October 2023

Report on the Impact of Artificial Intelligence and Associated Technologies on Arms Control, Nonproliferation, and Verification

DEPARTMENT OF STATE

International Security Advisory Board

DISCLAIMER

This is a report of the International Security Advisory Board (ISAB), a Federal Advisory Committee that provides the Department of State with a continuing source of independent insight, advice, and innovation on all aspects of arms control, disarmament, nonproliferation, outer space, critical infrastructure, cybersecurity, the national security aspects of emerging technologies, international security, and related aspects of public diplomacy. The views expressed herein do not represent official positions or policies of the Department of State or any other entity of the United States Government.

This report is accessible on the ISAB website.

https://www.state.gov/international-security-advisory-board/

United States Departement of State
Washington DC 20520

October 31, 2023

MEMORANDUM FOR UNDER SECRETARY BONNIE D. JENKINS

SUBJECT: Final Report of the International Security Advisory Board (ISAB) on the Impact of Artificial Intelligence and Associated Technologies on Arms Control, Nonproliferation, and Verification

This report responds to your request of October 18, 2022, that the Board undertake a study to advise the United States Government on how AI and associated technologies may impact arms control, nonproliferation, and verification. The report was drafted by members of a study group chaired by Dr. Shirley Ann Jackson. It was reviewed by all ISAB members and unanimously approved by all ISAB members present at the ISAB plenary meeting on October 31, 2023.

AI technologies are developing rapidly and they are transformative to security as well as to society as a whole. The Department of State and its U.S. government partners will need to make fundamental changes in order to address the risks and benefits of these rapidly developing technologies.

This report includes a series of finding and recommendations for the Department that address both risks and benefits of the application of AI. These findings and recommendations center on seven key areas: nuclear weapons and proliferation; biological and chemical threats; autonomous weapon systems; global supply chains, export, and trade; responsible state behavior and human rights; opportunities and applications in intelligence, and capacity building and human rights.

My board colleagues and I stand ready to brief you and other members of the Administration on this report.

Hon. Edwin Dorn
Chair
International Security Advisory Board

Contents

Executive Summary

Under Secretary of State for Arms Control and International Security, Bonnie Jenkins, charged the International Security Advisory Board (ISAB) to advise the United States on how artificial intelligence (AI), machine learning (ML), and associated technologies may impact arms control, nonproliferation, and verification, taking note of both the risks and benefits from its application.

In response, the ISAB Study Group on this topic framed its work around a set of key questions. These include: How is AI affecting the Under Secretary's core missions, such as nuclear, biological, and chemical (NBC) weapons treaty verification, use attribution, arms control, sensitive export controls, and responsible state behavior? What are the biggest risks, and how well does the current Department of State posture mitigate these? How might the Department better work with allies, and deter adversaries and potential competitors in this space? Finally, does the Department itself effectively engage on these topics internally and does it have the human capacity to ensure that it effectively manages these vital issues?

AI technologies are transformative to security as well as society as a whole. The Study Group believes that in many core areas the Department of State, the Under Secretary for Arms Control and International Security, and its U.S. government partners will need to make fundamental changes. The technologies themselves are developing rapidly, and a comprehensive account of them in this document would quickly be outdated. Rather, the Study Group has focused on potential threats, steps for Department leaders to take on policy, and capacity building within the Department itself.

The Study Group had meetings (in-person and virtual) with experts inside and outside of the Department of State, listed in Appendix C.

Our assessments have led to the following summary findings and recommendations for the Department of State that address both risks and benefits of the application of AI in arms control, nonproliferation, and verification, and in related areas. These areas include: 1) nuclear weapons and proliferation; 2) biological and chemical threats; 3) autonomous weapon systems;

4) global supply chains, export, and trade; 5) responsible state behavior and human rights; 6) opportunities and applications in intelligence; and 7) capacity building and human capital. This discussion of findings and recommendations is followed by a longer examination of each of these issues that details current challenges and potential future concerns.

Nuclear Weapons and Proliferation

Findings:

- U.S. programs to deter state and nonstate actors from pursuing new nuclear weapons programs focus on controlling weapons grade nuclear material, leveraging detection technologies, and denying access to said resources.
- New techniques leverage big data and AI to detect early warnings of emerging nuclear weapons programs through characterization of weapons-usable capabilities and/or consideration of advances in civilian, dual-use, and weapons-related nuclear science and technology, and through detection of the intent to change from civilian energy use to weapons use.

Recommendations:

- The Department of State, in partnership with the Intelligence Community (IC), should broaden its nonproliferation and deterrence approach to include early detection and deterrence, based on the use of big data, machine learning, and AI.
- The Department of State should leverage new methods developed by the Department of Energy's (DoE) National Laboratories to determine the weapons-usable capability of an emerging nuclear program, and to aid in detecting subtle indicators of strategic intent.
- The early indicator approach should be folded into deterrence and diplomatic strategies with regard to nuclear proliferation activities, such as sanctions and export controls, at an earlier stage of interaction with nation states.

Biological and Chemical Threats

Findings:

- The risks and potential timeline for risks at the intersection of AI/ML and biotechnology are uncertain, and the national security implications of the convergence are still unfolding.
- As with other technological advances, AI – especially generative AI – is potentially likely to lower barriers to biological and chemical weapons development through the synthesis of harmful pathogens and other malicious products.
- Currently, there is no consensus in the U.S. government regarding how to best approach the risks of the convergence of biotechnology and AI/ML.
- Domestic and global advances in biotechnology occur primarily in the private sector.

Recommendations:

- The Department of State should take this opportunity to lead international partners to prepare for the possible negative uses of AI/ML in biotechnology, to develop a shared understanding with allies about the risks of misuse, and to encourage the development of common levers to mitigate misuse.
- The Department of State should encourage allies and partners to further develop gene synthesis security domestically and through international mechanisms.
- The Department of State, along with other U.S. government entities, needs to foster new partnerships with the private sector to address national security risks related to AI and biotechnology in order to promote risk mitigation and responsible use of the technology, so as not to undermine the continued development of approaches, tools, and therapies crucial to disease mitigation and human health.
- The Department of State should support its own capacity building efforts to enhance understanding and management of critical and emerging technology-related national security risks, especially as these risks are only to become more pressing with time.

Autonomous Weapon Systems

Findings A:

- There is still no definition of Lethal Autonomous Weapon Systems (LAWS) in the international context.
- There is a general sense that states are making progress at the LAWS Group of Government Experts (GGE) as the various proposals, especially the Joint Proposal and the France/Germany-led proposal, get more detailed and come closer to alignment with each other. However, there is also growing external pressure to restrict the development and deployment of LAWS.
- U.S. policy on autonomous weapon systems (AWS) is consistent with U.S. interpretation of international and national law.
- The new U.S.-drafted Political Declaration on the Responsible Military Use of AI and Autonomy (hereafter referred to as the "Political Declaration") has increased salience given the undecided international legal environment.

Recommendations A:

- Continue working to bring additional states on board with the LAWS GGE Joint Proposal.
- The Department of State should not change its posture on the language of "human control."
- Develop a diplomatic strategy on how to approach the launch of an ad-hoc LAWS treaty writing process, potentially by Austria.
- Aggressively pursue agreement on the Political Declaration by a large group of states, and consider the following issues:
    - The appropriate venue for rolling out the group of states that have agreed to the Political Declaration.
    - Outreach to and leveraging of regional leaders, such as Singapore in Southeast Asia and Brazil and Chile in South America, to build greater regional support.

- Next steps for implementation following a roll-out, such as approaching the Department of Defense (DoD) about using the AI Partnership for Defense to discuss best practices and AI safety.
- Whether U.S. international outreach and activities on generative AI should include or exclude military applications of AI and autonomy. It will be important for any Department of State positions on large language models (LLMs) and generative AI to be consistent with the Joint Proposal and the Political Declaration, at a minimum.

Findings B:

- AI-informed systems may generate decision advantage with respect to both speed and quality of information for leaders as they balance risks and objectives in their decision making.
- Human use of AWS will rely upon decision support systems that themselves may be created using artificial intelligence approaches.
- Other countries might consider using AI-enabled autonomy in nuclear-capable systems.

Recommendations B:

- Diplomatic approaches to confidence building on the use of AWS should build on verifiable assurance of appropriate human judgment in the operation of such systems.
- Assessments of such systems within the U.S. government and with allies should focus on an agreed-upon approach to detecting the insertion of “poisoned data.” This could be a follow-on task for implementation of the Political Declaration.
- The U.S. government should work to increase its capacity to aggregate and synthesize situational information for human decision makers, which will enhance its ability to safely and responsibly use AI-enabled systems.

Global Supply Chains, Export, and Trade

Sensitive Export Controls

Findings A:

- The emergence and use of AI, especially generative AI and LLMs, raise questions about what commodities, software, and technology should be on the U.S. Munitions List (USML) – the training data, the software itself, or how it is embedded in military, or even dual-use systems.
- International Traffic in Arms Regulations (ITAR) was amended in 2020 to give more clarity to what confers military or intelligence advantage, thus warranting export and temporary import controls related to AI on the USML.

Recommendations A:

- The advent of generative AI requires a re-examination of ITAR. Partnering with the DoD, the Department of Commerce, and other agencies, the Under Secretary for Arms Control and International Security should lead a comprehensive interagency review of the definition of ITAR controlled data, with an eye to whether, and how, certain generative AI training data may confer military or intelligence advantage.
- The review should also include generative AI software, especially if it is embedded into critical infrastructure, command and control systems, and weapon systems.

Finding B:

- Much of the data aspects of ITAR are focused on data security and firewalls. Generative AI creates risks with respect to corruption of data and network architecture.

Recommendation B:

- In partnership with industry, the Department of State must have a stronger focus on training data for generative AI, data corruption, embedded software threats, and data and network architecture management, including AI-driven reconfigurable data and network architectures.

Findings C:

- Since most AI research and development occurs in the private sector, there will be increasing overlap and conflict between ITAR, which is administered by the Department of State, and the Export Administration Regulations (EAR), which are administered by the Department of Commerce. The EAR covers the export of items that are nominally commercial, but may have military applications, in a way that balances national security and commercial/research objectives.
- Given the dual-use nature of AI, continued strong alignment between the EAR and ITAR frameworks will be essential.

Recommendations C:

- The Department of State should participate with other U.S. government entities in the development of a harmonized national security approach regarding both ITAR and EAR requirements with regard to AI, especially generative AI and associated data sets.
- The Department of State should work with the Department of Commerce and the DoD to establish an interagency working group that monitors the application of export controls to AI, and ensures they are appropriately calibrated to emerging risks and support long-term U.S. economic and security interests.
- Consideration should be given to having one agency lead on AI export controls.

Finding D:

- More companies are requiring their supply chain members to be ITAR-compliant, but in AI software and training data, security vulnerabilities may be more difficult to detect.

Recommendation D:

- The Department of State should support work with industry on enhanced assurance within the AI software supply chain, including AI-enabled technology, with regard to embedded software in key military and commercial systems. This should include the certification and monitoring of software developers and their products, as well as the manufacturers of key physical technologies, including communication technologies and automated systems for logistics.

