will play a key role in the development of AI: the actions that countries take now will shape AI trajectories well into the future, and those nations that work together will have an advantage in setting the international agenda. However, nations thus far have adopted highly divergent approaches in their AI policies, and there is significant variation in how they are preparing for security threats and opportunities. For example, only half the strategies surveyed discuss the need for reliable AI systems that are robust against cyberattacks, and only two mention challenges associated with the rise of disinformation and manipulation online.

Our analysis highlights a variety of areas of convergence and divergence, policy gaps, and opportunities for coordination and cooperation among nations as they advance their respective AI strategies. Among the key insights detailed in this report:

- A comprehensive response to AI security is multifaceted and encompasses social, economic, and political dimensions that reach far beyond conventional considerations of physical and digital security.
- Some governments—particularly in France, India, and South Korea—are leading the way in acknowledging and preparing for the breadth of disruptive implications of AI in the future.
- Only two priorities are shared by all ten of the countries included in the analysis: promoting AI research and development, and updating training and education resources.
- Countries have many additional opportunities to coordinate AI security strategies. For example, most countries are trying to address transparency and accountability of AI as well as privacy, data rights, and ethics. Most countries also prioritize private-public partnerships and call for improving digital infrastructure and government expertise in AI.
- The United States and China share many priorities for advancing AI, including international collaboration; transparency and accountability; updating training and educational resources; private-public partnerships and collaboration; creating reliable AI systems; and promoting the responsible and ethical use of AI in the military.
- Critical gaps in leadership remain around key issues. For example, only two (or fewer) national strategies address inequality, human rights, disinformation and manipulation, and checks against surveillance, control, and abuse of power.

4 "National and International AI Strategies," The Future of Life Institute, 2018, https://futureoflife.org/national-international-ai-strategies/.

Recommendations

Based on the analysis of gaps and opportunities in national AI strategies and policies, we provide five recommendations for policymakers hoping to harness and direct AI technologies for a more resilient and beneficial future. These recommendations outline concrete actions that can be taken now to address a complex and quickly changing sociotechnical landscape:

1. **Facilitate early global coordination where common interests can be identified.** As autonomous systems become more ubiquitous and capable, their reach and effects will be more consequential and widespread. Global coordination and cooperation will be essential for ensuring sufficient oversight and control, but such cooperation will be harder to achieve the longer we wait due to technological and institutional "lock-in". The numerous areas of convergence identified in this report can be leveraged as opportunities for collaboration and innovation, sharing best practices, and preventing global catastrophic risks.

2. **Use government spending to shape and establish best practices.** Governments have an opportunity to establish standards and best practices while promoting AI development and use, for example by implementing guidelines for government procurement of AI systems, and by adding criteria such as safety, robustness, and ethics to AI R&D funding streams. Additionally establishing processes to support transparent and accountable government funding and use of AI technologies will help prevent misuse throughout public services and protect government actors from the limitations and vulnerabilities of AI tools.

3. **Investigate what is being left on the table.** The landscape of AI security is broad and complex, as indicated in the AI Security Map presented in this report. The analysis of policy documents identifies many gaps in different nations' current AI policy approaches. Governments may choose to prioritize a sub-set of issues, but they should recognize the opportunities and challenges they could be neglecting.

4. **Hold the technology industry accountable.** Many governments rightfully emphasize the importance of partnership and engagement with industry and other AI stakeholders. However, while some firms are addressing AI challenges, significant gaps remain. Policymakers have the unique primary responsibility to protect the public interest, and this responsibility carries even greater weight during periods of significant technological transformation.

Governments should ensure their citizens have access to the benefits that emerge from AI development and are proactively protected from harms.

5. **Integrate multidisciplinary and community input.** To support the widespread goal of improving government expertise in AI, policymakers should formalize processes to ensure multidisciplinary input from AI researchers and social-science scholars and practitioners. Community engagement should additionally form an integral part of any decision to implement AI in public services.

CLTC

Center for Long-Term Cybersecurity

UC Berkeley

Center for Long-Term Cybersecurity
cltc.berkeley.edu
@CLTCBerkeley