Second Quarter Recommendations

Quarterly Series, No. 2

NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

Commissioners

DR. ERIC SCHMIDT
Chairman

HON. ROBERT O. WORK
Vice Chairman

SAFRA CATZ

DR. STEVE CHIEN

HON. MIGNON CLYBURN

CHRISTOPHER DARBY

DR. KENNETH FORD

DR. JOSÉ-MARIE GRIFFITHS

DR. ERIC HORVITZ

ANDREW JASSY

GILMAN LOUIE

DR. WILLIAM MARK

DR. JASON MATHENY

HON. KATHARINA MCFARLAND

DR. ANDREW MOORE

Letter from the Commissioners

As it enters the third decade of the 21st century, the United States finds itself confronted by geopolitical, economic, ideological, technological, and military challenges—all at once. Artificial intelligence (AI) and its associated transformative technologies are central to meeting the demands of all of these challenges, and will help the United States navigate today's turmoil towards a healthier and more secure future.

Against this backdrop, Congress established the National Security Commission on Artificial Intelligence (NSCAI) in 2018. The United States Government must organize, resource, and train to understand, develop, and employ AI-enabled technologies. It must do so ethically, responsibly, and in close partnership with the private sector, academia, non-governmental organizations, and its international partners. In the context of recent events, excitement about the potential for AI to improve lives has increased in parallel with concerns about the danger of AI being misapplied or used for malicious purposes.

A Dynamic Approach: The Urgency of Today and the Work of a Generation

The Commission is pursuing a dynamic approach as it moves toward publishing its final report in March 2021. It is assessing and making recommendations about a technology in motion within a rapidly shifting global environment. Scientists, innovators, and government officials are still developing, seeking to understand, adopting, and establishing governing principles for AI-enabled technologies in all areas, including for national security purposes. We are trying to imagine a future altered by technologies that in some cases have not yet arrived. We are trying to build ethical guidelines while many of the implications remain hypothetical, not yet real. We are trying to separate hype from reality in how AI will be used and misused.

Last November, the NSCAI released an interim report articulating the overarching principles guiding our work and framing a research agenda for developing concrete recommendations for the legislative and executive branches to consider. In March, the Commission released a first set of quarterly recommendations.

Developing, adopting, and protecting AI advantages requires an expansive vision for promoting America's AI leadership. Successful and responsible adoption of AI requires more than technical progress. AI developments must progress in tandem with a larger reorientation of national security departments to compete in a world shaped by strategic competition. The Commission believes the national security challenge is urgent, but it recognizes that vision for dramatic change will take time to translate into action. Many of the ideas the Commission is developing will require consensus building and hard policy engineering. Re-imagining a digital workforce, overcoming ingrained bureaucracy, developing new operating concepts, and

synching visions with plans, strategies, organization, and action—that is the work of a generation. But it must begin now.

The NSCAI's second quarterly memo of 2020 is a compendium of recommendations that balance the urgency of the challenge with the recognition that the ambitious actions required to address it will take time. The recommendations are not a comprehensive follow-up to the interim report or first quarter memorandum. They do not cover all areas that will be included in the final report. This memo spells out recommendations that can inform ongoing deliberations tied to policy, budget, and legislative calendars. But it also introduces recommendations designed to build a new framework for pivoting national security for the AI era.

Each Tab of this document can stand alone as a discrete memo on a specific dimension of the AI-national security nexus. The Commissioners believe these recommendations are solidly grounded in analysis and ready for discussion with stakeholders and the general public. While the NSCAI does not anticipate major deviations from the proposals or the underlying assessments, the Commission will adjust as any new information comes to our attention, and we will render our final recommendations in March 2021 with the most up-to-date information available at that time.

Quarter 2 Recommendations

In the second quarter, the Commission has focused its analysis and recommendations on six areas:

- **Advancing the Department of Defense's internal AI research and development capabilities.** The Department of Defense (DoD) must make reforms to the management of its research and development (R&D) ecosystem to enable the speed and agility needed to harness the potential of AI and other emerging technologies. To equip the R&D enterprise, the NSCAI recommends creating an AI software repository; improving agency-wide authorized use and sharing of software, components, and infrastructure; creating an AI data catalog; and expanding funding authorities to support DoD laboratories. DoD must also strengthen AI Test and Evaluation, Verification and Validation capabilities by developing an AI testing framework, creating tools to stand up new AI testbeds, and using partnered laboratories to test market and market-ready AI solutions. To optimize the transition from technological breakthroughs to application in the field, Congress and DoD need to reimagine how science and technology programs are budgeted to allow for agile development, and adopt the model of multi-stakeholder and multi-disciplinary development teams. Furthermore, DoD

should encourage labs to collaborate by building open innovation models and a R&D database.

- **Accelerating AI applications for national security and defense.** DoD must have enduring means to identify, prioritize, and resource the AI-enabled applications necessary to fight and win. To meet this challenge, the NSCAI recommends that DoD produce a classified Technology Annex to the National Defense Strategy that outlines a clear plan for pursuing disruptive technologies that address specific operational challenges. We also recommend establishing mechanisms for tactical experimentation, including by integrating AI-enabled technologies into exercises and wargames, to ensure technical capabilities meet mission and operator needs. On the business side, DoD should develop a list of core administrative functions most amenable to AI solutions and incentivize the adoption of commercially available AI tools.

- **Bridging the technology talent gap in government.** The United States government must fundamentally re-imagine the way it recruits and builds a digital workforce. The Commission envisions a government-wide effort to build its digital talent base through a multi-prong approach, including: 1) the establishment of a National Reserve Digital Corps that will bring private sector talent into public service part-time; 2) the expansion of technology scholarship for service programs; and, 3) the creation of a national digital service academy for growing federal technology talent from the ground up.

- **Protecting AI advantages for national security through the discriminate use of export controls and investment screening.** The United States must protect the national security sensitive elements of AI and other critical emerging technologies from foreign competitors, while ensuring that such efforts do not undercut U.S. investment and innovation. The Commission proposes that the President issue an Executive Order that outlines four principles to inform U.S. technology protection policies for export controls and investment screening, enhance the capacity of U.S. regulatory agencies in analyzing emerging technologies, and expedite the implementation of recent export control and investment screening reform legislation. Additionally, the Commission recommends prioritizing the application of export controls to hardware over other areas of AI-related technology. In practice, this requires working with key allies to control the supply of specific semiconductor manufacturing equipment critical to AI while simultaneously revitalizing the U.S. semiconductor industry and building the technology protection regulatory capacity of like-minded

partners. Finally, the Commission recommends focusing the Committee on Foreign Investment in the United States (CFIUS) on preventing the transfer of technologies that create national security risks. This includes a legislative proposal granting the Department of the Treasury the authority to propose regulations for notice and public comment to mandate CFIUS filings for investments into AI and other sensitive technologies from China, Russia and other countries of special concern. The Commission's recommendations would also exempt trusted allies and create fast tracks for vetted investors.

- **Reorienting the Department of State for great power competition in the digital age.** Competitive diplomacy in AI and emerging technology arenas is a strategic imperative in an era of great power competition. Department of State personnel must have the organization, knowledge, and resources to advocate for American interests at the intersection of technology, security, economic interests, and democratic values. To strengthen the link between great power competition strategy, organization, foreign policy planning, and AI, the Department of State should create a Strategic Innovation and Technology Council as a dedicated forum for senior leaders to coordinate strategy and a Bureau of Cyberspace Security and Emerging Technology, which the Department has already proposed, to serve as a focal point and champion for security challenges associated with emerging technologies. To strengthen the integration of emerging technology and diplomacy, the Department of State should also enhance its presence and expertise in major tech hubs and expand training on AI and emerging technology for personnel at all levels across professional areas. Congress should conduct hearings to assess the Department's posture and progress in reorienting to address emerging technology competition.

- **Creating a framework for the ethical and responsible development and fielding of AI.** Agencies need practical guidance for implementing commonly agreed upon AI principles, and a more comprehensive strategy to develop and field AI ethically and responsibly. The NSCAI proposes a "Key Considerations" paradigm for agencies to implement that will help translate broad principles into concrete actions.

As the Commission moves toward a final report in March 2021, we will continue to solicit feedback from a diverse range of non-governmental organizations, businesses, scientists, and government officials, and work closely with our partners in the executive and legislative branches. The Commission is committed to driving changes that will maximize the role of AI in protecting U.S. security, extending American leadership in emerging technologies, and strengthening our core values.

Contents

TAB 1 — Accelerate AI R&D Across the DoD Research Enterprise

The Department of Defense (DoD) research enterprise encompasses a powerful and unique array of resources.[1] These research institutions have long been drivers of competitive advantage for the U.S. military, and engines of innovation for technologies that have transformed the U.S. economy and American society. However, outdated processes, funding policies, and organizational cultures limit the ability of these institutions to innovate at the pace of today's technological advances.[2] In the Commission's Interim Report, we found that bureaucratic and resource constraints are hindering government-affiliated labs and research centers from reaching their potential in AI research and development (R&D).[3] In our first quarter recommendations, we indicated that we would provide actionable recommendations to optimize the DoD research enterprise for AI R&D to enable strategic research investments and accelerate development and fielding of AI capabilities.[4]

To harness the potential of this enterprise to build and integrate the technologies that could transform U.S. forces and underpin their future competitive advantage, DoD must responsibly prioritize speed and agility, balancing incremental and disruptive research efforts. It must foster a culture of innovation that brings new capabilities to warfighters and their support organizations more rapidly, and involves end users in

---

[1] The Department's 63 owned laboratories cover the full lifecycle of research and development. The Services' "corporate" labs focus on discovering and transitioning technology to the warfighter, and the centers transform technology into fieldable systems and deliver them into the hands of the warfighter. The 14 University Affiliated Research Centers and 11 Federally Funded Research and Development Centers pursue cutting edge research and maintain the domain expertise essential to apply new technologies to DoD missions and systems. Extramural funding organizations such as the Office of Naval Research, Air Force Office of Strategic Research and Army Research Office fund broad portfolios of basic research at universities, small businesses, and government labs to advance the state of the art in technologies of interest to their mission areas. The storied Defense Advanced Research Projects Agency looks out even farther, investing in early-concept, game-changing capabilities.