Critical Minerals and Materials

Findings A:

- In its quest to be a dominant nation, the PRC has focused on technological leapfrogging; dominance in the production of key technologies; and control of critical materials such as lithium, cobalt, rare earths, semiconductors, and the associated supply chains. The PRC also has imposed its own export constraints on materials it controls.
- An example is lithium, a critical material for a myriad of technologies, including AI systems. Australia is a vital supplier of lithium, along with the so-called "lithium triangle" bordered by Argentina, Bolivia, and Chile. The PRC has gained a foothold in this area through its "Belt and Road" initiative, deals with governments, and lax regulations. Today, essentially all Australian and "lithium triangle"-extracted product is shipped to China for processing.
- The PRC has positioned itself in the extraction and refinement stages of the supply chain for lithium, cobalt, nickel, and other critical materials through its own export controls, loose environmental policies, price manipulation, and the use of state-owned enterprises given large subsidies by the government.

Recommendations A:

- The Department of State should work with countries where lithium and other critical minerals are mined to build up a domestic refining industry and develop export controls.
- The Department of State should work with the Department of Commerce and other U.S. government agencies to craft trade agreements to preferentially source refined lithium and cobalt from countries whose governance is strong.
- The Department of State should work to form international coalitions to support domestic refining in source countries.

Findings B:

- The production of semiconductors, critical for all modern technologies, occurs across a broad global supply chain, with different parts and processes concentrated in certain regions.

- The United States leads in core intellectual property for the most advanced designs in chip design, electronic design automation, and advanced manufacturing equipment. East Asia leads in wafer fabrication, and China leads in the less skilled and less capital-intensive areas of assembly, packaging, and testing.
- China is investing heavily to move up and throughout the value chain.
- The U.S. government has begun to use economic incentives to address strategic vulnerabilities in the supply chain of semiconductors and related technologies.

Recommendations B:

- The United States, led by the Department of State, should collaborate with our allies and partner countries to gain more visibility into the lifecycle of critical technologies across every production phase. The Department should consider using AI for evaluating intelligence; risk assessment, analysis, and reduction; and for action, as warranted.
- The Department of State, together with other U.S. government agencies, should continue to strengthen existing economic incentives and disincentives, in collaboration with global partners, to take collective action to mitigate vulnerabilities, and to build up manufacturing capacity domestically and in countries allied with the United States.
- The Department of State should consider recommending economic sanctions related to intellectual property (IP), chip design, electronic design automation, advanced manufacturing, wafer fabrication, and raw materials access if IP theft, price distortion, or violence are used to gain advantage.
- Other tools should be included, including investment screening, visa policy, grant restrictions, entity lists, customs inspections, and export controls.

Finding C:

- The PRC is demonstrating its willingness to leverage its critical mineral dominance against U.S. interests, including by restricting exports of gallium and germanium, with implications for the AI supply chain.

- There is emerging concern with respect to mining of previously inaccessible minerals (especially in the Arctic), by Russia and other nations due to environmental and strategic considerations.

Recommendations C:

- The Department of State should work with relevant agencies on using Defense Production Act Title III authorities to ramp up domestic production of gallium and germanium.
- The Department of State should work with countries, like Germany, Japan, and Australia, that have high geological potential in gallium and/or germanium to unlock new resources and diversify the market.
- The Department of State should re-examine the applicability, in this context, of the Law of the Sea as it relates to seabed mining and deep-sea mining.

Emerging Security Risks of Artificial Intelligence

Finding:

- Current controls to prevent adversaries accessing the most advanced AI capabilities have left gaps, including accessing AI computational power or resources (AI compute) through the cloud.

Recommendations:

- The Department of State should work with the Department of Commerce and relevant agencies and industry to develop a Know Your Customer (KYC) scheme for advanced AI cloud compute.
- The Department of State should work with like-minded international partners to introduce consistent KYC schemes and international governance for an aligned approach.

Responsible State Behavior and Human Rights

Findings A:

- The United States has begun leading on norms around AI and security, with the Political Declaration an important step. Aspects of the current international environment are

favorable for the United States to lead in developing important AI-related norms on security, safety, and human rights.

- AI capabilities are advancing more rapidly than the attendant diplomatic and regulatory frameworks.
- AI companies are expanding the capabilities of their technologies, often without a commensurate expansion of safety and security capabilities.
- In the absence of robust cyber and physical security, adversaries have the opportunity to monitor and steal algorithms and data from U.S. companies.

Recommendations A:

- The Department of State should use the Political Declaration with its close allies and partners to promote norms around the use of AI for national security.
    - The Department of State should encourage all allies to sign on to the Political Declaration and be prepared to make legitimate changes in response to allies' varied interests.
    - As close allies sign on, the United States should broaden its efforts to include greater regional engagements in South America and elsewhere.
- The Secretary of State should charge relevant parties, bureaus, and offices, in the Department, to develop norms and standards for data sharing with our allies, in consultation with the White House, the DoD, and others in the U.S. government.
- The Department of State should push several specific norms concerning the threat of AI to open, democratic societies. Two to consider are that "deepfakes" generated by AI systems should not be used to encourage conflict, and that deepfakes should not be used to undermine another government's democratic processes.
- The Department of State should elevate the importance of AI safety and assign appropriate high-level responsibility and resources.
- The Department of State should drive global engagement on cyber and physical security, as well as safety, in AI-related work.

- In order to assess their potential impact in its areas of responsibilities, the Department of State should consider ways to monitor and estimate increases in new AI capabilities through surrogate variables such as computer power.

Finding B:

- The United States and the PRC have shared interests in AI safety, such as avoiding accidental launches of weapon systems.

Recommendations B:

- The Department of State should continue attempts to engage the PRC bilaterally on AI safety and promote Track 1.5 and Track 2 efforts to do so.
- The U.S. government should identify mutually beneficial research on AI safety for collaboration with the PRC and a process for weighing the advantages and disadvantages of such collaboration. It will be important to focus only on measures that increase the safety of PRC AI systems but do not inadvertently enhance PRC military capabilities by making PRC AI systems more reliable and robust.

Opportunities and Applications in Intelligence

Findings A:

- Although new work is underway, there has so far been limited current application of AI tools to the arms control mission within the Department of State.
- With the New Strategic Arms Reduction Treaty (New START Treaty) entering its final years, and with limited arms control structure governing U.S.-PRC relations, the Bureau of Arms Control, Deterrence, and Stability (ADS – formerly the Bureau of Arms Control, Verification and Compliance) problem set is likely to shift to more risk reduction and threat management, where more advanced technologies are likely to be valuable in a risk-reduction and threat-management context governing U.S.-PRC and U.S.-Russian relations going forward.

- The U.S. government has much less hands-on insight into PRC strategic programs due to the lack of a bilateral arms control treaty with in-person monitoring, less overall history operating with their strategic programs due to their comparatively recent ramp up, and the general opacity of PRC military activities.

Recommendations A:

- The Department of State, with partners in the U.S. government, should enlarge its use of AI in applications to risk reduction, confidence building measures, and threat management, where advanced technologies, such as AI are likely to be valuable.
- New technologies, such as generative AI, should be used to gain a better understood baseline of PRC-backed programs, activities, and patterns of life. Technical insight into novel programs would also be valuable.
- To the extent that it is not already being pursued, the Department of State should advocate for U.S. government programs which conduct AI/ML-based investigations of indicators of nuclear, chemical, and biological weapons programs.

Finding B:

- The nature of potential future strategic competition between the United States and the PRC could vary widely.

Recommendations B:

- The Department of State should be preparing simultaneously for multiple potential futures that capture everything from a re-opening to a post-conflict environment.
- The Department of State should plan for future uncertainty through structured assessment of potential future states, including early signs of pathways to desired end states.

Capacity Building and Human Capital

Findings A:

- AI and other advanced technologies are rapidly changing, and thus knowledgeable personnel are vital to adapt and adjust programs and policies.
- The Department of State needs more knowledge of AI, biotechnology, and other emerging technologies, and of science, technology, engineering, and mathematics (STEM) topics in general. This is difficult as the personnel system is underfunded and under-supported.

Recommendations A:

- Foreign service and civil service job descriptions, examinations, and hiring processes should stress STEM more. This would include more direct hiring of individuals with science and engineering expertise and ensuring that applicants overall have stronger STEM literacy. This is consistent with current approaches on international affairs, writing skills, and economics.
- The Department of State should have fellowships for students, similar to the Thomas R. Pickering Foreign Affairs Graduate Fellowship and Rangel Graduate proc programs focused on underrepresented communities, to provide expedited hiring of qualified STEM candidates.
- The Department of State personnel offices should identify those with STEM backgrounds (not information technology) currently in the Department and determine their satisfaction levels.
- There should be additional resources given to recruit, promote, and retain employees in STEM areas. Those with technical skills in vital but under-staffed areas could receive bonuses, accelerated promotion, and other benefits in order to retain them.
- The Department of State should offer longer-term internships to take advantage of emerging technical talent. A longer internship would enable the participants to make a far greater contribution to the Department of State and give the Department access to emerging professionals with impressive technological skills.

- The Department of State should incentivize the development of skills related to emerging technologies in the current workforce. As it does with foreign language acquisition, there should be bonuses and additional opportunities for those who learn skills such as data science, Python or other AI-related languages, advanced biochemistry, and so on.

Finding B:

- Because the DoE National Laboratories, the DoD Defense Threat Reduction Agency (DTRA), the military services, and other parts of the government have significant expertise on emerging technologies. The Department of State needs to gain access to, and to partner with, such government entities.

Recommendation B:

- The Department of State should develop liaisons, working groups, rotations, and other means of accessing the capabilities of the National Laboratories, DTRA, and other government experts outside the Department.

Findings C:

- Because much of the expertise is outside government, the Department of State needs more linkages and access to private sector expertise on AI, biotechnology, and other emerging technologies.
- The Department of State currently benefits from the rotation of experts from academia.

Recommendations C:

- The Department of State should identify individual offices and officers with the responsibility for engaging with the leading AI companies. The companies should also be encouraged to assign a set of people to engage with the Department of State to ensure regular interactions.
- Programs such as the Intergovernmental Personnel Act (IPA) Mobility Program or those modeled on the Defense Science Study Group could be expanded to attract more academic talent.

- The Department of State should develop a diplomatic “reserve” program of people working in civilian organizations that are advancing AI and emerging technologies. The program would be informed by the U.S. military reserve programs but tailored to the needs of the Department of State. The individuals involved could provide links to private sector expertise and could be called upon in a crisis situation.
- The Department of State should encourage the U.S. government to subsidize certain types of national security-related research, especially those related to technology competition with the PRC.
- The Department of State should explore programs that allow private company officials with AI expertise to participate in decision-making on export controls and negotiations, subject to appropriate controls and oversight.
- For any personnel brought into the Department of State on special programs (Reserve, IPAs, long-term interns, etc.) the Department should have a plan for post-recruitment placement to ensure they are used optimally.