[2] A Defense Science Board study found that "in an era of globalization, the Labs continue to fulfill vital missions on behalf of the warfighter," but "rapidly changing technology landscape means that the Labs also must adapt their mission to continue to serve and ready themselves for their evolving needs of the warfighter." See *Defense Research Enterprise Assessment*, Defense Science Board (Jan. 2017), https://apps.dtic.mil/dtic/tr/fulltext/u2/1025438.pdf [hereinafter Defense Research Enterprise Assessment]. Similarly, a 2017 Government Accountability Study on the Defense Science and Technology enterprise found that "DOD's ability to adopt leading commercial practices in its approach to managing science and technology investments is limited by its funding policies and culture." See *Defense Science and Technology: Adopting Best Practices Can Improve Innovation Investments and Management*, U.S. Government Accountability Office, GAO-17-499 (June 2017), https://www.gao.gov/assets/690/685524.pdf.

[3] *Interim Report*, NSCAI at 28 (Nov. 2019), https://www.nscai.gov/reports.

[4] *First Quarter Recommendations*, NSCAI at 7 (Mar. 2020), https://www.nscai.gov/reports [hereinafter First Quarter Recommendations].

prototyping, experimentation, and adaptation. Across its Components and Services, DoD must develop, deploy, and move faster than our competitors.

At the same time, DoD must continue to leverage and invest in existing AI expertise wherever it resides—whether that is in affiliated and sponsored research organizations, academia, private sector partners (large and small), national laboratories, Federally Funded Research and Development Centers (FFRDCs),[5] or University Affiliated Research Centers (UARCs).[6]

To improve its internal ability to accelerate research, development, and fielding of AI-enabled capabilities, the Department should urgently: 1) Equip the enterprise with necessary resources, tools, and infrastructure to support AI R&D; 2) Invest in test and evaluation, verification, and validation capabilities to responsibly accelerate development of robust capabilities; 3) Optimize transition of breakthroughs from the laboratories to the field; and 4) Unlock innovation at the defense laboratories through partnerships.

Issue 1: Equipping the Enterprise for AI R&D

The ability of DoD research entities to accelerate AI research and development is limited by a lack of enterprise-wide access to data, sufficient computing support, cloud-based tools and resources, and state of the art software, as well as experience in DevSecOps[7] and change management. Ready access to these tools and the

---

[5] FFRDCs are government-owned, contractor-operated research centers designed to meet a "special long-term research or development need which cannot be met as effectively by existing in-house or contractor resources." DoD has three R&D laboratory FFRDCs: the Lincoln Laboratory, the Software Engineering Institute, and the Center for Communications and Computing. Across the government, 12 agencies support a total of 42 FFRDCs. See *Master Government List of Federally Funded R&D Centers*, NSF (Mar. 2020), https://www.nsf.gov/statistics/ffrdclist/#agency.

[6] UARCs are strategic DoD research laboratories associated with universities that include education as part of their overall mission. These not-for-profit organizations maintain essential research, development, and specific engineering core capabilities, and enter into long-term strategic relationships with their DoD sponsoring organizations. DoD sponsors 14 UARCs. See *Federally Funded Research and Development Centers and University Affiliated Research Centers*, Defense Innovation Marketplace, https://defenseinnovationmarketplace.dtic.mil/ffrdcs-uarcs/ [hereinafter Federally Funded Research and Development Centers and University Affiliated Research Centers].

[7] DevSecOps is an organizational software engineering culture and practice that aims at unifying software development (Dev), security (Sec) and operations (Ops). The main characteristic of DevSecOps is to automate, monitor, and apply security at all phases of the life cycle: plan, develop, build, test, release, deliver, deploy, operate, and monitor. DoD's Chief Information Officer issued the DoD Enterprise DevSecOps Reference Design in August 2019. See *DoD Enterprise DevSecOps Reference Design: Version 1.0*, Department of Defense Chief Information Officer (Aug. 2019), https://dodcio.defense.gov/Portals/0/Documents/DoD%20Enterprise%20DevSecOps%20Reference%20Design%20v1.0_Public%20Release.pdf?ver=2019-09-26-115824-583 [hereinafter DoD Enterprise DevSecOps Reference Design: Version 1.0]. The DoD's Joint AI Center (JAIC) is building a joint common foundation (JCF) to create a specialized AI/machine learning (ML) DevSecOps environment on an enterprise cloud construct. See *About the JAIC*, JAIC (last accessed July 13, 2020), https://www.ai.mil/about.html.

environments in which they are developed will enable innovation to take off, providing researchers and developers across the Department with the ability to leverage AI to solve problems and build new capabilities. Without them, the Department's AI ambitions will not be achievable.

Security processes and laboratory funding models slow—and, in some cases, impede—the ability of researchers to gain access to cloud-based storage, computing services, and optimal software tools. While industry and academia reap the benefits of widespread democratization of AI software and tools through the open source community, DoD researchers find themselves cut off from this vast resource of cutting-edge capabilities. Commercially licensed software is also advancing rapidly, and current DoD methods for clearing third-party software, sometimes taking more than a year, cannot keep pace.

To establish a world-class ability to develop AI-enabled solutions, an organization must invest in the building blocks of data, computing support, and software and digital tools—and make them all easily accessible to developers and users. The DoD is no different. In this memo, we recommend investments the Department can make now to start building that necessary enterprise infrastructure and establish shared resources to enable AI R&D across the enterprise. The Commission continues to develop a comprehensive vision for a future DoD digital ecosystem architected to enable widespread innovation in AI at all levels.

Recommendation 1: Create an AI software repository to support AI R&D.

DoD needs an enterprise-level software repository, supported with continuous Authorization to Operate (ATO)[8] to accelerate AI R&D.[9] With the widespread use of AI to drive innovation, DoD researchers in key domains of science and application find themselves in need of the same core set of tools—many of which are open source—to stand up local AI development pipelines.[10] These include tools to support

[8] See Recommendation #2 below for discussion of continuous ATO. ATO is an authorization granted by a designated authorizing authority for a DoD information system to process, store, or transmit information. An ATO indicates a DoD information system has adequately implemented all assigned information assurance controls to the point where residual risk is acceptable to the designated authorizing authority. See John Grimes, *DoD Information Assurance Certification and Accreditation Process (DIACAP),* DoD Instruction 8510.01 (Nov. 28, 2007), http://www.acqnotes.com/Attachments/DoD%20Instruction%208510.01.pdf.

[9] This could be modeled after RepoOne, an Air Force pathfinder effort that is part of its Platform One DevSecOps service, which serves as a central repository for the source code to create hardened and evaluated containers for DoD networks and includes various open-source products. See *Platform One: DoD Enterprise DevSecOps Services*, U.S. Air Force, (last accessed July 13, 2020), https://software.af.mil/dsop/services/ [hereinafter Platform One: DoD Enterprise DevSecOps Services].

[10] These tools fall into many categories. Service-oriented tools are needed to support the data pipeline, storage, development pipeline (particularly DevSecOps), TEVV, machine learning, and data analytics. Tools should be developed and hosted to support red teaming of AI technologies, the ethical

cloud-based research collaboration, science workflows and AI-driven experimentation, and test and evaluation.

A centrally managed repository of the latest AI-supporting software and releases, where software would be tested, tagged, and containerized for authorized use on designated levels of government networks, would provide researchers and developers across the enterprise access to the necessary tools to accelerate AI R&D. Additionally, it would overcome current barriers caused by lab funding constructs that preclude this access.

A repository would enable the Department to build a robust community of AI software developers and users alongside leading AI researchers across the DoD research enterprise, helping to overcome stovepipes and focus efforts towards the state of the art. As the resource matures, it could become a type of AI market, populated with swappable solutions—licensed from vendors, open source, or tunable models. This would provide researchers and developers across the enterprise, and notably those at the edge, with the ability to innovate and leverage AI to solve problems.

Proposed Executive Branch Action

DoD should create a repository either under the purview of the Office of the Under Secretary of Defense for Research and Engineering (OUSD R&E), or by leveraging the Joint Artificial Intelligence Center's (JAIC) Joint Common Foundation (JCF).[11] Either option should build on the successful model implemented by the Air Force's Platform One, which provides access to hardened containers and open source tools, and enables a DevSecOps pipeline with a continuous ATO for rapid deployment and scalability.[12]

This enterprise AI software platform should be executed in close coordination with the DoD Chief Information Officer (CIO) and Chief Data Officer (CDO) and should

development of AI, the generation of synthetic data, and AI modeling and simulation. Tools are needed to support the development of autonomy at rest in virtual environments as well as autonomy in motion supported by AI embedded hardware in physical environments. To support AI at the tactical edge, tools are needed to support harvesting data with store, forward, and integration appliances.

[11] The JCF is intended to break down barriers to entry and scale access to AI technologies for the warfighter by creating a secure digital environment for developers to work and train AI models. As currently conceived, the JCF's AI software tools will prioritize support for the JAIC's AI development mission and community. Software will be vetted and added based on the relevance and level of demand to the JAIC's mission areas. The Department could expand and scale engagement underway between JAIC and Platform One to instantiate such a repository in order to have the effort to support a broader research-focused user base to include Service labs, FFRDCs, UARCs, and other cleared researchers.

[12] Platform One is the Air Force's fee for service DevSecOps capability that provides cloud-based collaboration tools, cybersecurity tools, source code repositories, artifact repositories, and development tools, as well as managed software factories. See Platform One: DoD Enterprise DevSecOps Services.

align with the Department's general software strategies and implementation plans. The goal should be an enterprise-wide, modern digital infrastructure.

Recommendation 2: Promote ATO reciprocity as the default practice within and among programs, Services, and other DoD agencies to enable sharing of software platforms, components, infrastructure, and data for rapid deployment of new capabilities.

To better equip the enterprise—including its research components—the Department must strengthen adoption of the policies and processes that allow for the use of modern software components, tools, and infrastructure across multiple programs, once they are accredited as secure.[13] For commercial off-the-shelf or open source tools, this would mean that once the accreditation was done, the tools would be available for use across the Department, as appropriate.

Historically, DoD's approach to security accreditation for software components has considered each network and context as distinct from the outset, which drives time- and effort-intensive processes. While current Departmental policies state that ATO reciprocity should be exercised to the maximum extent possible,[14] default practices and behaviors across the enterprise have been slow to change.[15] The current scale of reciprocity adoption across DoD programs, Services, and agencies remains inadequate to enable DevSecOps and presents a barrier to the nimble approach necessary to support AI R&D.

The Department must accelerate the move to a posture that eliminates perceived trade-offs between cyber security and modern development. The standard expectation should be for Authorizing Officials (AOs) to accept reciprocity as the default, placing the onus on the AO to prove why another Component's ATO is insufficient. DoD CIO, Service CIOs, and Component heads should maintain full visibility on reciprocal ATOs, establish reporting metrics and measures relative to

---

[13] Working groups across the DoD are currently focusing on aspects of this issue. Meaningful progress is contingent on the Department expeditiously translating their work into formal guidance and processes, coupled with robust training and education.

[14] The Department's current policy contained in DoD Instruction 8510.01 Risk Management Framework (RMF) for DoD Information Technology (IT) places emphasis on the promotion of ATO reciprocity to the maximum extent possible, but stops short of making reciprocity default and rejection of reciprocity an exception requiring justification. See Teresa Takai, *Risk Management Framework (RMF) for DoD Information Technology (IT)*, Department of Defense Instruction 8510.01 (July 28, 2017), https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/851001p.pdf?ver=2019-02-26-101520-300.

[15] Select organizations have developed guidance to streamline the Department's decision structure for cyber risk management, thereby providing options to reduce time to initial ATO—but this guidance is implemented at the Component-level and fails to drive reciprocity across the enterprise, See Memorandum from Deputy Chief Information Officer of the Air Force, to Authorizing Officials (Mar. 18, 2019), https://federalnewsnetwork.com/wp-content/uploads/2019/04/AF-fast-track-ATO-memo-march-2019.pdf.

accepted and denied ATOs, and assess the data on a routine basis to inform policy, guidance and training.[16]

In parallel with the above, the Department should prioritize scaled adoption of shared service platforms, processes and workforce training to enable a "continuous ATO" approach. Continuous ATO allows the AO to authorize a platform's security and testing *process* instead of the release of each *product*. In a continuous ATO construct, checks are automated and performed on every build and coupled with appropriate instrumentation and continuous run-time monitoring of operational software; accreditation is revoked if performance strays outside defined boundaries.[17]

While the Department is making progress to more widely embrace ATO reciprocity and establish the foundation for continuous ATO,[18] implementation has lagged behind guidance. The lack of ATO reciprocity still consistently arises in discussions across the spectrum of AI stakeholders—from researchers to developers and end-users—as a primary contributor to delayed fielding of new capabilities. For AI researchers, this often means that they must either opt for easily accessible tools that may not be at the cutting edge or best suited to their project, or incur months of delay by pursuing a new authorization for a software component on their specific network.

---

[16] For these steps to be meaningful, they must be based on a common operating picture. DoD's Enterprise Mission Assurance Support Service (eMass) is the recommended system of record, but Components' use of the tool is inconsistent. To promote reciprocity at scale, DoD must prioritize a common tool that is an accessible, accurate reflection of cyber security assessments and authorizations. A 2018 DoD Inspector General Report found that several Components maintain duplicate systems and processes for cybersecurity documentation, citing functionality limitations within eMass as a primary reason for selecting an alternate tool. DoD Instruction 8510.01 states security authorization documentation should be made available in eMass or another tool with the means of providing visibility. As written, the instruction is insufficient to achieve the necessary data accuracy and cross-Component visibility. See *DoD Information Technology System Repositories*, U.S. Department of Defense Office of the Inspector General, DODIG-2018-154 (Sept. 24, 2018), https://media.defense.gov/2018/Sep/26/2002045060/-1/-1/1/DODIG.

[17] Initiatives such as National Geospatial Intelligence Agency's ATO in a day effort, and the Air Force's Platform One and Kessel Run programs have served as pathfinders to develop the processes, tools, and infrastructure to support continuous monitoring and continuous ATO approaches. The Defense Security/Cybersecurity Authorization Working Group DevSecOps sub-group is working to build an implementable DoD-wide continuous-ATO policy for integration into the DoD Risk Management Framework. See *Team 6: Continuous ATO*, U.S. Air Force (Mar. 25, 2020), https://repo1.dsop.io/dsawg-devsecops/continuous-ato-guidance/team6_artifacts/-/blob/6f2e6f586408875dca96ccbd63bcd43cbccc734c/team_6_details.pdf.

[18] Notable progress includes a provisional authorization, released Fall 2019 by the Defense Information Systems Agency, that allows ATO reciprocity in the DoD of FEDRAMP-approved cloud service providers at IL2, and the Air Force's Fast-Track ATO, RMF Now and Ongoing Authorization Risk Management Framework (RMF) pathways; and through the continuous-ATO process through Platform One. Platform One has been recently designated a DoD Enterprise Service Provider for DevSecOps. See Memorandum from Department of Defense Chief Information Officer, to Chief Management Officer of the Department of Defense, et al., (May 22, 2020), https://software.af.mil/wp-content/uploads/2020/05/DoD-CIO-Signed-Memo-Enterprise-Service-Provider-for-DevSecOps.pdf [hereinafter May 22, 2020 Memorandum from DoD CIO, to DoD CMO, et al.]

Proposed Executive Branch Action

Recognizing the promising efforts underway across the Department, the Commission underscores the need for DoD to make reciprocity the default practice, and promote maximum use of infrastructure as code and automation of security controls to enable continuous ATO.[19] The Commission also recommends the DoD CIO expedite and scale efforts toward a single, enterprise repository for ATO artifacts that supports data across classification levels and is complete with tools and access rules that enable Components to discover existing and continuous ATOs.[20]

Recommendation 3: Create a DoD-wide AI data catalog to enable data discoverability for AI R&D.

The Commission recommends that the CDO[21] build and manage a secure, online DoD-wide AI data catalog that would enable DoD researchers and developers to identify data resources that could fuel new research and development opportunities for a range of AI approaches, including machine learning, model-based, and symbolic.

The Commission continues to examine broader enterprise requirements around DoD data access and management for the development and application of AI solutions. An effective posture will be critical not only for AI, but also as the foundation for broader DoD modernization efforts. Creating a catalog represents a first step towards leveraging the Department's data for AI and providing a key resource to the DoD-affiliated AI research community.

---

[19] This recommendation is consistent with the DoD Enterprise DevSecOps Reference Design, which outlines additional tools, approaches, metrics, and thresholds as modern software development best practices and sets preconditions for authorization and assessment inheritance at the enterprise and local levels, including continuous ATO. See DoD Enterprise DevSecOps Reference Design: Version 1.0.

[20] This echoes a recommendation made by the Defense Innovation Board's 2019 Software Acquisition and Practices (SWAP) Study. See *Software is Never Done: Refactoring the Acquisition Code for Competitive Advantage*, Defense Innovation Board (May 3, 2019), https://media.defense.gov/2019/Apr/30/2002124828/-1/-1/0/SOFTWAREISNEVERDONE [hereinafter Software is Never Done: Refactoring the Acquisition Code for Competitive Advantage]. The repository should have the ability to ingest from other data sources, including Component-specific security assessment and documentation tools, for one common operating picture. DoD CIO should make any policy changes necessary to support data completeness and accuracy.

[21] In the National Defense Authorization Act for Fiscal Year 2020, Congress directed DoD to move the position of CDO to report directly to the CIO, away from its prior position under the Chief Management Officer. The legislation also gave the CDO the principal responsibility for providing for the availability of common, usable, Defense-wide data sets. See National Defense Authorization Act for Fiscal Year 2020, Public Law 116-92. In an undated memorandum to the Deputy Secretary of Defense, DoD CIO Dana Deasy underscored an initial priority for the CDO Office to ensure that data policies, standards, and implementation are fully aligned to the needs for all-domain operations against a capable adversary. See Memorandum from Department of Defense Chief Information Officer, to Deputy Secretary of Defense, https://federalnewsnetwork.com/wp-content/uploads/2020/01/012719_cio_cdo_memo.pdf.

Proposed Executive Branch Action

The Department should first undertake an inventory at the military department and defense wide organization level to gain a picture of the type, volume, structure, and location of the Department's data assets. This should include both internally generated and collected data as well as existing commercial datasets for which the Department has ongoing access. In the process, a mechanism could be instituted to capture and maintain the inventory as a living, online resource—integrating it into the catalog itself. This effort would scope the ultimate size and help set metrics for populating the catalog, while also serving as a mechanism to develop a Department-wide data cataloging and sharing strategy.[22]

The CDO could leverage the proof of concept data cataloging activity underway through the JAIC's JCF effort,[23] to build out an enterprise-wide solution that could support a broader user base of researchers, including those at Service labs, FFRDCs, UARCs, and others.

Such a resource would provide DoD AI researchers a tool for data set discovery, pointing to the data stewards[24] who host the data. It would enable researchers to identify and request access to existing, annotated training data, processed AI-ready data with weights, as well as raw data. The catalog should contain relevant knowledge about the data, including taxonomic information about how the fields are related, and how the data is currently being extracted, transformed, used, deidentified, and kept up to date. Each data set should adhere to the minimum data documentation standards as recommended in the Commission's First Quarter Recommendations.[25]

As the catalog expands and the CDO formalizes policies and processes around it, it could become possible in time to assign the system the ability to broker data access requests and provide direct authorized access to data sets.[26] Once it matures as a

---

[22] Such a catalog and sharing strategy should include clear access guidelines and de-identification and privacy standards for any data sets involving personally identifiable information.

[23] The JAIC JCF data catalog is intended as a resource modeled off data.gov to support the development community aligned with its national mission initiatives.

[24] For NSCAI's purposes, data stewards are those who are responsible for implementing data governance for an organization including data content, context, and upholding rules for authorized sharing and access.

[25] NSCAI recommended that minimum documentation must reveal what the data is; why, how, and from whom it was collected; and what it could be appropriately used for. See First Quarter Recommendations at 71.

[26] The catalog system will require the formation and evolution of a foundational set of policies and processes. Many different authorities are at play with data sharing and access. These include organizational, user, legal, privacy, and human protection policies. They apply to the data, the research purpose, and the capability being developed. These authorities often conflict with each other, creating the need for a more agile process for deconfliction and decision making. However, AI R&D can begin to take advantage of the catalog while new policies and processes are developed to support more sensitive data and usage.

centralized resource, the Department could consider the added benefits of scaling to a federation of distributed catalogs that are automatically synchronized and supported by an appropriate knowledge framework, as well as augmenting the system to provide inventory services for trained AI models in addition to data. Furthermore, it could evolve to provide visibility and point researchers to additional standard public and private data sets to fuel their research, with documentation of availability and appropriate use.

Success rests on the actions of data stewards. Today, managers and data stewards are incentivized to securely harbor the data for which they are responsible. DoD leaders should promote a culture of sharing and implement incentives to better leverage data as a strategic asset.[27] DoD should allocate funds across its components to explicitly support data labeling, curation, and skill-building efforts. Prioritizing this investment would send a message from senior leadership that future military strength is critically dependent on a data-enabled force.