Overview and Context: AI and Emerging Technologies

Artificial intelligence/machine learning and other emerging technologies can prove beneficial to U.S. diplomacy and national and international security interests, but also pose risks. With recent advances in large language models, such as ChatGPT, AI has been the focus of great attention and debate lately. These technologies have applicability across multiple domains, including national security, global stability, the rule of law, commerce and trade, bias, privacy, and human rights. The questions of when and where these technologies are most useful, which ones should be used, and how to do so, have important implications for U.S. diplomacy with allies, competitors, and enemies. The Department of State and the Under Secretary of State for Arms Control and International Security should play an important role in addressing these questions. Its work should be carried out in partnership with the Under Secretary of State for Economic Growth, Energy, and the Environment which leads the Department of State's efforts to develop and implement international policies related to economic growth, energy, agriculture, the ocean, the environment, and science and technology. The Cold War dynamic of technology developed for military uses being adapted to civilian uses has evolved. Governments today leverage commercial technologies for their own purposes – for good or ill. Most AI advances have come from academia and especially technology companies, and the Department of State must engage with the private sector in new ways. Historically, the Department's engagement with the private sector has been led by the Under Secretary for Economic Growth, Energy, and the Environment's Office of Global Partnerships.

For the Under Secretary for Arms Control and International Security, and the Department of State overall, the broader national and geopolitical context within which AI and other emerging technologies will be operative is complex. At the forefront of this context are the national strategies and geopolitical intent of the PRC. In October 2022, President Xi directed that PRC's science and technology (S&T) development emphasize self-reliance rather than foreign investment and international collaboration. This stance, and concern about the PRC's focus on access to, and exploitation of, sensitive S&T developed in the United States, represent a real

threat to the U.S. economy and security and from an intelligence and counterintelligence perspective.

In a number of technology and resource arenas, the PRC already has a leading position. Examples include photovoltaics, telecommunications companies such as Huawei, mobile grid equipment, and access to and control of key resources like lithium and rare earth supply chains. Now the PRC is giving particular attention to AI and quantum information science. The limited high-level engagement between the United States and the PRC exacerbates the overall geopolitical situation.

In addition, Russia remains, and in many ways has re-emerged, as a strategic threat, particularly since the beginning of the conflict in Ukraine. Adding to the danger is the apparent involvement of both Iran and the PRC on behalf of Russia in the conflict.

The risks are myriad. Of particular focus are nuclear, chemical, and biological weapons proliferation, including the possible use of generative AI in synthetic biology. Also of vital importance are autonomous weapon systems; export, trade, and supply chains; and the use of AI in violations of privacy and human rights.

The traditional mission of the Under Secretary for Arms Control and International Security has rested on monitoring military equipment and nuclear installations, using treaties and well-established methodologies to verify compliance or violations. Now, with Russian aggression in Ukraine and intransigence on the New START Treaty, with no real arms control structure vis-a-vis the PRC, and the advent of technologies that make attribution and verification more difficult and that can lead to new threats, the focus of the Under Secretary has to enlarge to focus more on risk assessment and threat reduction through the appropriate use of these same new technologies.

This broadened perspective by the Under Secretary for Arms Control and International Security must also be accompanied by greater cross-Department collaboration, especially with the Under Secretary for Economic Growth, Energy, and the Environment and the regional bureaus, to address the security dimensions of technological and economic competition and the dual-use risks and benefits of AI and these other emerging technologies.

Each risk area also represents an opportunity to apply AI and emerging technologies to mitigate risks and reduce threats. To date, there has been limited use of new tools derived from AI and emerging technologies in the Department of State, although there are a number of nascent activities.

In this report we delineate our findings and recommendations with regard to risk recognition and threat reduction in seven key areas:

1. Nuclear weapons and proliferation
2. Chemical and biological threats
3. Autonomous weapon systems
4. Global supply chains, export, and trade
5. Responsible state behavior and human rights
6. Opportunities and applications in intelligence
7. Capacity building and human capital

Nuclear Weapons and Proliferation

Proliferation detection is challenging; indicators are sparse against a complex and noisy background. Advances in AI-enabled technologies and the availability of new data sources present new opportunities to further enhance U.S. nuclear proliferation detection. Through measures including diplomacy, policies and treaties, threat reduction assistance, and export controls, nuclear nonproliferation seeks to dissuade or prevent state and nonstate actors from proliferating nuclear-weapons usable capabilities or make more costly their access to sensitive technologies, material, and expertise.

Ongoing research and development is underway to build AI and analytics systems that are suitable for the challenges and requirements of national security missions.

These next generation AI technologies may enable detection of strategic changes in the intent of foreign nuclear programs earlier than before to inform U.S. competition with Russia and the PRC.

Foundational to U.S. nuclear nonproliferation and arms control is the use of technologies and scientific capabilities to detect weapons-usable nuclear material and whether existing nuclear programs are intended for peaceful or military applications.

As such, many nuclear proliferation detection technologies focused on specialized nuclear material and equipment unique to nuclear weapons development. Leveraging advances in data science and computing as well as new data sources, ongoing research efforts are developing innovative, AI-enabled techniques to reveal additional indicators of nuclear proliferation. These next generation methods reveal subtle clues that may indicate a change in capability or even a change in strategic intent of foreign nuclear weapons programs.

New techniques leverage big data and AI to detect early warnings of an emerging nuclear weapons program by characterizing the weapons-usable capability of advances in civilian, dual-use, and weapons-related nuclear science and technology and detecting subtle indicators of changes in intent from civilian to military use. This may enable intervention when a program first diverges from peaceful purposes and deterrence is more likely to be successful.

Dual-use Nuclear Science and Technology: Compliance and Verification

It is likely that any new nuclear weapons program will leverage dual-use research and nuclear energy science and technology to clandestinely advance weapons-usable capabilities. Researchers have demonstrated new methods to determine the weapons-usable capability of an emerging nuclear program and detect subtle indicators of change in strategic intent from peaceful to military use. Earlier proliferation detection methods may provide policy and decision makers with essential, timely information to develop deterrence strategies for a wider range of nuclear proliferation activities, like sanctions and export controls, to deny key resources beyond nuclear material and impose greater costs to acquiring them. Earlier

proliferation detection will enable the United States to deter programs at early stages of development when the stakes are lower and there is limited investment in purely weapons-applicable capabilities. At this point, before public declaration of malintent by a withdrawal from the Nuclear Nonproliferation Treaty (NPT) and long before the celebration of a successful nuclear weapons test, it may be possible to devise strategies to deter further advances toward military use.

Impact of Generative AI

Generative AI tools offer, and are beginning to demonstrate, transformational impacts in the broad domain of nuclear weapons and proliferation, both by integrating huge bodies of information for enhanced decision support and by delivering rapid technical advances relevant for weapon design, development, and deployment. Enhanced decision support approaches are being enabled by generative AI, and LLMs in particular, due to the enhanced capability to identify subtle correlations across extremely large data sets. This will enable major improvements in proliferation detection, characterization, and understanding. For weapons, generative AI tools offer capabilities for rapid discovery of critical weapons materials and optimized system designs that can counter (or be) evolving threats.

Biological and Chemical Threats

The convergence of AI and ML platforms with cutting-edge biotechnology and biochemistry is an emerging and rapidly evolving field with promising development for biomedical research. ML techniques have been applied to accelerate drug discovery and development processes, unlock precision medicine advances, and revolutionize medical imaging and diagnostics. ML algorithms have been instrumental in analyzing and interpreting vast amounts of genomic and biological data, screening of compound libraries, predicting drug-target interactions, streamlining genetic engineering workflows, and designing synthetic biological systems.

However, the application of AI within the biotechnology arena could also enable the potential accidental misuse or potential abuse of machine learning techniques for synthesizing harmful pathogens, chemicals, and other malicious products. The dual-use nature of AI and

biotechnology advancements presents new challenges for those who assess national security risks associated with intentional technological misuse, such as state-sponsored biological and chemical weapons programs, and actions by non-state actors.

For the majority of pathogen research conducted worldwide, there are laws and regulations that govern, guide, and oversee these benign research activities. Within the United States, all research with biological select agent and toxins (BSAT) and research on a small subset of pathogens that have pandemic potential is subject to additional scrutiny to assess and mitigate biosecurity and biosafety risks. Globally, the risk of unintended biosecurity and biosafety consequences for research on potential pandemic pathogens, especially those that may also utilize AI technology, remains unclear.

Additionally, the dissemination of misinformation about biological security has an impact on nonproliferation and the Department of State's mission. The Department plays a critical role in understanding and mitigating national security concerns that may arise from these emerging and converging technological applications. The convergence of several technological advancements has expanded the applications of and outcomes that result from the convergence of AI and biotechnology. Those advancements are the acceleration of computing power to run AI applications, the sophistication of open-source AI applications themselves, and the proliferation of multi-omics biological databases upon which to train AI applications.

The risks and potential timeline for risks at the intersection of AI and biotechnology are uncertain. It is clearly being developed for positive reasons to identify potential drug targets for new medicines, and this same approach could be used to develop targets for weapons. However, it is still unknown how immediately successful either positive or negative endeavor will be, and if there is enough biological scientific information in the right format available to be useful for AI techniques to aid with weapons development.

AI is likely to lower barriers to biological and chemical weapons development. The convergence of AI and biotechnology is likely to further accelerate the research and development process for novel biological and chemical weapons development, by reducing the need for research and development to develop possible candidate weapons. This could lower the resource-intensive

nature of biological weapons development and could make this option more attractive to state and nonstate actors with malicious intent.

Currently, there is no consensus in the U.S. government regarding how to best approach the risks of the convergence of biotechnology and AI/ML. The Department of State currently has a small number of personnel who are monitoring the national security risks related to AI and biotechnology. We found that they have a sophisticated understanding about how the technologies and capabilities are advancing, and are connected with the IC with respect to possible implications. However, a plan of action to take advantage of both potential risks and benefits has yet to be formed.

By evaluating intelligence and collaborating with intelligence agencies and research institutions, the Department of State can gain insights into emerging threats and vulnerabilities. It can then share this information with relevant stakeholders to facilitate informed decision making and risk mitigation strategies.

The Department of State has an opportunity to lead international partners to prepare for the possible negative uses of AI in biotechnology, to develop a shared picture with allies about the risks of misuse, and to encourage the development of common levers to mitigate misuse. The pace of AI has been rapid, but it is likely just the beginning of its exploration for potential benefits and risks in biotechnology. Building a foundation of shared understanding will be useful in the years to come. Advances in biotechnology have often been hyped at first, and disappoint observers when advances do not come immediately to fruition. However, longer term advances often exceed the initial projections. The Department of State should consider that this trajectory may be duplicated with AI in biotechnology, giving time to pursue these aims with international partners. The Department can create, promote, and advance international collaboration and information sharing mechanisms to effectively address biosecurity risks effectively and build consensus responses to rapidly evolving threats. This involves sharing best practices, harmonizing regulations, and facilitating communication between governments, research institutions, and industry stakeholders. International agreements and partnerships at the early stage of this technology convergence can establish global norms and standards for the responsible use of machine learning and biotechnology.