Recommendation 4: Expand Section 219 Laboratory Initiated Research Authority funding to support AI infrastructure and software investments at DoD laboratories.

The Commission recommends that Congress update the authorities it has granted to defense labs through the Laboratory Initiated Research Authority, provided in Section 219 of the Duncan Hunter National Defense Authorization Act (NDAA) for Fiscal Year 2009, to enable higher-level dollar investments in infrastructure and software assets to support AI research, prototyping, and testing.[28]

The Section 219 Laboratory Initiated Research Authority provides lab directors a means—through service charges to customers or a percentage of funds available to the laboratory —to fund projects they consider to be a priority, in four categories: (1) basic and applied research, (2) technology transition, (3) workforce development, and (4) revitalization, recapitalization, or repair or minor construction of lab

[27] The Department should establish clear guidance from the Secretary of Defense level on expectations around data sharing, as well as policy and practice that requires DoD researchers to curate and register their data with the catalog in order to continue to receive research funding, and build performance metrics and data goals within the performance plans for managers and data stewards.

[28] Since 1995, Congress has granted authorities that address hiring, infrastructure, and technology transition challenges to defense labs. These authorities provide defense lab directors with certain flexibilities within the established legal framework to manage their operations. See *Duncan Hunter National Defense Authorization Act for Fiscal Year 2009*, Public Law 110–417 [hereinafter 2009 NDAA]. This authority was made permanent in the 2017 NDAA, with an accompanying rise in the rate that labs are authorized to charge to customers or collect from available funding to finance the fund from the previous allowance of no more than 3% to a mandated charge between 2% and 4%. Moreover, cost compliance requirements for infrastructure projects were updated, capping expenses at $6 million (updated from $4 million), as codified in 10 U.S.C. § 2805(d).

infrastructure.[29] These projects include those not specifically tied to defined requirements outside of the normal two-year budget planning process.[30]

These innovation funds equip lab directors—who possess the most intimate knowledge of their labs' means and potentials—with an ability to invest not only in higher-risk, curiosity-driven research efforts that can unlock the next generation of capabilities, but also in lab infrastructure that would likely not make it through the major military construction (MILCON) requirements weighting process.[31]

Proposed Legislative Action

Congress should raise the authorized cap for laboratory infrastructure investments, currently set at $6 million, in order to provide laboratories with the ability to invest in equipment and testbed infrastructure necessary for robust AI research, prototyping, and testing. Furthermore, Congress could mandate that laboratories use the full four percent service charge to support the innovation funds, which would provide additional capital to support AI-related research and infrastructure investments while eliminating the comparative disadvantage associated with charging customers a fee higher than that of other DoD labs.[32]

Proposed Executive Branch Action

To further strengthen the ability of the defense laboratories to maximize this authority, the DoD Comptroller should create accounts to allow the labs to bank Section 219 funds from year to year in order to fund infrastructure projects that

---

[29] See 2009 NDAA.

[30] The Defense labs are managed by a range of funding models. Air Force and Army labs rely on appropriated funding provided from the Service—often referred to as mission funding—and from customers. External customers, typically program offices, provide funding to Defense labs for technology development activities and related research. The Air Force and Army funding structure is in contrast to Navy R&D activities, which operate under the Navy Working Capital Fund—a revolving fund that finances Department of the Navy activities on a reimbursable basis.

[31] The Defense Science Board found that between FY13-15, Section 219 authority funding supported 39% of the total laboratory infrastructure investments. See *Final Report of the Defense Science Board Task Force on Defense Research Enterprise Assessment,* Defense Science Board at 52 (Jan. 25, 2017), https://apps.dtic.mil/dtic/tr/fulltext/u2/1025438.pdf [hereinafter Report of the Defense Science Board Task Force on Defense Research Enterprise Assessment].

[32] A 2018 GAO report found that most labs were not using the full 4% of all funds available, or charging customers the full fixed percentage fee of 4% of costs, as allowed by law. They found that Navy labs were charging 2% and Army between 2-3%; while the Air Force often utilized the entire 4% of funds available, it charged customers nothing due to weak mechanisms of financial management and accounting. The customer fee functions similar to an overhead charge: if a program office were to require services worth $10,000 at an army research lab testing facility, it would be charged $10,400, with the $400 made available for reinvestment in basic research, infrastructure projects, or other activity permissible under the lab authority. *Defense Science and Technology: Actions Needed to Enhance Use of Laboratory Initiated Research Authority*, U.S. Government Accountability Office, GAO-19-64 (Dec. 2018), https://www.gao.gov/assets/700/696192.pdf.

exceed the $6 million (or adjusted level) cap.[33] This should be paired with implementation of robust accountability measures to ensure individual laboratories are fully leveraging innovation funds to advance DoD modernization priorities.

Issue 2: Establishing AI Test and Evaluation, Verification and Validation Capabilities

Test and Evaluation, Verification and Validation (TEVV) represents a critical, and cross-cutting factor in the process to develop, deploy, and maintain new capabilities responsibly, reliably, and at speed. Robust and readily available AI TEVV capabilities will provide the Department the ability to more aggressively pursue AI R&D and speed delivery of new capabilities to the warfighter.

For AI systems and solutions, the DoD must rethink its approach to TEVV. TEVV must be integrated as a *continuous* component of development, deployment, and maintenance processes, which breaks the traditional DoD paradigm.[34] It is difficult to assure the behavior of an AI system when encountering unanticipated use cases in unfamiliar environments, which necessitates use case-specific validation and regular revaluation and re-certification.[35] This puts a premium on TEVV that enables operators to make informed decisions around employing the system for specific use-cases and environments. Furthermore, the diversity of AI methods and applications demands a diversity of TEVV methods, many of which require significant research and development to advance the state of the art.

In our First Quarter Recommendations, the Commission noted research on AI TEVV as a priority area for federal investment. We emphasize that the DoD must continue to support this research, while it formalizes processes and builds the tools and infrastructure to support TEVV for responsible application of today's AI systems and solutions.

---

[33] Currently, labs cannot carry over Section 219 funds from one fiscal year to the next. See Report of the Defense Science Board Task Force on Defense Research Enterprise Assessment.

[34] Traditionally, DoD conducts TEVV in the final stages of development and as a system is readied for operation. AI development is based on an agile process that embodies an iterative cycle in which testing and evaluation plays a continual role. Once in use, a machine learning system should be subject to rapid, iterative updates and releases that are tested and quality checked in a controlled environment, before being pushed out.

[35] We note that the DoD will initially employ narrow AI for defined use cases, and not likely allow for individual systems to learn independently in the field for safety-critical tasks. Rather, the Department could adopt a posture similar to that of Tesla, where telemetry is collected from the dispersed fleet of privately-owned vehicles and used to update ML models in a controlled environment. Updates are then pushed back to the full fleet only after tests and quality checks are run on any changes. See *Tesla Vehicle Safety Report*, Tesla (last accessed July 13, 2020), https://www.tesla.com/VehicleSafetyReport; *Autopilot*, Tesla (last accessed July 13, 2020), https://www.tesla.com/autopilotAI.

DoD should accelerate development of the test infrastructure to support AI across the stages of the research, development, and validation and verification cycle.[36] At Service labs, such test resources are difficult to set up and maintain due to project-based funding models, which do not provide stable funds for technical or maintenance support. The Department should explore funding mechanisms to ensure investments in AI TEVV infrastructure are made across the Service labs and warfare centers.

DoD must invest in new TEVV capabilities and move from traditional "waterfall" development processes to agile approaches, where testing and feedback from use are continuous and integrated as key components of the development and sustainment of AI tools and solutions.[37] The Department should build an infrastructure, framework, and tool set that supports the diversity of AI applications, doing so in a manner that embraces best practices from industry and ensures an ability to evolve and adapt as the technology and the science behind TEVV matures. Fostering these resources and approaches enterprise-wide will significantly accelerate the successful research, development, and transition of AI-enabled capabilities to the warfighter, transform logistics, and bring efficiencies to business operations.

Recommendation 5: Establish an AI testing framework.

The Commission recommends that DoD establish a foundational and adaptable AI testing framework to provide necessary assurance, guidance, and capabilities to the enterprise, overcoming a critical barrier to fielding AI capabilities at the speed of relevance. As AI R&D accelerates and the technology matures, the need for AI TEVV will grow. AI applications are extremely diverse and thereby necessitate a wide range of testing methods. Establishing common approaches to tailoring appropriate processes and tools to the type of AI application at hand will support the ability of DoD components to embrace and scale AI solutions by shortening the testing cycle and making test results interpretable and comparable across the Department.

---

[36] Next-generation AI test infrastructure needs to support TEVV covering all the ways AI is applied in military applications including: autonomy at rest and in motion, in virtual and physical environments, and AI at the tactical edge. It must support cloud-based AI as well as hardware embedded AI, configurable instrumentation, and be scalable, mobile, and replicable. Furthermore, evaluating a wide range of AI technologies and applications requires a repository of well-curated, diverse, and large data sets. The data pipeline needed requires automation for harvesting and collection, curation and tagging, and posting to a data repository for discovery. The infrastructure should be able to leverage synthetic data, modeling, simulation.

[37] DevSecOps for AI-driven capabilities requires TEVV to be embedded in the development process at a speed that changes how test beds need to be designed and architected to account for AI systems that are typically non-deterministic and dynamically changing over the course of their operation. These AI systems are also often deployed in operational environments where real-world data is non-stationary and prone to drift in quality and characteristic from the data used in laboratory testing. AI technology test beds therefore need to support the evaluation of continuously living software requiring new forms of TEVV automation. This is an evolving area of research requiring new methods and metrics incorporating synthetic data as well as consideration of adversarial threats.

An AI testing framework should:

1. Establish a process for writing testable and verifiable AI requirement specifications that characterize realistic operational performance.[38]
2. Provide testing methodologies and metrics that enable evaluation of these requirements--including principles of ethical and responsible AI, trustworthiness, robustness, and adversarial resilience.[39]
3. Define requirements for performance reevaluation related to new usage scenarios and environments, and distribution over time.
4. Encourage incorporation of operational usage workflow and requirements from the defined use case into the testing.
5. Issue data quality standards to appropriately select the composition of training and testing sets.
6. Support the use of common modular cognitive architectures within suitable application domains that expose standard interface points for test harnessing—supporting scalability through increased automation along with federated development and testing.
7. Support a cyclical DevSecOps-based approach, starting on the inside and working outward, with AI components, system integration, human-machine interfaces, and operations (including human-AI and multi-AI interactions).
8. Remain flexible enough to support diverse missions with changing requirements over time.

A Department-wide, core suite of TEVV practices for similar types of AI-enabled systems and applications (e.g., object detection in overhead imagery) would enable a comparison analysis among AI technology solutions, help determine the best options and where they may be best deployed, and help identify alternatives in the event that a particular AI algorithm or model is compromised. Given the diversity of use cases, the framework would not embody a one size fits all approach, but rather provide core capabilities and guidance adaptable across application areas.

An existing effort that would benefit this work is an initiative recently launched by the Office of the Director of National Intelligence, in partnership with Carnegie Mellon University's Software Engineering Institute and the University of Maryland's Applied Research Laboratory for Intelligence and Security,[40] to establish a National AI Engineering Initiative. The initiative will build and implement an R&D roadmap to advance the science of AI engineering, including in areas of system verification and validation, software engineering, and information assurance.

---

[38] This should be framed broadly, providing left/right limits that provide guidance but do not limit innovation.

[39] These testing methodologies and metrics should support robust red teaming, meeting the DoD's particular needs for solutions hardened to adversarial actions.

[40] Carnegie Mellon University Software Engineering Institute is a FFRDC and University of Maryland Applied Research Laboratory for Intelligence and Security is a UARC.

Proposed Executive Branch Action

The Secretary of Defense should appoint and resource a lead entity, with the applicable expertise and remit, to harness the AI TEVV community to develop and formalize a joint, common framework for AI TEVV. This effort should be completed within six months of tasking.

Recommendation 6: Expedite the development of tools to create tailored AI test beds supported by both virtual and blended environments.

Use of virtual and blended environments allows for more realistic testing and evaluation earlier in the development process leading to quicker, less costly, and more effective deployments. DoD should develop a robust set of common tools to support the TEVV of a wide range of AI-powered systems and capabilities—spanning AI in autonomy, at rest and in motion; cloud-based AI and hardware embedded AI; human-AI, human-machine, and machine-machine teaming; combined environmental domains of land, air, sea, and space; and more. Developed as joint, shared capabilities, they should be transferable, able to support flexible platforms of diverse types and sizes, and tailorable to specific groupings of technologies, environments, and use cases. Further, performance validation capabilities should also be made available at the edge, in abridged formats. The Department should integrate generally accepted AI test and verification methods employed in the private sector, where appropriate.

This tool set should enable a standardized, robust, and smart approach to iterative testing of digital technologies, to include:

- Virtual environments, and ability to blend live and virtual environments;[41]
- Robust modelling and simulation services;
- Instrumentation for increased understanding and transparency of AI modules;
- Digital twinning;
- DevSecOps environment;[42]
- System integration testing;
- Data capture for continuous development; and
- Generation and use of synthetic data as appropriate.

---

[41] These blended environments are described, collectively, as Live, Virtual, and Constructive (LVC). LVC enclaves combine simulation with physical interaction, enabling dynamic, tailored, safe, and holistic testing environments.

[42] The Air Force's Platform One DevSecOps stack and suite of services provides a model that could be replicated across the enterprise. On 22 May 2020, DoD CIO designated Platform One as one of the DoD Enterprise Service Providers for DevSecOps. See May 22, 2020 Memorandum from DoD CIO, to DoD CMO, et al.

Investing in these tools will support an enterprise-wide capability to conduct TEVV of AI systems, setting the foundation for scaling AI solutions as the technology rapidly matures as a key component of U.S. military competitiveness.[43] Establishing this ability will require a level of technical expertise not yet present across much of the TEVV enterprise, thus benefiting from a top-down push driven from the Departmental level that could bring together the technical, policy, and domain expertise needed.

Proposed Executive Branch Action

The Secretary of Defense should appoint and resource a lead entity to develop a roadmap and implementation plan and oversee its execution to build the enterprise-wide set of tools and resources for AI TEVV.

Recommendation 7: Create test beds to focus on evaluation of commercially available AI solutions that could serve DoD missions.

The Department could bolster its portfolio of AI TEVV tools by creating third party test beds—at FFRDCs, UARCs, or other contracted entities[44] to evaluate existing market and market-ready AI solutions for DoD-relevant missions.[45] These test beds would focus on:

1. Identifying representative DoD-specific embedded applications which are enabled by AI technologies;

---

[43] It would support the TRMC's efforts in realizing a new LVC Autonomy and AI test range; help the JAIC build out its JCF T&E tool set and capabilities, and enable the Services, FFRDCs, and UARCs to stand up virtualized and tailored AI test beds to support their R&D efforts.

[44] DoD should invest in and leverage AI talent where it exists, be it in FFRDCs, UARCs, or elsewhere. FFRDCs and UARCs are sponsored research entities under long-term contracts to accomplish tasks integral to the mission and operation of the sponsoring agency, free from profit motive or conflict of interest. FFRDCs are operated by universities or not-for-profit organizations and UARCs by universities. DoD sponsors 11 FFRDCs in total, 3 of which are research and development labs, which maintain long-term competencies in key technology areas, and 14 UARCs. In addition to these, DoD sponsors 3 systems engineering and integration FFRDCs and 5 studies and analysis FFRDCs. MIT Lincoln Laboratory, CMU Software Engineering Institute and IDA Communications and Computing Center are the 3 R&D laboratory FFRDCs. For the full list, see Federally Funded Research and Development Centers and University Affiliated Research Centers.

[45] The Defense Business Board recommended in 2016 that the DoD better leverage FFRDCs by giving them a greater role in tracking and evaluating new science and technology in order to enhance military capabilities, avoid strategic or technological surprise, and counter threats from potential adversaries. It recommended the Department use FFRDCs to vet and prototype scientific breakthroughs and the advanced technologies being offered by defense industry and private sector to ensure the capability meets DoD's requirements and is technologically mature. See *Future Models for Federally Funded Research and Development Center Contracts*, Defense Business Board (Oct. 2016), https://dbb.defense.gov/Portals/35/Documents/Reports/2017/DBB%20FY17-02%20FFRDCs%20Completed%20Study%20(October%202016).pdf.

2. Evaluating and leveraging commercially available and academically viable AI solutions to solve these applications; and
3. Defining technological gaps that are not being addressed by the commercial sector but are critical for the DoD community.

This effort would accelerate the ability of the Department to identify and assess applicability of commercial solutions, while building an understanding of technological gaps and limitations, as well as future investment opportunities.[46]

Proposed Executive Branch Action

The Department of Defense should fund the creation of an AI test bed capability at FFRDCs, UARCs, or other contracted entities to accelerate an ability to identify new military and national security capabilities that are immediately realizable using commercially available or academically viable AI solutions.

Issue 3: Accelerating the Transition of Technology Breakthroughs

To develop and deploy AI solutions at the pace of technological change and ahead of U.S. competitors, DoD must improve its ability to transition viable advances from research centers to acquisition programs and/or directly into the field. DoD must embrace an agile approach that enables development at the speed of operational relevance and incentivizes early delivery of minimally viable products to the end user to ensure AI-enabled solutions solve the right problems and are easily accessible to the user.

The budget structure and the sequential nature of the Department's management of the research and development cycle, paired with stove-piped communities and authorities, hinders DoD's ability to embrace a nimble, iterative, multi-stakeholder approach that could more effectively steward and bridge technology from the lab to the field.[47]

Optimized for the traditional large-scale weapons system paradigm in which transition from research to a fielded capability takes years and sometimes decades,

[46] Such resources should support entities such as the Defense Innovation Unit, AFWERX, and SOFWERX, who focus on identifying and scaling commercial technology to address mission priorities; and not detract from or duplicate their engagement with industry partners.
[47] The Navy's R&D framework calls out this process as a key obstacle to continued maritime superiority, namely that the "structure and cadence of budgeting activities drive near-term, fragmented decision-making and foster a protectionist mindset at the expense of strategic program effectiveness." And that "prototyping, experimentation and demonstration are misallocated in acquisition vice earlier in development." See *Naval Research & Development: A Framework for Accelerating to the Marine Corps After Next,* Office of Naval Research (Feb. 2018), https://www.onr.navy.mil/en/our-research/naval-research-framework.

the DoD budget construct is not well-suited for an AI development cycle that can take months, weeks or even days. The separation of funding for research, development, prototyping, and fielding in the Department's traditional R&D budget activity categories runs counter to the optimal development cycle for AI, which is rooted in a tightly-coupled, *iterative* process of researching, developing, and fielding.[48] This process is driven by an ability to prototype and test early with end users, building a capability through iterative improvements.

By distinguishing between research and development funds and operating funds, appropriations law that inflexibly pairs each DoD spending category with its allowable uses further complicates the continuous development cycle necessary to derive value from AI applications.

Recommendation 8: Support the DoD software and digital technologies budget activity pilot and its expansion to include an S&T development effort.

Congress should support the DoD software and digital technologies pilot program designed to allow for flexibility in funding the full lifecycle of development, procurement, deployment, assurance, modifications, and continuous improvement for digital technologies.[49] Furthermore, DoD should expand the pilot in

---

[48] A Congressional Research Service defense budget primer includes a table summarizing DoD research, development, test, and evaluation budget categories 6.1 through 6.7. See *Defense Primer: RDT&E*, Congressional Research Service at 1 (Apr. 29, 2020), https://crsreports.congress.gov/product/pdf/IF/IF10553. These budget categories become a limiting factor for DoD Service labs that primarily receive 6.1 and 6.2 basic and applied research funding, but do not receive sufficient 6.3 and 6.4 development and prototype funding. This hinders their ability to prototype early in the development process.

[49] This is being led by the DoD Office of the Under Secretary of Defense for Comptroller (OUSD C) and Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD A&S), based on the findings and recommendations of the Defense Innovation Board's Software Acquisition and Practices Study. See Software is Never Done: Refactoring the Acquisition Code for Competitive Advantage. Jeff Boleng, Special Assistant for Software Acquisition to the Under Secretary of Defense for Acquisition and Sustainment, publicly stated the goal of the pilot as "simplifying the budget process, increasing the visibility, accountability of the funding." See Billy Mitchell, *DOD has OMB Support for Special Software-only Appropriations Pilots*, FedScoop (Sept. 10, 2019), https://www.fedscoop.com/dod-omb-support-special-software-appropriations-pilots/. In public remarks made March 3, 2020, Undersecretary of Defense for Acquisition and Sustainment, Ellen Lord, underscored the significance of the pilot, asserting "we will begin to see results almost instantaneously, because the administrative burden of making sure you are charging the right development number, the right production number, the right sustainment number, slows things down." Jared Serbu, *Pentagon Teeing Up Nine Programs to Test New 'Color of Money' for Software Development*, Federal News Network (Mar. 4, 2020), https://federalnewsnetwork.com/acquisition/2020/03/pentagon-teeing-up-nine-programs-to-test-new-color-of-money-for-software-development/; *West 2020: 3 March 2020 Morning Keynote with The Honorable Ellen Lord*, WEST Conference, YouTube (Mar. 3, 2020), https://www.youtube.com/watch?v=VGlqjyMhtok&list=PLFZb4znlHwx0TcsirmyYD6k5BAYxDRwU0&index=6&t=0s.