The Department of State can encourage allies and partners to further develop gene synthesis security in their nations and through international mechanisms. Gene synthesis products are commonly used in research and biotechnology pursuits, and are likely to be increasingly important as AI in biotechnology advances. To date, the United States is the only nation that has guidance for gene synthesis companies to screen orders and customers to prevent misuse. The acquisition of gene synthesis products could allow unauthorized people to avoid export controls and mechanisms (including Australia Group) to more easily synthesize dangerous pathogens de novo. There is an opportunity for nations to make this process more challenging by requiring gene synthesis companies to screen and to only allow national research monies to be spent on companies that screen. The state of California, through the law (AB 1963, 2022), has taken this approach to only allow research money to be used for companies that have implemented appropriate security measures. Steps taken in this area will be immediately useful to reduce bioterrorism risks but as AI progresses may become even more important.

In conjunction with other U.S. government entities, partnerships with the private sector are critical to addressing national security risks related to AI and biotechnology and promoting risk mitigation and responsible use of the technology. These collaborative discussions with technology companies, research institutions, and industry associations could result in a range of outcomes including, regulatory frameworks, global norm-setting, standard security protocols, and risk assessment requirements.

Autonomous Weapon Systems

The United States leads the world in the depth, transparency, and responsibility of its policy surrounding AWS, and U.S. international engagement on AWS reflects a strong interagency partnership on the issue. The best path forward on AWS for the United States is to continue promoting the Joint Proposal in Geneva, as part of the LAWS GGE process in the Convention on Certain Conventional Weapons (CCW). If the LAWS discussions ever fall out of the GGE due to a prominent-enough country pulling together an ad-hoc convention designed to create a LAWS equivalent to the nuclear ban treaty, the United States should advocate that like-minded countries use the Political Declaration as the base text.

U.S. Policy on LAWS

U.S. policy on autonomous weapon systems is designed to be consistent with U.S. interpretation of international and national law. One of the foundations is the DoD Directive 3000.09, Autonomy in Weapon Systems, which was update in January 2023.

The Directive is one element of policy and guidance that demonstrates U.S. commitment to developing and deploying its weapon systems and other advanced capabilities in a responsible and lawful manner. The Directive was established to minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements. The requirements established in the Directive include the following:

- AWS, unless specifically exempted, undergo senior leader reviews prior to development and fielding.
- AWS will be designed to allow senior civilian policymakers to exercise appropriate levels of human judgment over the use of force.
- Persons who authorize the use of, direct the use of, or operate autonomous and semi-AWS will do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement.
- AWS will go through rigorous hardware and software verification and validation (V&V) and realistic system developmental and operational test and evaluation (T&E) to ensure they function as anticipated in realistic operational environments against adaptive adversaries and are sufficiently robust to minimize failures.
- The design, development, deployment, and use of AI capabilities in autonomous and semi-autonomous weapon systems will be consistent with international law and presidential guidance.

U.S. International Engagement on LAWS

The United States and several allies and partners (Australia, Canada, Japan, Poland, South Korea, and the United Kingdom) support the LAWS GGE process at the CCW. The Joint Proposal is the most detailed existing proposal by any country or set of countries in the LAWS process, and is recognized as positive by most states. A key difference between states within the LAWS GGE involves whether a legally binding instrument (LBI) is necessary to achieve these goals. The United States views international humanitarian law (IHL) as strong and the U.S. commitment to IHL as universal, whether a weapon system is autonomous, semi-autonomous, or anything else, and thus an LBI is not necessary. Moreover, there is still no definition of a LAWS in the international context.

One key question for the United States moving forward involves the language used to describe what the international community should prefer in thinking about lawful weapon systems. U.S. policy is to favor "appropriate levels of human judgment" over the use of force, which is consistent with IHL. Many other countries, including many European allies and partners, favor the language of "human control" to describe the obligation of the human in the use of force.

The United States should continue resisting pressure to shift to the language of control. Human control as a phrase implies a degree of direct supervision to the point of impact of a weapon that is inconsistent with decades of IHL, and established practice. For example, while human operators launch precision-guided munitions, those munitions often have their own guidance systems and not many have direct human supervision to the point of impact. While it might seem attractive diplomatically to accede to the language of human control in the LAWS context to build support there, the second-order consequences could be substantial, including placing the legality of precision weapons at risk.

The potential of the use of generative AI and LLMs in autonomous weapon systems adds additional complications.

International State of Play on LAWS

There is a general sense that states are making progress at the LAWS GGE as the various proposals, especially the Joint Proposal and France/Germany-led proposal, get more detailed and come closer to alignment with each other. However, there is also growing external pressure to restrict the deployment of LAWS. This campaign has been working to convince countries, including Chile, Costa Rica, Pakistan, the Philippines, and others to seek a full prohibition on LAWS through a legally binding instrument.

The goal of the conference would be to draft an international ban on LAWS similar to the land mines treaty or the nuclear ban treaty. Since the CCW operates based on consensus, it would only take one state refusing to continue the conversation in the GGE to make the GGE no longer an option for dialogue. Thus, some degree of progress in the LAWS GGE process is necessary to persuade countries to remain on board.

If an ad-hoc treaty effort on LAWS is created, it will make the new U.S.-drafted Political Declaration even more important. The Political Declaration reflects interagency collaboration and creates strong norms. It is much broader than LAWS, including all military uses of AI and autonomy, which makes sense since LAWS are only an extremely small set of the potential military applications of AI and autonomy.

The Department of State is currently leading efforts to bring other countries on board with the Political Declaration as the interagency works to finalize the text of the Political Declaration in response to feedback by allies and partners. The more countries that endorse the Political Declaration, the stronger the position the United States will be if an ad-hoc process on LAWS forms. The Political Declaration commits states to responsible behavior, a degree of transparency on their AI policies, and next steps on implementation.

Considerations for AI Use in Conventional Weapons Versus Nuclear Capable Systems

In the use of lethal autonomous weapon systems, the authority to perform certain functions is delegated to an embedded algorithmically driven set of actions. Barring any international agreements, such weapons will be developed and used when the benefit of not having a human in the loop (e.g., very fast decisions, operations in denied areas) outweighs the consequence of a possible bad decision by the algorithm. This will require the systems to have very high levels of safety and reliability. As the consequences of a poor autonomous decision increase (as they would in going from conventional to nuclear weapons), the set of circumstances in which one would be willing to delegate that authority diminish greatly. U.S. policy, as described in the Nuclear Posture Review, makes clear the U.S. commitment to human involvement in decisions involving the employment of nuclear weapons.

Hence, LAWS requires serious analyses of decision uncertainties in their applications to avoid unwanted or even devastating outcomes. This need for uncertainty quantification (UQ) will drive the development of the underlying AI technologies for autonomous weapon systems as well as the systems that control their use.

It is conceivable that conventional weapons could use AI-driven autonomy to greatly shorten response times and increase effectiveness. For nuclear weapons, on the other hand, the prudent route to improvement likely lies in aggregating and synthesizing situational information to yield decision superiority for leaders with a human remaining the ultimate decision maker.

Next Steps

Alongside all of the above, there is growing pressure for international engagement on generative AI connected to public discourse on ChatGPT and related LLMs. The Department of State should continue working to bring additional states on board with the LAWS GGE Joint Proposal and not change its posture on the language of “human control.” The Department of State should also develop a diplomatic strategy for how to approach the launch of an ad-hoc LAWS treaty writing process, potentially by Austria.

The Department of State should aggressively pursue agreement on the Political Declaration by a large group of states. It should consider the appropriate venue for rolling out the group of states that have agreed to the Political Declaration; outreach to and leveraging of regional leaders, such as Singapore in Southeast Asia and Brazil and Chile in South America, to build greater regional support; and next steps for implementation following a roll-out, such as approaching DoD about using the AI Partnership for Defense to discuss best practices and AI safety. Finally, it will be important for any Department of State positions on LLMs and generative AI to be consistent with the Joint Proposal and the Political Declaration, at a minimum.

In ongoing discussions of autonomous weapon systems, there should be discussions about the importance and need for rigorous processes and checks to gain ongoing confidence in the operations of autonomous systems. Even an approach that emphasizes human judgment will rely upon decision support systems that may well be created using artificial intelligence approaches. Citizens and leaders will want assessments that are checked and on-guard for potential insertion of "poisoned data," that are augmented with new data from recent encounters in relevant theaters, and whose algorithms are developed and tested for relevant scenarios (e.g., topography, infrastructure, and environmental conditions). In crisis, AI-informed systems may generate decision advantage with respect to both speed and quality of information for leaders as they balance risks and objectives in their decision making, but there must be confidence building with respect to the operations of such systems and the data they ingest.

Global Supply Chains, Export, and Trade

Exports: Arms Control and Nonproliferation and ITAR

The Under Secretary for Arms Control and International Security's responsibilities include administering ITAR. ITAR controls exports and important defense-related goods and services on the USML. All manufacturers, exporters, and brokers of defense goods and services, or related data must be ITAR-compliant.

With the emergence of AI in weapon systems and dual-use technologies, and with the advent of generative AI, there is a complication of what should be on the USML – the software itself or how it is embedded in military (including command and control and weapons systems), critical infrastructure, and dual-use systems. In particular there needs to be greater clarity of the role of AI training data in AI-enabled systems, and of data used in generative AI, and generative AI algorithms themselves in defining what constitutes ITAR-controlled data.

While much of the data aspects of ITAR are focused on data security and firewalls, generative AI creates risks with respect to the corruption of data and of network architecture, and other embedded software threats. In addition, data and network architecture management will become increasingly important in managing threats.

More companies are requiring their supply chain members to be ITAR-compliant. But, in the software area, this is difficult, especially with respect to the certification and monitoring of software developers, and software validation and verification. Interestingly, AI itself can be deployed in certain instances to assure software validation and verification, and to detect malicious generative AI.

Since most AI research and development occurs in the private sector, there will be increasing overlap and conflict between ITAR and EAR, which is administered by the Department of Commerce. The EAR covers the export of items that are nominally commercial, but may have military applications, in a way that balances national security and commercial/research objectives. To date, the EAR has been the framework of choice to administer AI export controls, given the dual-use nature of AI.

Global Supply Chains and AI

The advent of AI and other emerging technologies require a broader examination of global supply chains and their inter linkages – broader than heretofore has been undertaken. Of necessity, consideration must include key materials for new technologies, the supply chains for new technologies themselves, as well as how AI and related technologies are, or can be, embedded in key systems.

Areas of concern include key metals such as cobalt and lithium, especially important for advanced computational systems, including AI-based systems, modern micro/nano-electronic devices, automated and autonomous systems, and advanced control systems. Other critical materials such as rare earths and semiconductors such as gallium and germanium are also of importance.

Other key concerns are the possibility of embedding AI-driven malware into critical imported technologies and infrastructure, including communication networks and automated systems for logistics. For example, at the Port of Los Angeles, and other key ports in the global trade supply chain, the automation system for ship loading and unloading comes from China. If such a system is controlled remotely, or is improperly AI-driven, it can affect the delivery and flow of imported and exported goods. Such a system could also be designed to ingest critical data with respect to the development, vulnerabilities, and capabilities of key areas of the U.S. economy. There is always the concern of the development of, and illicit trade in, AI-enabled weapon systems.

Maintaining U.S. Leadership in Artificial Intelligence

It is in U.S. interest to take active steps to maintain technological leadership in AI for security as well as for economic reasons. With the United States and the PRC continuing to compete in these fields, it is clearly in the U.S. interests to safeguard critical components and developments from proliferating and feeding into adversarial capability. The October 2022 export controls on advanced semiconductors – the key component in AI hardware – were designed precisely to manage this risk. Yet at the same time, overly restrictive actions risk dampening U.S. innovation, or incentivizing other countries to fill gaps left by U.S. restrictions.

Critical Minerals and Materials

With respect to critical minerals or materials, an illustrative example is lithium, which is important for lithium ion batteries. These batteries are important across a very broad front, from consumer electronics to business infrastructure to military command and control systems to weapon systems themselves, and to the performance of AI-based systems. While half of all

lithium is mined in Australia, much of it comes from the so-called “lithium triangle” bordered by Argentina, Bolivia, and Chile, which together with Peru, contains about two-thirds of proven lithium reserves, and produces about half of the global lithium supply. It is a region where the PRC has been working to gain a foothold through its “Belt and Road” initiative, even building a transportation link across South America to ensure its control of exports of critical materials. In addition, essentially all of Australia’s lithium is shipped to China, which dominates the rest of the global supply chain. The PRC has positioned itself as a market leader in the various manufacturing stages of the lithium supply chain through loose environmental policies, price manipulation, and the use of state-owned enterprises given large subsidies by the government.

Beyond lithium, PRC-backed companies control most of the cobalt mines in the Democratic Republic of the Congo in conformance with its strategy of accessing and controlling key resources around the world.

Supply Chain Vulnerabilities for AI

The PRC is demonstrating its willingness to leverage its critical mineral dominance against U.S. interests, with implications for the AI supply chain. In early July 2023, the PRC Commerce Ministry announced that from August 1, 2023, it will restrict the exports of gallium and germanium – essential for the semiconductor industry. While the new restrictions do not explicitly ban exports – instead requiring exporters to first obtain a license – it is common for the PRC to use ambiguous restrictions to amount to de facto bans. China produces 60 percent of germanium and 80 percent of gallium globally. These restrictions will likely result in delays and higher costs for semiconductor manufacturing, impacting economic and strategic interests.

There is emerging concern with respect to mining of previously inaccessible minerals (especially in the Arctic), by Russia and other nations due to environmental and strategic considerations. There is a question about the applicability of the United Nations Convention on the Law of the Sea (UNCLOS) in this context as it relates to seabed mining and deep-sea mining, although this agreement has not been ratified by the U.S. Senate.

Semiconductor Supply Chains

The production of semiconductors occurs across a broad global supply chain, with different parts and processes concentrated in certain regions. The United States still leads in core intellectual property for the most advanced designs, in chip design, electronic design automation, and advanced manufacturing equipment. Such chips, like those developed by U.S. chip designers Nvidia and AMD, offer enormous computational power designed to support the specific needs and calculations of AI systems. East Asia leads in wafer fabrication, based on huge government investments and incentives to create infrastructure and workforce skills. China leads in the less skilled and less capital-intensive areas of assembly, packaging, and testing. But the PRC is investing heavily to move up and throughout the value chain.

Given global market concentration – with only a few providers of leading-edge chips – as well as U.S. leadership in this space, controlling access to these chips has been a key way to counter AI proliferation.

Restrictions on trade in advanced AI chips was implemented through the October 2022 U.S. export controls, which imposed restrictions on the PRC's ability to "obtain advanced computing chips, develop and maintain supercomputers, and manufacture advanced semiconductors." These restrictions are expected to set the PRC back years in its development of AI systems, helping to cement U.S. leadership in this strategic technology race. International engagement and collaboration with key international partners have been key to making these restrictions effective. By successfully lobbying Japan and the Netherlands to join in the U.S.-led export controls (where both countries control other strategic points of the AI chip supply chain), the United States looks to have effectively prevented other nations stepping in to take over supplying the Chinese market. Given the centrality of advanced AI to the PRC's strategic interests, the PRC will seek to evade these export controls. Strong enforcement will be essential.

In beginning to mitigate critical supply chain risks, the United States has drawn up a strategic plan to guide efforts to limit its dependence on China. The United States has also signed an agreement with Australia on energy transition, including accessing critical resources.

AI Cloud Compute: Increasing Visibility and Control

Compute refers to the computational power designed to train and run AI models and systems. While compute requires the hardware of advanced semiconductors, the intensity of compute required for AI makes it impractical for each AI designer and operator to own their own data processing center. This has meant compute has formed its own part in the AI supply chain, with entities such as Cloud Service Providers (CSPs) renting out the hardware infrastructure and access to advanced AI chips. Incentivizing AI innovators to use cloud for their compute, rather than developing their own hardware capabilities, is in U.S. interests, particularly if the data centers and CSPs are located in the United States. Centralizing AI compute power within fewer, more mature companies, and those subject to U.S. regulations, offers a useful channel for policy interventions as future AI risks emerge.

Current controls to prevent adversaries accessing the most advanced AI capabilities have left gaps, including accessing AI compute through the cloud. While Department of Commerce export controls prevent the PRC from accessing advanced AI chips in their hardware forms, PRC backed entities can still rent these capabilities through the cloud. But a blanket ban on PRC-backed entities accessing advanced AI would likely be ineffective and diminish U.S. cloud dominance, with unscrupulous actors instead seeking cloud services from other countries or resale providers.

The most significant issue is a lack of visibility: Cloud Service Providers are not currently required to identify and monitor entities using their advanced AI capabilities. Awareness of who is using significant levels of AI computational power would be beneficial in broader AI safety and security efforts, both domestically and internationally.

Introducing a KYC scheme could help improve visibility of entities seeking access to significant amounts of advanced AI computing power. Drawing on lessons learned from KYC in the financial sector, this scheme would help the U.S. government develop and apply targeted restrictions where there is significant risk to the national interest.

However, such a scheme will have a regulatory impact and risks a dampening effect on industry at a time when continued U.S. technology leadership is central to our national interests. This risk can be managed by targeting KYC requirements towards levels of AI compute that pose credible risks, with thresholds updated periodically in consultation with industry.

Using this scheme, instead of a blanket ban, will allow the United States to develop a flexible lever that can be adjusted to address changing risks. For example, as AI capabilities advance further and grow more dangerous, the U.S. government might wish to undertake greater scrutiny of a variety of actors, domestic and international, seeking to access these frontier technologies.

To implement this scheme, the Department of State should engage the Department of Commerce to leverage existing work underway to enhance the role of CSPs in customer identification and monitoring. The Trump administration's 2021 Executive Order titled, "Taking Additional Steps to Address the National Emergency with Respect to Significant Malicious Cyber-Enabled Activities" charged the Department of Commerce to introduce an obligation for providers of U.S. Infrastructure-as-a-Service to verify and record the identity of persons applying for an account, in order to combat malicious cyber activity. More recently, the 2023 National Cybersecurity Strategy commits the Administration to implementing this Executive Order. While the scope of this measure is much broader – applying to all cloud compute – there is an opportunity to build on this scheme to create a tiered approach. Light touch KYC could be implemented across the board, with rigorous, comprehensive due diligence obligations applied to just the highest levels of AI compute power, proportionate to the greatest risk.

International cooperation and alignment would significantly enhance the efficacy of U.S. actions in this arena. Implementing a KYC program unilaterally could result in perverse incentives: diminishing the attractiveness of U.S. CSPs and pushing those with greater privacy premiums (including less scrupulous entities) to lower regulatory environments. Buy-in from a broader grouping of like-minded partners could also help implement international engagement on emerging risks, trends, and suspicious activity – much like the Financial Action Taskforce does for the financial sector. To maximize international support, the Department of State could

message the opportunity of KYC initiative as a country-agnostic AI governance mechanism, thereby reducing the risk of potential members shying away from PRC retaliation.

Large, multi-use models, like those used for generative AI, could be used for significant harm in the wrong hands. The release of the leading generative AI tool ChatGPT in November 2022 by U.S. company OpenAI, powered by a multimodal LLM GPT-3, demonstrated unprecedented natural language processing, pattern recognition, information synthesis, and predictive power – something reinforced when Open AI released GPT-4. OpenAI, like most U.S.-based AI industry leaders, has a stated aim of protecting its data, models, and products and preventing misuse. But there are currently no restrictions against the export of these models. It would be to the detriment of U.S. security, as well as economic interests, should an adversarial military or intelligence service gain access to these advanced models, though other countries also have advanced AI companies that can build powerful models.

Risks will only heighten as these leading models grow in capability. Even less capable models, when trained on especially sensitive data, have demonstrated significant misuse risks. A 2022 experiment by U.S. company Collaborations Pharmaceuticals demonstrated how their AI-enabled pharmaceutical design software could be tweaked to instead discover highly toxic compounds. As models advance further, it is likely they could be used to design novel pathogens and bioweapons more broadly. Given the potentially catastrophic impacts of such AI models, it will be important to prevent them from falling into the wrong hands. Yet currently, there are no restrictions on companies choosing to export sophisticated, or sensitive, AI models. Working with the Department of Commerce, as well as other relevant agencies, industry, and AI experts to develop export license requirements for the highest risk advanced AI models is both necessary and prudent.

Such controls would need to be carefully targeted to capture only the most advanced or high-risk models, so to not inadvertently dampen AI innovation in the United States. The U.S. government should explore the plausibility of these types of measures.

Generative AI also increases the ease of generating deepfakes and disinformation. While these are not new threats, the risk of increasing scale and persuasiveness of deepfake media could

increase the challenge to public trust in government and social cohesion. While industry is already developing mechanisms to identify and counter deepfake content, there remains a risk of low public awareness. Should the prevalence of deepfakes suddenly start to increase, further action may be required.

Responsible State Behavior and Human Rights

The U.S. government seeks to promote Responsible AI (RAI), which one expert defined as "ensuring that AI systems enter into human-centric frameworks that are defined by humans to maintain human agency and responsibility." There is no consensus on what RAI entails in practice, but U.S. leaders have already articulated useful principles. These include Department of Defense guidelines for AI, remarks by senior U.S. leaders such as Under Secretary of State Bonnie Jenkins, and other statements and written instructions. These documents lay out a set of principles for RAI, stressing the need for systems to accord with applicable international law, the necessity of human judgment, ensuring that AI capabilities have explicit uses and are engineered carefully to fulfill these functions, and minimizing unintended bias. As Under Secretary Jenkins noted, "our attempts to harness the benefits of AI have to be accompanied by a focus on safe and responsible behavior that is consistent with the law of war and international humanitarian law."

The U.S. government is ahead of most other countries, in some cases far ahead, in articulating its approach to AI and embodying it in government policy documents that guide AI development, deployment, and use. Fortunately, on security-related issues, many democratic allies, including all of NATO, have issued some statements or conducted informal studies consistent with principles articulated by the DoD that AI should be responsible, equitable, traceable, auditable, reliable, and governable. Even the PRC has articulated principles on RAI that, on paper at least, are not too far from U.S. principles. Allies, however, may disagree on specific definitions of these core terms and have additional concerns that the United States must consider as it moves forward on AI.