Fiscal Year 2022 to include a program that explicitly supports an S&T development effort.

This pilot capability, proposed as the creation of a new budget activity (BA 8), seeks to overcome the barrier that DoD spending categories pose to the development and sustainment of digital technologies. The Office of the Under Secretary of Defense for Acquisition and Sustainment and the Office of the Under Secretary of Defense for Comptroller selected nine programs to begin to pilot the BA 8 for Fiscal Year 2021. If formalized, the BA 8 would be established for each Service and Defense-wide under the Research, Development, Test, & Evaluation appropriation and enable two-year funding.

Selected based on nominations from each Service, the Office of the Secretary of Defense, and Defense Agencies, the proposed programs for the Fiscal Year 2021 pilot include both weapons systems and defense business systems and represent efforts that are fully-funded with a high likelihood of success. However, none of the selected programs embody efforts at earlier stages in the development process.[50] By including a Science and Technology (S&T) development effort in Fiscal Year 2022, the Department would effectively test the impact of the single funding mechanism for the entirety of the AI research and development process.

Proposed Legislative Action

The Commission recommends that Congress appropriate funds to support the BA 8 pilot program for Fiscal Year 2021, in order to begin to test the construct as a mechanism to fund the full life cycle of development, procurement, deployment, assurance, modifications, and continuous improvement for digital technologies.[51]

Proposed Executive Branch Action

The Commission further recommends that in Fiscal Year 2022 the Department expand the pilot to include a program that explicitly supports an AI S&T development effort.

Recommendation 9: Encourage Services to build AI development models that integrate AI experts, domain experts, acquisition experts, and end users.

[50] Programs are: Risk Management Information (Navy), Maritime Tactical Command and Control (Navy); Space Command and Control (Space Force); Operational Medicine Information System (Defense Health Agency); National Background Investigation Services (Defense Counterintelligence and Security Agency); Global Command and Control System - Joint (Defense Information Systems Agency); Defensive Cyber Operations (Army); and Project Maven.
[51] At time of publication, the House Appropriations Committee has approved funding for 8 of the 9 proposed projects for the BA 8 pilot.

The Department should adopt multi-stakeholder and multi-disciplinary development models across the research enterprise. Multi-disciplinary teams with early user interaction are the engines of AI development. Without the domain knowledge and end user context, the resulting AI-based system risks failure. Additionally, domain scientists and engineers without AI expertise may not appreciate the full benefit and applicability of AI technology, and again, the end result suffers.

Furthermore, integration of acquisition experts and transition partners early in the development process can set the foundation, and expedite, successful transition. For example, the Navy has created the "AI DevRon" concept, a single entity accountable start to finish for the life cycle of capability development.[52] Another successful model to address near-term operational requirements are the Tactical Data Teams used by Army Futures Command (AFC) and Army Special Operations Command (USASOC).[53]

Proposed Executive Branch Action

The Secretary of Defense should issue guidance to the Services to adopt AI development models that integrate AI experts, domain experts, acquisition experts, and end users. This approach should become the default, rather than the exception.

Issue 4: Innovation across DoD Laboratories

AI is a fast-evolving field. Better coordination across the DoD research community and more robust connections with outside researchers would bolster the ability of DoD researchers to move quickly and stay on the cutting edge.[54]

Researchers in government labs must be able to connect with counterparts in other government labs, academia, and the commercial sector and participate in the conferences where the latest breakthroughs are presented. However, administrative hurdles around public release and ad hoc connections to academic counterparts hinder the ability of DoD researchers to engage fulsomely with the non-DoD research community and stay abreast of developments in the field of AI.[55]

---

[52] The new entity is accountable for requirements, acquisition, contracting, T&E, delivery, and monitoring, among other things.

[53] This model brings AI/ML expertise forward to the field in the form of 3 to 6 person teams to build AI solutions for real-time operational problems. Executed by a small business, Striveworks, under contract with AFC and USASOC, they are currently supporting efforts in Central Command and Indo-Pacific Command Areas of Responsibility.

[54] The Defense Science Board assessed this as a key component in the future success and value proposition of the DoD labs, calling on them to "embrace open innovation and technology defense -- security need not equal isolation." See Defense Research Enterprise Assessment.

[55] For detail of the administrative burdens around conference approval, See id. at 21.

Recommendation 10: Direct the Services to adopt open innovation models through the Service labs.

The Secretary of Defense should direct and incentivize the Services to adopt open innovation models at their laboratories, akin to the Army Research Labs' (ARL) Open Campus. ARL's Open Campus, launched in 2013, is a framework through which ARL scientists and engineers work collaboratively and side-by-side with visiting scientists in ARL's facilities and as visiting researchers at collaborators' institutions.[56] These collaborative endeavors, driven by mutual scientific interest and investment by all partners, work toward the Army's goal of building an S&T ecosystem that encourages groundbreaking advances in basic and applied research areas of relevance to the Service.

ARL has since complemented this initiative with the creation of ARL Extended—comprising four regional hubs that house ARL researchers and staff to facilitate more interaction between researchers, academic institutions, and regional companies.[57] With this distributed hub and spoke model, ARL has established a presence in some of the leading technology hubs across the country. Furthermore, it serves as a recruiting mechanism for talent, bringing young researchers into contact with national security problems early in their training.

Such models of academic exchange and collaboration would complement Navy and Air Force efforts to increase collaboration with industry and small business partners through the NavalX Tech Bridges[58] and Air Force Research Lab Innovation Institutes.[59]

Proposed Executive Branch Action

The Secretary of Defense should direct and incentivize the Services to replicate these innovative models at their labs to overcome barriers between the military research community and the wider research environment.[60]

---

[56] See *Army Research Laboratory's Open Campus Effort Forges Ties with Academia, Industry*, Government CIO Magazine (Mar. 15, 2018), https://governmentciomedia.com/joe-mait-ARL-open-campus.
[57] ARL regional hubs have been established in Chicago, Boston, Austin, and Los Angeles. See *Open Campus, Regional Sites*, U.S. Army (last accessed June 16, 2020), https://www.arl.army.mil/opencampus/ARLExtended.
[58] *Spanning the Gap, Tech Bridges*, NavalX (last accessed June 16, 2020), https://www.secnav.navy.mil/agility/Pages/techbridges.aspx.
[59] See *AFRL Innovation Institutes*, Doolittle Institute (last accessed June 16, 2020), https://doolittleinstitute.org/about/afrl-innovation-institutes/.
[60] Notably, the Air Force's 2030 Science and Technology Strategy, released in 2019, includes the objective to "[e]valuate service pilots similar to the U.S. Army Research Laboratory's Open Campus, potentially expanding engagement and formally integrating them into Air Force procedures." *Science and Technology Strategy: Strengthening USAF Science and Technology for 2030 and Beyond*, U.S. Air Force at 18 (Apr. 2019), https://www.af.mil/Portals/1/documents/2019%20SAF%20story%20attachments/Air%20Force%20Science%20and%20Technology%20Strategy.pdf.

Recommendation 11: Create a DoD research and development database.

The DoD research enterprise is a rich, multi-stakeholder environment, with a range of organizations involved. Lack of coordination and communication among R&D efforts disadvantages the community's collective ability to share progress and expertise, build upon each other's work, and accelerate innovation.

While some duplication of effort is desirable, better coordination across the DoD research community and more robust connections with outside researchers would bolster the ability of DoD researchers to move fast and stay on the cutting edge.

DoD should create a searchable database to capture and make available to the enterprise a comprehensive view of ongoing R&D efforts. Such a resource, accessible on secure DoD networks, would provide a mechanism to collaborate and avoid duplication of effort while also enabling data-informed resource decisions, tracking and measurement of R&D investments, and the ability to more deliberately target specific capabilities at the Department level.

The resource should provide detail on projects, including types of data used, as well as points of contact for additional information. Population of the database should be a requirement tied to execution of a funding vehicle or development agreement. The Department of Energy's external-facing lab partnering service portal and internal AI exchange database could provide models for a DoD database.[61]

Proposed Executive Branch Action

The Commission recommends that the Secretary of Defense task OUSD R&E to build a research and development database as an enterprise resource to enable greater return on investment and collaboration across the DoD R&D ecosystem and provide a tool for assessment and data-informed decision-making around research portfolio management.

---

[61] See *Lab Partnering Service Discovery*, Lab Partnering Service, U.S. Department of Energy (last accessed June 16, 2020), https://www.labpartnering.org/search?q=artificial+intelligence.

TAB 2 — Accelerate Artificial Intelligence Applications for National Security and Defense

The United States must identify, develop, and integrate artificial intelligence (AI)-enabled applications for national security and defense faster and more effectively than its competitors. NSCAI's Interim Report assessed that AI is key to the next technological leap that will allow the Department of Defense (DoD) and the Intelligence Community (IC) to understand, operate, and execute their missions faster and more effectively.[62] Making the leap requires broad understanding of how AI can address core national security challenges and what is needed to achieve an AI advantage. Without clear communication linking vision to organizational change, progress and adoption will stall, causing capabilities to fall behind.

In our first quarter recommendations, we offered our strong recommendations on ways to improve DoD's organizational approach to adopting AI-enabled applications by recommending increased senior leader oversight and support for the Department's AI initiatives. This quarter's recommendations focus on accelerating DoD adoption of AI-enabled applications through clear technology development and fielding plans, and greater experimentation.

To maintain advantage, DoD and the IC must have enduring means to jointly identify, prioritize, and resource the AI-enabled applications necessary to fight and win. They also must adapt their traditional approach in order to effectively integrate these technologies into emerging warfighting concepts and operations. To meet this challenge, our recommendations seek to establish: 1) a strategic approach for identifying, resourcing, and ultimately fielding AI-enabled applications that address clear operational challenges; 2) mechanisms for tactical experimentation to ensure technical capabilities meet mission and operator needs; and 3) paths to accelerate adoption of business AI applications essential to institutional agility.

---

[62] *Interim Report*, NSCAI at 30 (Nov. 2019), https://www.nscai.gov/reports [hereinafter Interim Report]. Secretary of Defense Mark Esper supported this assertion in his November 2019 remarks at the NSCAI public conference, stating "Whichever nation harnesses AI first will have a decisive advantage on the battlefield for many, many years." *Remarks by Secretary Esper at National Security Commission on Artificial Intelligence Public Conference*, Department of Defense (Nov. 5, 2019), https://www.defense.gov/Newsroom/Transcripts/Transcript/Article/2011960/remarks-by-secretary-esper-at-national-security-commission-on-artificial-intell/.

Issue 1: A Strategic Approach for Technology Identification and Integration

As the 2018 National Defense Strategy (NDS) highlights, the convergence of new technologies on future battlefields will likely lead to dramatic changes in the character of war. This fact is not lost on America's great power rivals.[63] With its focused, determined, and heavily resourced military modernization, China has made clear its determination to dictate the shape of this emerging revolution in military affairs. China believes AI, big data, swarm intelligence, automated decision-making, along with AI-enabled autonomous unmanned systems, and intelligent robotics will be the central features of the emerging military-technical revolution. The People's Liberation Army (PLA) has developed a warfighting concept for what it calls "intelligentized operations" with AI at its core.[64] Within this construct, China theorizes that in future conflict, the central contest will be between adversarial battle networks rather than traditional weapons platforms, and that information advantage and algorithmic superiority will be a determinant of victory.[65] Russia has established research and development institutes to advance the military applications of AI, and it has already utilized armed systems with autonomous features on the battlefield without regard for ethical considerations. It will likely employ AI to accelerate its hybrid warfare tactics ranging from cyber-attacks to information operations.[66]

We face a situation where we could be outnumbered on the battlefield, denied our preferred method of fighting by our adversaries' capabilities, and consistently behind our adversaries in our understanding of the environment and ability to effectively conduct information operations. To address this, the NDS states that DoD "will invest broadly in military application of autonomy, artificial intelligence, and machine learning, including rapid application of commercial breakthroughs, to gain

---

[63] As China's President Xi has said: "A new technological and industrial revolution is brewing, a global revolution in military affairs is accelerating, and the pattern of international military competition is experiencing historic changes." Chris Buckley & Paul Mozur, *What Keeps Xi Jinping Awake at Night*, New York Times (May 11, 2018), https://www.nytimes.com/2018/05/11/world/asia/xi-jinping-china-national-security.html.

[64] Elsa Kania, *Chinese Military Innovation in Artificial Intelligence*, CNAS at 1 (June 7, 2019), https://www.cnas.org/publications/congressional-testimony/chinese-military-innovation-in-artificial-intelligence (testimony before the U.S.-China Economic Security Review Commission).

[65] The main characteristics of their intelligent combat operations include: intelligent ordnance, intelligent platforms, intelligent systems, intelligent command decision making, intelligent logistics, and intelligent equipment support. See id. See also Elsa Kania, *Learning Without Fighting: New Developments in PLA Artificial Intelligence War-Gaming*, The Jamestown Foundation (Apr. 9, 2019), https://jamestown.org/program/learning-without-fighting-new-developments-in-pla-artificial-intelligence-war-gaming; Elsa Kania, *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power*, CNAS (Nov. 2017), https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power.

[66] Interim Report at 11, 15, 18.

competitive military advantages."[67] In its 2018 AI Strategy, DoD emphasizes the importance of those investments.[68]

To be successful, DoD and the IC must fundamentally embrace and plan for algorithmic warfare—*the notion that a new era of conflict will pit algorithms against algorithms in a contest dominated more by the speed and accuracy of knowledge and action than by traditional military factors*. DoD and the IC must also be able to move at the speed of relevance to gain superiority by adopting, integrating, and iterating on emerging technologies as rapidly as possible. While industry has long realized that the commercial advantage comes from updating and deploying smart algorithms faster than the competition, our government has struggled to evolve and to operationalize our national security strategies to adapt our force.

As the NSCAI's Interim Report indicated, United States Government strategies recognize the importance of emerging technologies such as AI, but struggle to effectively drive implementation—challenged by bureaucratic impediments and inertia.[69] There are multiple processes within DoD to identify requirements and manage their development. However, to drive action, clear guidance is needed to identify, prioritize, and chart a path forward for developing and exploiting emerging technologies such as AI to enable new disruptive capabilities that can help solve critical operational challenges. We offer recommendations to help achieve this goal.

Recommendation 1: As part of the National Defense Strategy (NDS), DoD, with support from the Office of the Director of National Intelligence, should produce a classified technology annex that outlines a clear plan for pursuing disruptive technologies and applications that address the operational challenges identified in the NDS.

A classified technology annex to the NDS focused on development and fielding is more than a simple list of technologies. The annex should identify emerging technologies and applications that are critical to enabling specific capabilities for solving the operational challenges outlined in the strategy. **The main objective of the annex should be to chart a clear course for identifying, developing, fielding, and sustaining those critical emerging and enabling technologies, and to speed their transition into operational capability.** Doing so will advance NDS implementation by connecting strategic vision to priority investments, and ensure technological advances are integrated into future concept

---

[67] *Summary of the 2018 National Defense Strategy of the United States of America*, Department of Defense at 7 (2018), https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf.

[68] The strategy asserts that "[f]ailure to adopt AI will result in legacy systems irrelevant to the defense of our people, eroding cohesion among allies and partners, reduced access to markets that will contribute to a decline in our prosperity and standard of living, and growing challenges to societies that have been built upon individual freedoms." See id. at 5.

[69] Interim Report at 31.

development. Additionally, a technology annex aligned with the NDS will help focus and coordinate the multiple entities within DoD and the IC that each play a part in resourcing, developing, fielding, and iterating such technologies.[70] Better coordination and integration will ensure that both communities can stay abreast of emerging trends and iterate as fast as industry on critical technologies. Even small gains in performance can bring an outsized advantage.

Proposed Executive Branch Action

The Secretary of Defense, with support from the Director of National Intelligence, should develop a comprehensive classified technology annex to the NDS focused on development and fielding by January 2021. The annex should lay out roadmaps for designing, developing, fielding, and sustaining critical technologies and applications necessary to address the specific operational challenges identified in the NDS. DoD should have primary ownership of the document. The Department should also establish a reporting structure and metrics to monitor implementation of the annex to ensure each effort is resourced properly and progressing sufficiently. The annex should be reviewed annually and ensure both guidance and implementation iterate at the pace of rapidly changing technologies.

The technology annex should set clear guidance that drives prioritization and resourcing, while allowing enough flexibility for disparate and decentralized entities to implement that guidance as best suits their organization. At a minimum, the technology annex should include:

- Identified intelligence support requirements, including how the IC analyzes the global environment and monitors technological advancements, adversarial capability development, and emerging threats.
- Identified functional requirements and technical capabilities necessary to enable concepts that address each challenge.
- A prioritized, time-phased plan for developing or acquiring such technical capabilities, that takes into account research and development (R&D) timelines, a strategy for public private partnerships, and a strategy for connecting researchers to end users for early prototyping, experimentation, and iteration.
- Identified additional or revised acquisition policies and workforce training requirements to enable DoD personnel to identify, procure, integrate, and operate the technologies necessary to address the operational challenges.
- A prioritized, time-phased plan for integrating technology into existing DoD exercises that support the NDS, per Recommendation 3 below.
- Identified infrastructure requirements for developing and deploying technical capabilities, including data, compute, storage, and network needs; a

[70] These entities include, but are not limited to: USD(R&E), USD(A&S), CAPE, CIO, CDO, JAIC, SCO, DARPA, the Services, and Combatant Commands; and ODNI, CIA, NSA, NGA, NRO, and IARPA.

resourced and prioritized plan for establishing such infrastructure; and an analysis of the testing, evaluation, verification, and validation (TEVV) requirements to support prototyping and experimentation and a resourced plan to implement them.[71]

- Identified joint capability and interoperability requirements and a resourced and prioritized plan for implementation.
- Consideration of human factor elements associated with priority technical capabilities, including user interface, human-machine teaming, and workflow integration.
- Consideration of interoperability with allies and partners, including areas for sharing of data, tools, and operational concepts.
- Flexibility to adapt and iterate annex implementation at the speed of technological advancement.

Recommendation 2: The Tri-Chaired Steering Committee on Emerging Technology NSCAI recommended in March 2020 should steward the implementation of the technology annex described above.

The Commission proposed a Steering Committee tri-chaired by the Deputy Secretary of Defense, the Vice Chairman of the Joint Chiefs of Staff, and the Principal Deputy Director of National Intelligence to drive innovation and action on emerging technologies. Drafting and implementing the NDS technology annex will require significant policy and investment decisions. While there are multiple complex processes associated with DoD's formal planning, programming, budget, and execution (PPBE) process, implementation of the technology annex would benefit from focused attention and oversight by senior leadership across DoD and the IC. The Steering Committee provides an appropriate forum to manage the annex, with both senior leaders responsible for the process and key technical expertise in its members. The Steering Committee should establish a reporting structure and metrics to monitor the implementation of each technology roadmap to ensure each effort is resourced properly and progressing sufficiently.

Proposed Executive Branch Action

Once established, the Secretary of Defense and the Director of National Intelligence should direct the Tri-Chair Steering Committee to steward implementation of the technology annex described above and establish a reporting structure and metrics to monitor the implementation of each technology roadmap to ensure each effort is resourced properly and progressing sufficiently.

[71] This requirement addresses a preliminary judgment from the Interim Report which asserted that AI is only as good as the infrastructure behind it and that DoD's infrastructure is severely underdeveloped. Interim Report at 33-34.

Issue 2: Integrating AI-Enabled Applications into Military Operations and Tactics

Every military develops doctrine, the foundational principles that describe how it fights wars. Historically, the most successful militaries develop doctrine through rigorous experimentation of potential war fighting concepts designed to address specific military challenges posed by a nation's adversaries. These warfighting concepts begin as unproven theories that propose solutions to military and intelligence problems for which no doctrine exists. For example, the German doctrine of Blitzkrieg began with the idea that a mechanized force could break through defenses more quickly than the defending force could reinforce or counter-attack, providing a potential solution to overcome the stalemate of trench warfare during World War I. As concepts mature, they are prototyped and tested through wargames and military exercises until they are either validated or discarded as ineffective. Those concepts that are validated, like Blitzkrieg, are then codified as doctrine which ultimately drives decisions on how military forces are organized, trained, and equipped for combat.