Several aspects of AI have the potential to jeopardize or at least undermine human rights outside of the national security use context. The possibilities are numerous, but some of the most pressing include:

- Potential for surveillance. Image recognition, and the ability to rapidly track and assess large amounts of data, allows governments to implement widespread surveillance programs.
- Infringement of civil rights. AI technologies might be improperly used for criminal justice, to shape which political opinions might be expressed and how, and be vulnerable to bias against protected groups for important areas like employment and housing, among many other possibilities.
- Potential for propaganda and false information. The ability to rapidly generate fake yet realistic audio and video, and to do so at a mass scale, gives governments the ability to manipulate the perceptions of their citizens and of foreign audiences. This, in turn, can dramatically distort the information environment and potentially subvert legitimate governments.
- Risks to individual privacy. AI can threaten privacy both through the data on which it is trained and the ways in which it is used.
- Safety. A particularly important area, and one that may position the Department of State differently from other U.S. government agencies, is ensuring AI safety, both for U.S. systems and for creating global standards. AI safety risks include failures, accidents, or unexpected emergent behaviors. Human operators may also interface imperfectly with AI systems in multiple ways. One danger is simply not understanding how the system operates and its potential for errors. As a result, operators may trust the AI too much, leading to automation bias, even when it produces potentially dangerous results.

Many RAI concerns play into arms control and other tasks of the Under Secretary of State for Arms Control and International Security. Export controls may be appropriate for systems that have the potential to violate privacy and allow dangerous surveillance. Other countries might develop AI-enabled systems that could confuse or accelerate decisions in a way that makes

accidental and inadvertent escalation more likely. The United States also needs to ensure that allies' systems are safe to prevent unintended crises tied to U.S. security commitments.

The threatening aspects of AI are a concern that demand strong diplomacy, but they also present an opportunity for U.S. leadership and for negotiations with potentially hostile powers. Although the United States and the PRC are likely to disagree on many aspects of AI as they relate to national security, both have a strong interest in AI safety.

Several aspects of the current deployment of AI pose challenges for efforts to ensure responsible state behavior. These include:

- Private sector dominance. In contrast to nuclear weapons and most advanced conventional military technologies, the overwhelming majority of the development of AI systems is occurring and is likely to continue to occur in the private sector. Much of the compute power and datasets are also currently controlled by the private sector.
- Domestic Drivers. Because AI affects privacy, health care, employment, discrimination, and a host of other vital issues, it is likely that both legislation and regulation will shape its development in democratic countries.
- Dual-use technologies. Many of the technologies integral to AI, such as advanced computer chips and large datasets, are useful for both military and non-military advances. Biological and health care datasets, for example, can be used for highly beneficial pharmaceutical development, or, potentially, for designing more sophisticated biological weapons. Among such datasets are included genomic data, patient data, medical research records, etc.
- AI capabilities are advancing more rapidly than the attendant diplomatic and regulatory frameworks.
- AI companies are expanding the capabilities of their technologies, often without a commensurate expansion of safety and security capabilities.
- In the absence of robust cyber and physical security, adversaries have the opportunity to monitor and steal algorithms and data from U.S. companies.

However, the very dominance of U.S. companies in current AI deployment is favorable to U.S. efforts to encourage responsible use of artificial intelligence. The U.S. lead in AI development and deployment gives greater weight to the leading position of the U.S. government than to other governments.

The Role of Allies

To advance Responsible AI, discussions, and ideally agreement, with allies is vital. As Secretary Blinken remarked at the National Commission on AI summit in July 2021, "We need partners." Policy alignment can encourage joint research and development, ensure robust legal frameworks, and advance human rights and civil liberties issues as well as create international standards and norms to guide private sector research. A failure to engage allies, in contrast, decreases U.S. influence when pressing competitors, risks problems when allies and the United States are using AI systems in joint operations, and increases the risks of accidents.

Fortunately, the United States is in an advantageous position to promote AI in rough harmony with its most important allies – but this window will not stay open indefinitely. Therefore, it is important that the United States monitor and estimate increases in new AI capabilities through multiple direct and indirect means. Related to this is the imperative to drive global engagement on cyber and physical security, as well as safety, in AI-related work. The rapid advances in AI and machine learning are creating both excitement and concern, while no existing approach has gelled: a rare combination of high-level attention and opportunity.

Allies, of course, have their own interests and foreign policy preferences and thus their own views of what a future AI-powered world should look like. Even when allies agree in general on principles, they often single out different issues or approach regulation in different ways. France, for example, appears more concerned about bias and that humans may not trust AI systems whose results they do not understand and also worries that it will depend on technology where the United States (and the PRC) have the lead, essentially forcing France to rely on foreign technology it cannot control for its sovereignty and survival. Some allies seek Chinese markets and investment, and the PRC has invested in some of their AI companies.

Privacy is a particularly difficult issue. The United States and many key democratic allies have different standards for privacy, with many European countries having far stricter standards for both private and government use of personal data. Germany, for example, opposes many aspects of biometric recognition. In contrast, some allies may seek to sell AI-enabled systems that pose significant privacy and surveillance risks.

Negotiating with Competitors and Adversaries

Although countries like the PRC and Russia have many goals that are opposed to U.S. interests, it is still important to engage them selectively on AI-related topics, when possible, particularly those related to stability and crisis management. There is a long history of adversaries cooperating in peacetime to ensure safety. During the Cold War, the United States offered to share permissive action links with the Soviet Union to prevent an unauthorized use of nuclear weapons.

AI safety is one mutual concern: no one wants an emergent AI to set off a confrontation or for AI to push a rapid rate of decision making that prevents de-escalation. Some might worry that the fear of falling behind would lead to the deployment of AI-enabled military systems too quickly. Alternatively, concerns about safety and reliability could lead to algorithm aversion and militaries holding back on deploying AI-enabled systems.

Confidence building measures for AI systems may also be of interest. Adversaries too may feel pressure to deploy systems that are not fully tested even as they recognize the potential dangers because they fear U.S. systems would otherwise overwhelm them. In addition, adversaries may seek limits on systems where the U.S. has advantages. Similarly, there may be areas such as offensive cyber operations or the targeting of adversary decision-making systems that carry a high risk of inadvertent escalation and accidents where a commitment to limit or avoid certain types of activity would be beneficial. A particularly important area is to limit the role of AI in nuclear command and control to prevent emergent behavior or overly rapid decision making from creating or worsening a crisis, which is consistent with existing DoD policy.

Negotiations with competitors will, of course, face many barriers. Inherent suspicions on all sides will make progress difficult, even on basic areas. Competitors' definitions of key AI terms may differ considerably from those of the United States and its allies. For example, the PRC's definition of "safety" includes the impact of a technology on the regime's power.

AI's potential abuse on human rights grounds will be an issue of disagreement. Already, countries like the PRC are using AI-powered facial and voice recognition systems, predictive policing, and other tools that are part of the comprehensive surveillance state the country has established, with AI systems used in Xinjiang and other regions. The PRC is also a major exporter of surveillance systems, including ones powered in part by AI.

At times negotiations will fail but provide diplomatic opportunities to the United States. For example, if the PRC and Russia refuse to sign protocols prohibiting AI's use for extensive surveillance or that require humans be in the loop for nuclear authorization, it would highlight the positive U.S. position on these issues.

The PRC, however, has so far resisted U.S. outreach. The DoD has tried to foster dialogue with the PRC military on AI risk reduction, but has been refused. The Department of State, in coordination with other government agencies, should identify the appropriate venues, whether multilateral or bilateral, for AI negotiations with the PRC.

The Department of State should initiate Track 2 dialogues on AI safety and other Responsible AI issues with the PRC and other potential competitors if possible. These might focus on issues such as reward hacking, robustness, and verification and validation. Private sector AI leaders have indicated a willingness to participate, particularly if encouraged by the U.S. government. The Department should also prepare for shifting from Track 2 to Track 1.5 should there be progress, with formal negotiations being the long-term goal.

Private Sector Cooperation

The Department of State, ideally with its allies, should also work with the private sector and academic research community in the United States and in allied countries on AI-related issues. This stems from simple necessity: most of the AI-related innovation is in the private sector, and

the products and talent are outside government ranks. In addition, and in contrast to other defense-oriented fields, many leading AI researchers have little experience working with the U.S. government, much of their work focus is on business imperatives or more general research. Finally, large private companies control the large datasets and immense computing power that enable advances in AI research.

Many leading companies are U.S. based, and many of their top officials support U.S. goals and share American values, but this broad commonality will at times be in conflict with basic issues of profit and loss. Making this more complex, many of the companies involved are global. As such, they have work forces and sensibilities that do not always reflect the goals of the U.S. government. This may make some companies reluctant to work on defense-related or similar programs. In many cases, some of the leading AI engineers at global companies like Google or leading universities will not be U.S. citizens, and some may be nationals of competitor nations, like the PRC. Even when they are U.S. citizens, leading experts are likely to be engaged in collaborative research with nationals from other countries, including (indeed, especially) the PRC. A failure to engage in such collaborative research would be potentially harmful to U.S. interests, decreasing U.S. research and overall knowledge.

Safety will often prove an easier conversation than topics more explicitly tied to national security, despite their overlap. For brand and liability reasons, companies have a strong interest in improving AI safety, and their research and products can assist government-oriented efforts.

At least some AI-focused companies are willing to work with the U.S. government, sharing expertise and ideas. In our interviews, several leading companies expressed a willingness to work with the Department of State. In addition, they often seek government guidance as they develop their standards and prefer government cover when they join Track 2 or other engagements with countries like the PRC.

Norms and Procedures for AI Governance

Confidence building measures represent a promising potential arena to build international cooperation surrounding artificial intelligence. Discussion on issues such as testing and evaluation standards, rules for verification and validation, and sharing of information on AI accidents can be done through discussions with the European Union and the Organization for Economic Co-operation and Development, but it will also require engagement with more technical bodies such as the International Organization for Standardization, among many others. Indeed, there is reason to move forward quickly. The PRC is putting forward a set of rules governing AI development, deepfakes, and other AI uses, and the United States should not allow the PRC to set the pace and shape the pathways for AI governance and development.

Norms will not solve many RAI-related problems, but they can help, particularly when attempting to speak with one voice with allies. For example, democratic governments might agree that deepfakes should not be used in ways that could lead to international conflict. The United States and its allies might also agree not to disrupt power grids, broadband networks, financial networks, or medical facilities via AI-powered cyber weapons.

The Department of State should work with other U.S. government agencies and lawmakers to develop international organizations and procedures for better AI governance. Some governance measures might spell out minimum procedures to ensure RAI, while others might involve processes to balance competing security and RAI concerns.

Because of the many safety risks, sharing data on AI-related accidents is vital. Some experts recommend as models the National Transportation Safety Board database for aviation accidents and the public-private Information Sharing and Analysis Centers that share cyber intelligence.[1]

[1] Zachary Arnold and Helen Toner, “AI Accidents: An Emerging Threat,” Center for Security and Emerging Technology, (2021), p. 17.

The Vulnerabilities Equity Process, which the U.S. government developed to manage its response to the discovery of zero-day cybersecurity vulnerabilities, aims to ensure that vulnerabilities are leveraged only when there is a compelling reason to do so, with a large interagency process with various equities considered in the discussion. Such an approach might be used for deepfakes and other AI-enabled information operations, which have considerable human rights implications.

There seems to be no urgency in sorting out data sharing approaches with allies – especially in Europe. The United States is missing many opportunities to share data for important objectives with allies owing to the absence of data sharing and protection norms in the United States. As Europe and other nations proceed with their rule making and data implementation – like the General Data Protection Regulation (GDPR), and others – these allies are hesitant, restricted, or are forbidden to work with the United States on AI projects that entail sharing of large health databases or other sensitive data given our own lack of a more structured approach. The Secretary of State should charge the relevant parties, bureaus, and offices, in the Department, to develop norms and standards for data sharing with our allies.

Opportunities and Applications in Intelligence

The Under Secretary for Arms Control and International Security is seeking to fortify arms control, nonproliferation, disarmament, and related activities. In the current tense geopolitical environment, this requires developing new tools to enhance the intelligence community's ability to detect behavior of interest with the assistance of AI and ML. A desired end state in this area is that the United States can enter into arms control agreements that advance national priorities with less concern about negotiation of intrusive monitoring regimes because enhanced intelligence tools will provide high confidence that violations of potential agreements will be detected.

To date, the Study Group has identified only limited current application of AI tools to the arms control mission within the Department of State. In part, this is due to the traditional arms control mission being focused on monitoring of major military assets specified for accountability under treaties that were themselves designed to be verifiable using national

technical means available at the time of negotiation. However, with the New START Treaty entering its final years, and with very limited arms control structure governing U.S.-PRC relations, the ADS Bureau problem set is likely to shift to more risk reduction and threat management, where more advanced technologies for monitoring are likely to be valuable.

Although the U.S. defense industry has been demonstrating capability to harness ML for functions relevant to Department of State missions like large scale target recognition, it is not apparent that the Department has yet articulated a need for improved capability to support future ADS Bureau problems to the IC. It is likely that key attributes of the requirements set for ADS Bureau missions overlap partially but not completely with those of the IC's strategic warning mission, and similarly that there is only a partial overlap with the tactical reconnaissance needs of the military. The former mission likely requires lower confidence, while the latter likely demands much more rapid processing to meet tactical timelines. If the Department of State needs in this area are not fully congruent with other U.S. government actors', a failure to fully understand and model them will leave the ADS Bureau's mission at a disadvantage.

Thus far, the Department of State appears to make limited use of intelligence collection and analysis generated with AI/ML tools. The whole of the Department can become a better customer for AI enabled technical intelligence including in support of the Under Secretary mission set. Key opportunities are available for compliance monitoring in existing arms control arrangements, but likely even better applications exist for risk reduction relative to less-well-controlled relationships in the context of strategic competition with the PRC.

This opportunity is of increasing importance because the Department of State has much less hands-on insight into PRC strategic programs due to the lack of a bilateral arms control treaty with in-person monitoring, less overall history operating with their strategic programs due to their comparatively recent ramp up, and the general opacity of PRC military activities. Preparatory to any potential future agreement, and for the purposes of risk reduction activities in the absence of a U.S.-PRC agreement, the Department would benefit from having a more well-understood baseline of PRC programs, activities, and patterns of life. Technical insight into novel programs would also be valuable.

Advanced sensors may offer opportunities to identify signals of weapons development and to expose proliferation activities, particularly with cross-sensor and cross-phenomenology fusion. Space sensing companies have already demonstrated the applicability of multi-intelligence tracking supported by ML for detection of human trafficking activity, which shares some signals with proliferation. Improved ability to detect proliferation signals is particularly valuable as growing focus on achieving climate objectives leads to increased investment in nuclear power. Ensuring that these activities are compliant with both the proliferation and the peaceful uses elements of the NPT will be increasingly important.

Ideally the United States would be able to collect intelligence information to predict malevolent state behavior before it occurs (i.e.., indications and warning). Sometimes these signals may be buried in noise, especially for undesired activity below the level of a full-scale invasion or major nuclear weapons program. However, even for these harder predictive tasks, ML may offer useful investigative tools for attributing bad actions after they occur. Knowing that undesired activity X has occurred, identifying signals of which actor may have been involved is a notably easier problem than advance prediction of an undesired event based on all the available signals.

The nature of potential future strategic competition between the United States and the PRC could vary widely – the Department of State should be preparing simultaneously for multiple potential futures that capture everything from a re-opening to a post-conflict environment. It would be wise to explore structured assessments of potential future states, identification of what is likely needed to achieve the end state (and what early signals of that might be), and explore structured assessments of potential future states, including identification of measures to achieve the end states, early indicators, and contributions and intersections of multiple technologies.

Capacity Building and Human Capital

One challenge for the U.S. government in general, and for the Department of State in particular, is ensuring that it has personnel who understand the latest developments in artificial intelligence and other emerging technologies and to incorporate this understanding into daily

and long-term Department operations. AI is rapidly advancing, and the Department of State needs informed personnel who can adjust or even radically change current policies as new developments emerge.

Much of the talent and innovation in AI is in the private sector, and it is often concentrated among younger professionals, including many in or just out of university. In addition, the salary disparity between the private sector and government is particularly immense in this area. As a result, it is often difficult to attract and retain the top technological talent and for the Department of State and other U.S. government agencies to be aware of the cutting edge of various technologies.

No single approach will ensure the proper level of expertise within the Department of State: a range of approaches is necessary. An important approach is to bring together offices in the Department who currently work on advanced science and technology topics, especially those under the purview of the Under Secretary for Arms Control and International security and the Under Secretary for Economic Growth, Energy, and the Environment.

Other approaches to strengthening the Department of State scientific and technological capacity involve hiring and other personnel procedures at the Department itself and greater coordination with other departments and outside entities.

One step is to prioritize technological talent in hiring. This requires resourcing hiring and personnel offices to ensure they have the necessary capabilities. When it hires foreign service officers and civil servants, the Department of State emphasizes a range of skills, such as English expression and situational judgment. Technological knowledge could be emphasized more explicitly in the screening process; a technological background could be prioritized for hiring. Ideally, all new professionals would have a stronger technological background in general, and a significant subset would have enough expertise to work on technological issues in more detail.

The Department of State must also foster longer-term internships to take advantage of emerging talent. The Department internships currently occur during the summer or one semester. Because of the short duration of these programs, students often can make only limited contributions, as they spend much of the time familiarizing themselves with the

Department. Such programs could be expanded to include longer internships (six to nine months) prioritizing students with technological skills. A longer internship would enable the participants to make a far greater contribution to the Department of State and give the Department access to emerging professionals with impressive technological skills.

The Department of State needs to place greater emphasis on retaining technological talent. Because of the salary differential with the private sector, the Department, like other government agencies, is also vulnerable to disproportionate losses of technological talent. Those with technical skills could receive bonuses, accelerated promotion, and other benefits in order to retain them.

Expanding rotations of outside experts into the Department of State is also vital. The Department currently benefits from the rotation of experts from academia. The Department of State could expand the size of existing programs that target experts from universities, non-profits, and national laboratories. The Department also could consider reprising prior programs of this nature if warranted. The Department could also explore the use of special hiring authorities for rotational assignments of technical experts such as those employed by the Defense Advanced Research Projects Agency (DARPA) and some of the newer Congressionally created Advanced Research Projects Agencies such as Advanced Research Projects Agency-Energy (ARPA-E), Advanced Research Projects Agency for Health (ARPA-H), and Advanced Research Projects Agency-Infrastructure (ARPA-I).

The Department of State also might explore programs that allow private company officials with AI expertise to participate in decision making on export controls and negotiations, subject to appropriate controls and oversight with respect to potential or perceived conflicts of interest, the need to protect pre-decisional and diplomatic information, potential counter-intelligence targeting, etc. For example, the Center for Drug Evaluation and Research at the Food and Drug Administration (FDA) allows for private partners to have a staff presence at FDA, enabling them to better understand the regulatory and approval processes and allowing the government to understand private sector concerns.

The Department of State must ensure robust procedures to tap into other government agencies. Some parts of the U.S. government have many personnel with excellent technological skills. In addition to their nuclear expertise, the DoE National Laboratories and DoD DTRA have a huge workforce with many experts in artificial intelligence, data science, and biology, among others important areas. The Department of State could develop liaisons, working groups, rotations, and other means of accessing the National Laboratories and other government experts outside the Department.

The Department of State should identify individual offices and officers with responsibility for engaging with the leading AI companies. The companies should also be encouraged to assign a set of people to engage with the Department to ensure regular interactions.

Most ambitiously, the Department of State might consider a reserve system for professionals with technical expertise. A more ambitious program would be to develop a "diplomatic reserve" program that is (very) loosely modeled on the U.S. Army Reserve (and even more so its specialized units, such as the Cyber Protection Brigade). Professionals with valuable experience in the private sector could sign on to serve in a part-time capacity with the Department of State, working on relevant issues for a few days, a month, or a longer, concentrated period once a year. They could also be surged during a crisis. The Department of State officials with valuable expertise who leave the Department could also join such a Reserve. Reserve members would receive limited pay and other benefits, and it would enable them to do public service in a limited capacity even as they continue in the private sector. It would also infuse the values and concerns of the Department more broadly through the private sector, helping drive innovation in a government-friendly way.

Conclusion

In this new era of global competition, robust scientific and technological innovations are foundational for both national security and economic security. Strategies for international engagement with both allies and partners as well as with non-allies need updating to ensure U.S. strength and scientific prowess. Updated and productive public-private partnerships are also needed.

The Department of State needs to develop a strategy to enable success in strategic competition with the PRC and in response to other challenges. The strategy should employ a comprehensive, holistic, and integrated approach to international engagement that explicitly recognizes the dual-use nature of AI and associated technologies.

A part of an updated technology competition strategy should include better harmonization of science and technology policies and practices between open and restricted science and technology applications, and updated collaborations within the Department of State, and among the Department and other parts of the U.S. government.

All of this requires greater capacity building within the Department of State, including human capital development and partnerships with other departments and agencies with specialized expertise. A new and robust partnership with the technology industry is needed to keep up with innovations with dual-use capabilities that may create national security concerns, discern risks associated with them, develop risk-informed confidence building measures, and lead in promulgating sensible regulations and norms to balance risks and opportunities globally.

Appendix A – Terms of Reference

UNDER SECRETARY OF STATE FOR
ARMS CONTROL AND INTERNATIONAL SECURITY
WASHINGTON

October 18, 2022

MEMORANDUM FOR THE CHAIRMAN, INTERNATIONAL SECURITY ADVISORY BOARD (ISAB)

SUBJECT: Terms of Reference - ISAB Study on the Impact of Artificial Intelligence and Associated Technologies on Arms Control, Nonproliferation, and Verification

The International Security Advisory Board (ISAB) is requested to undertake a study to advise the United States on how artificial intelligence and associated technologies (hereinafter referred to as "AI") may impact arms control, nonproliferation, and verification, noting both the risks and benefits from its application.

> *"If AI and emerging technologies are] going to be used as part of our national defense, we want the world to have a shared understanding of how to do that responsibly, in the same way that we've hammered out rules for how to use conventional and nuclear weapons. That's how we reduce the risk of proliferation. It's how we prevent escalation or unintended incidents."*
>
> - Antony Blinken, Secretary of State, July 2021

Emerging technologies represent a wide range of evolutionary as well as disruptive innovations that have national security relevance. As the 2022 National Security Strategy states, "emerging technologies [are] transform[ing] warfare and pose novel threats to the United States and our allies and partners." Technological development will play a critical role in defining the national security posture and competitive position of the United States. Emerging technologies also present new opportunities that can assist the U.S. government with issues that relate to national security, but these opportunities are not without risks.

Artificial intelligence (AI), is of particular interest, given its potential to transform decision-making and national security capabilities. Per the National Artificial Intelligence Act of 2020, "artificial intelligence" refers to "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments." The speed of technological development is generating both enthusiasm for AI' s benefits and alarm over the potential for misuse.

The military and national security applications of AI, particularly related to nuclear weapons, are generating attention within government and other interested parties. Applications of machine learning algorithms to issues of arms control, nonproliferation, and verification could pave the way for innovative solutions within the field of nuclear policy. Emerging data science methods and advanced analytical tools can expose proliferation activities and can be a useful tool for U.S. government programs on nonproliferation, particularly in areas related to existing strategic trade controls on dual-use goods, determining the origin of illicit material or items, or detecting potential violations.

There also are questions about how AI could impact crisis stability and escalation, decision making, command and control, communication, and the verification of arms control measures and agreements. The rapid evolution of AI capabilities raises the concern that a focus on near-term applications and risks might result in the United States failing to anticipate long-term destabilizing impacts or developments that leave the United States at a strategic disadvantage or with ineffective technology. Some experts have raised concerns that applying AI to decision-making processes could lead to inadvertent escalation between nuclear powers or other states. The fast-paced development and implementation of new AI technologies is driving the discussion of the establishment of national AI ethical standards to encourage responsible state behavior and policies to realize its benefits while discouraging detrimental consequences. Other discussions concern how export controls and investment screening might be utilized to curb our adversaries' access to U.S. goods and technologies that could provide them military advantages or result in other destabilizing national security consequences.

Another challenge that the Department faces is retaining, attracting, or having access to the talent needed to understand the trajectory of AI development and its potential applications, highlighting the need to build partnerships with industry and academia. As of now, discussion and policy framework focused on the use cases, limitations, and capabilities of AI in nuclear and other military technology is still emergent. The Department needs the right technical and subject matter expertise to understand and successfully manage potential risks, and pursue potential opportunities derived from AI applications.

It would be of great assistance if the ISAB conducts a study on how artificial intelligence is impacting, and will impact, international security, noting both the risks and benefits from its application.

- Provide an assessment of the current state of AI application to international security, nonproliferation, and arms control missions within the Department of State and expected new capabilities in the coming decade.
- Identify potential interactions between AI and other emerging technologies (e.g., biotechnology, quantum information science) that might impact international security, arms control, and proliferation risks.
- Identify potential risks and opportunities from the application of AI in the military domain in a manner that affects strategic stability and nonproliferation including, but not limited to, nuclear operations.
- Identify avenues that the U.S government should explore to help build international norms of responsible state behavior and risk reduction measures related to AI.
- Identify supply chain chokepoints that would limit our adversaries' access to AI-enabling technologies.
- Identify ways AI can be used to enhance arms control and nonproliferation, specifically ways it can detect proliferation risks and/or enhance verification.

- Identify barriers to the successful application of AI technologies to arms control, nonproliferation, and verification, and how such obstacles can be addressed.
- Review the Department's acquisition and development strategies for limitations or gaps that would inhibit the Department from procuring the necessary attributes that are needed in this fast-moving technology for the acquisition of relevant data and training of personnel.

In the conduct of its study, as it deems necessary, the ISAB may expand upon the tasks listed above. I request that you complete the study in 180 days. Completed work should be submitted to the ISAB Executive Directorate no later than April 2023.

The Under Secretary of State of State for Arms Control and International Security will sponsor the study. The Assistant Secretary for Arms Control, Verification and Compliance will support the study. Anne Choi will serve as the Executive Secretary for the study and Michelle Dover will represent the ISAB Executive Directorate. Angela Sheffield will provide support as a subject matter expert.

The study will be conducted in accordance with the provisions of P.L. 92-463, the "Federal Advisory Board Committee Act." If the ISAB establishes a working group to assist in its study, the working group must present its report or findings to the full ISAB for consideration in a formal meeting, prior to presenting the report or findings to the Department.

Bonnie D. Jenkins

Appendix B – Members & Project Staff

Board Members

- Hon. Edwin Dorn (Chair)
- Ms. Sherri Goodman (Vice Chair)
- Dr. Daniel Byman
- Hon. Patricia Falcone
- Dr. Julie Fischer
- Dr. James Goldgeier
- Dr. Gigi Kwik Gronvall
- Dr. Gregory Hall
- ADM Cecil Haney, USN (ret.)
- Dr. Eboni Haynes
- Ms. Julie Herr
- Dr. Michael Horowitz
- Ms. Heather Hurlburt
- Hon. Shirley Ann Jackson
- Amb. (ret.) Laura Kennedy
- Dr. Susan Koch
- Dr. Edward Levine
- Dr. Jeffrey Lewis
- Hon. Jamie Morin
- Hon. Eric Rosenbach
- Dr. Ian Simon
- Ms. Lyric Thompson
- Dr. Paul Walker
- Dr. Heather Williams
- Mr. Jon Wolfsthal

Study Group Members

- Hon. Shirley Ann Jackson (Chair)
- Dr. Daniel Byman
- Hon. Patricia Falcone
- Dr. Gigi Kwik Gronvall
- Dr. Michael Horowitz
- Hon. Jamie Morin
- Hon. Eric Rosenbach
- Dr. Ian Simon

Project Staff

- Ms. Anne Choi, Executive Secretary
- Ms. Angela Sheffield, Subject Matter Expert
- Ms. Michelle Dover, Executive Director, ISAB
- Mr. Scott Bohn, Deputy Executive Director, ISAB
- Ms. Thelma Jenkins-Anthony, Senior Advisor, ISAB

Appendix C – Individuals Consulted by the Study Group

December 16, 2022

Mr. Eric Desautels	Bureau of Arms Control, Verification and Compliance, State
Expert	Bureau of Political-Military Affairs, State
Expert	Office of the Legal Advisor, State
Experts	Bureau of Oceans and International Environment and Scientific Affairs, State
Expert	Bureau of European and Eurasian Affairs, State
Experts	Bureau of International Security and Nonproliferation, State
Experts	Intelligence Community
Dr. Allison Schwier	Acting Science and Technology Advisor to the Secretary, Office of the Science and Technology Adviser, State

February 2, 2023

Expert	Department of Defense
Experts	Intelligence Community
Experts	National Nuclear Security Administration, DoE
Ms. Gisele Irola	Foreign Affairs Officer, Bureau of International Security and Nonproliferation, State
Mr. Wayne Mei	National Nuclear Security Administration, DoE

Mr. Shawn Steen	Senior Policy Advisor, Force Development and Emerging Capabilities Office, Office of the Secretary of Defense (Policy), DoD

February 27, 2023

Dr. Anthony Bak	Palantir
Dr. Sarah Shoker	OpenAI

March 8, 2023

Mr. Eric Desautels	Bureau of Arms Control, Verification and Compliance, State
Experts	Office of the Special Envoy for Critical and Emerging Technology, State
Mr. Chris McGuire	National Security Council
Mr. Dan Oates	Global Technology Policy Advisor, Bureau of Cyberspace and Digital Policy, State
Mr. Corrie Robb	Senior Policy Advisor, Bureau of Cyberspace and Digital Policy, State

March 23, 2023

Experts	Office of the Special Envoy for Critical and Emerging Technology, State
Ms. Andreea Paulopol	Physical Scientist, Bureau of Arms Control, Verification and Compliance, State

March 28, 2023

Experts	Intelligence Community
Dr. Ben Buchanan	National Security Council

May 8, 2023

Mr. José E. Colón	Technical Advisor, Bureau of International Security and Nonproliferation, State
Ms. Daniela Cooper	Bureau of International Security and Nonproliferation, State
Experts	Bureau of International Security and Nonproliferation, State
Dr. Gopal Sarma	Technologies Office, Defense Advanced Research Projects Agency
Dr. Katherine Sixt	Office of the Under Secretary of Defense for Research and Engineering, DoD

May 17, 2023

Dr. Jonathan Dordick	Institute Professor of Departments of Chemical and Biological Engineering, Biomedical Engineering, and Biological Sciences, Rensselaer Polytechnic Institute

June 6, 2023

Dr. Philip Bingham	Division Director, Oak Ridge National Laboratory
Experts	Bureau of Global Talent Management, State
Experts	Sandia National Laboratories
Expert	National Nuclear Security Administration, DoE
Expert	Los Alamos National Laboratory
Expert	Lawrence Livermore National Laboratory
Expert	Pacific Northwest National Laboratory
Ms. Katie Gibb	National Security Advisor, Pacific Northwest National Laboratory

Mr. Michael Goldman	Associate Program Leader in Global Security, Lawrence Livermore National Laboratory
Ms. Amanda Johnson	Bureau of Global Talent Management, State
Mr. Christoper Park	Director, Biological Policy Staff, Bureau of International Security and Nonproliferation, State

June 16, 2023

Dr. Kimberely Sablon	Principal Director for Trusted Artificial Intelligence and Autonomy, Office of the Assistant Secretary of Defense for Critical Technologies, U.S. Department of Defense
Dr. Brian Spears	Director, AI Innovation Incubator (AI3), Lawrence Livermore National Laboratory
Dr. Michael Schneider	Associate Program Leader, Decision Superiority Laboratory, Lawrence Livermore National Laboratory

June 24, 2023

Expert	Bureau of International Security and Nonproliferation, State
Dr. Herbert Lin	Senior Research Scholar and Hank Holland Fellow, Stanford University

September 29, 2023

Dr. Samantha Anderson	Bureau of International Security and Nonproliferation, State
Ms. Daniela Cooper	Bureau of International Security and Nonproliferation, State
Expert	Bureau of Political-Military Affairs, State
Expert	Office of the Legal Advisor, State
Expert	Office of the Special Envoy for Critical and Emerging Technology, State

Ms. Alexis Frisbie	Senir Technical Advisor, Global Engagement Center, State
Mr. Michael Garth	Bureau of International Security and Nonproliferation, State
Ms. Carrie Goux	Senior Advisor, Global Engagement Center, State
Mr. David Phillips	Deputy Director, Analytics and Research, Global Engagement Center, State
Mr. Douglas Weinstein	Senior Technology Strategist, Global Engagement Center, State