AI will not only bolster the way militaries have traditionally fought; it will also drive completely new ways of fighting. As a result, integrating AI into concept development is a critical step both in bringing AI-enabled capabilities into use and in enabling next generation military concepts and doctrine. Experimentation and iteration are central to this process. Operators, commanders, and analysts need to understand how these technologies function in practice, how they impact and enable user capabilities, and their overall mission impact in realistic and novel scenarios. Hands-on approaches to test the concepts and technologies, such as wargames, exercises, and fielding prototypes, generate outcomes that both users and senior leaders can see and reference. They also move forward the technology itself by generating data that drives development and iteration to better serve warfighter needs—a critical part of the AI development life cycle. The recommendations below establish mechanisms to create a field-to-learn cycle that generates capabilities and concepts faster, maintaining our military edge.

Recommendation 3: DoD should integrate AI-enabled applications into all major Joint and Service exercises and, as appropriate, into other existing exercises, wargames, and table-top exercises.

To accelerate experimentation and learning, DoD should direct that existing exercises, wargames, and table-top exercises develop plans to integrate AI-enabled applications. This includes large-scale joint exercises and smaller, more frequent events at all echelons. Such exercises align with DoD's development of a Joint Warfighting Concept and Joint All Domain Command and Control. AI will play a critical role in realizing these concepts by enabling connectivity between systems and

sensors, rapid data analysis, faster and more informed decision-making, and more distributed operations.

The exercises should be integrated into the technical annex under Recommendation 1 above and be used to experiment with the most promising concepts and technical capabilities to address the operational challenges articulated in the NDS. The results of the exercises should be reported back to the Tri-Chair Steering Committee to inform policy and resource decisions.

These exercises should include Live Virtual Construct (LVC) environments that allow for user interface with actual systems and experimentation in a realistic simulated environment at a scale not always possible on physical ranges. Developing the testbed infrastructure to support such LVC environments requires significant investment, as outlined in Tab 1 Recommendation 6 of this report. Such infrastructure is essential to effectively developing AI-enabled capabilities, and to supporting the exercises that bring them into the hands of users.[72] Additionally, the R&D community, including the Under Secretary of Defense for Research and Engineering (USD (R&E)) and the Service Labs, should have an active role in the design and execution of the exercises. This enables real-time interaction between developers and operators that allows both communities to better understand the needs of the other, thus allowing technologies to better serve user needs and informing future research.

Proposed Executive Branch Action

The Secretary of Defense should direct all major existing exercises, wargames, and table-top exercises to develop plans to integrate AI-enabled capabilities following the guidance of the technology annex outlined in Recommendation 1. The Tri-Chair Steering Committee should oversee integration plan development, as well as time phasing and resources for implementation.

Exercise / Wargame Objectives related to AI should include:

- Applying AI tools and applications to concrete operational challenges in physical or simulated environments.
- Understanding human-machine and machine-machine teaming dynamics in operational environments.
- Understanding how AI applications augment current processes and capabilities and where they present opportunities for different ways of operating.
- Generating data that supports future development and testing of AI applications.

[72] Configurable instrumentation for data harvesting and collection is also an essential component of this infrastructure.

- Generating data that furthers virtual exercise environments—including visualization, modeling, and simulation capabilities—enabling more realistic and comprehensive exercises at lower cost.
- Capturing lessons learned that inform concept and capability development.
- Integrating allies and partners where appropriate and capturing lessons learned to inform existing multinational exercises and interoperability opportunities, including the sharing of data, tools, and operational concepts.[73]
- Demonstrations of new technologies where full incorporation is not possible.

Recommendation 4: DoD should incentivize experimentation with AI-enabled applications through the Warfighting Lab Innovation Fund, with oversight from the Tri-Chaired Steering Committee.

DoD should incentivize experimentation with AI applications across the Department at every level possible. The Warfighting Lab Innovation Fund (WLIF), established in 2016, is one existing mechanism to do so, with the express intent to "spur field experiments and demonstrations to evaluate, analyze and provide insight into more effective ways of using current capabilities, and to identify new ways to incorporate technologies into future operations and organizations."[74]

The structure and content of the proposals are classified; however, it can be noted that: 1) proposals can be submitted by the "Service Warfighting Labs, Combatant Commands (CCMDs), Joint Staff, Office of the Secretary of Defense (OSD), Defense Agencies, Federally Funded Research and Development Centers, University Affiliated Research and Development Centers, and the Defense Industrial Base;" and 2) proposals must have a "warfighting sponsor (CCMD, Service and/or Defense Agency)" and a plan to "transition to operational capability."[75]

Currently, the Joint Staff and the Director of Cost Analysis and Program Evaluation (CAPE) evaluate and prioritize WLIF proposals annually and provide execution reports to the Deputy Secretary of Defense and Vice Chairman of the Joint Chiefs of Staff highlighting insights from each exercise they fund. The Joint Staff and CAPE should present their prioritized list of proposals for the upcoming year as well as the execution reports from previous exercises directly to the Tri-Chair Steering Committee for approval and guidance as part of a standing Committee process. As a standing member of the Committee, the Director of the Joint Artificial Intelligence Center (JAIC) should provide guidance on prioritization of funding requests to incorporate AI-enabled applications. This process will give the Committee greater

---

[73] NSCAI's first quarter recommendations propose an AI Wargame and Experimentation Series with allies and partners. See *First Quarter Recommendations*, NSCAI at 67, (Mar. 2020), https://www.nscai.gov/reports.

[74] *Warfighting Lab Innovation Fund*, Defense Innovation Marketplace (last accessed May 28, 2020), https://defenseinnovationmarketplace.dtic.mil/business-opportunities/warfighting-lab-incentive-fund/.

[75] See id.

visibility on innovation across the force, informing policy and resourcing decisions. Insights from experimentation across the force would also inform decisions on how to integrate AI-enabled applications into exercises and wargames.

DoD should either a) develop a special category within the existing WLIF to provide funding augmentation to any qualifying entity who wishes to incorporate AI applications into existing exercises or wargames, or b) incorporate AI applications as one of the prioritized evaluation criteria. Either of these options would incentivize efforts to get AI applications into the hands of users to accelerate the lab-to-field transition. The Department's current evaluation criteria for WLIF proposals are:

- Potential for disruptive innovation;
- Potential contribution to off-set key US vulnerabilities;
- Potential for cost imposition/enhancements to US national interests across conflict continuum;
- Potential cost/benefit for the Department;
- Amount of funding requested;
- Time required to execute and generate results;
- Potential for advancing US national interests (e.g., improving ally integration); and
- Past performance of requesting organization.[76]

Proposed Executive Branch Action

The Tri-Chair Steering Committee should have oversight of the WLIF and should establish either a special category or prioritized evaluation criteria for proposals that incorporate AI applications in their proposal to incentivize experimentation with AI applications throughout the Department. WLIF funds should also be provided to incentivize and enable the integration of AI-enabled applications into exercises and wargames as outlined in Recommendation 3.

Issue 3: Business AI Applications

Institutional agility can provide warfighters and intelligence professionals with a competitive edge that allows them to adapt faster than their adversaries. AI-enabled capabilities are as vital in an institutional setting as they are on the battlefield because they support the systems behind the mission. Yet, DoD enterprise systems often struggle when faced with complexity and the need for speed, failing to keep pace with technological change, adaptive adversaries, and complex emergencies. The institutional functions of DoD are hindered by outdated business processes and systems; the department must modernize to become more effective and cost-efficient.

[76] Memorandum from Deputy Secretary of Defense, to Director, Cost Assessment and Program Evaluation, et al., (May 6, 2016), https://defenseinnovationmarketplace.dtic.mil/wp-content/uploads/2018/02/DSD_memo.pdf.

Proven off-the-shelf commercial AI solutions can make core processes such as human resources, financial management, contracting, and logistics more efficient and cost-effective. Business process modernization will contribute to institutional agility through faster, evidence-based decision-making across the range of national security agencies and missions, supported by automation of simple, repetitive tasks. Greater institutional agility will require structural, cultural, and process changes that go well beyond new software; however, business process modernization is a critical first step.

Recommendation 5: DoD should develop a prioritized list of core administrative functions that can be performed with robotic process automation and AI-enabled analysis and take specific steps to enable implementation.

The NSCAI's Interim Report noted that DoD is not adequately leveraging basic commercial AI to improve business practices and save taxpayer dollars.[77] Robotic process automation and AI-enabled analysis can generate significant labor and cost savings, speed administrative actions, and inform decision-making with superior insights into core DoD business processes. To realize these benefits, DoD should initiate the digital transformation of its core administrative functions and assign responsibility to a senior DoD executive, such as the Chief Management Officer or a similar senior official. The Department should begin this process by assembling enterprise-wide datasets that will allow effective training and deployment of AI algorithms.

The current state of data governance within DoD (and government writ large) includes numerous overlapping and conflicting regulations and policies for the collection, storage, and sharing of data that would impose insurmountable procedural obstacles and delays on efforts to build enterprise datasets. This will require a coordinated top-down effort to modernize data governance. The effort should leverage AI technology to analyze the corpus of governance documentation and develop new streamlined rulesets.

Once policy obstacles have been overcome, significant resources will be required to access, clean, and label enterprise data from the range of legacy business platforms. These will include skilled data engineers, cloud and high-performance computing, data labeling software, contract vehicles to secure these resources, and end users who are conceptually grounded in the principles of data science.

As DoD gains experience deploying commercial AI and builds a workforce of internal AI developers,[78] it will need to invest in further classes of commercial AI applications for generating bespoke AI solutions. These include: 1) data preparation

[77] Interim Report at 34.
[78] DoD's AI workforce requirements are addressed in the NSCAI's Nov 2019 Interim Report (pp. 61-65) and March 2020 First Quarter Recommendations (pp. 30-31). They will also be developed further in future NSCAI recommendations.

and labeling applications; 2) model building, compilation, and maintenance applications; 3) imaging applications for object recognition and anomaly detection; and 4) language applications including advanced methods for speech recognition, machine translation, and text to speech.

Proposed Executive Branch Action

DoD should prioritize dataset construction across the following DoD business administration areas: human resources, budget & finance, logistics, retail, real estate, and health care, assigning a lead organization with primary responsibility for developing enterprise-wide datasets.[79] Prioritization should initially go to processes that directly support the DoD audit. The Secretary of Defense should issue a department-wide directive for DoD agencies to proactively provide all requested data to the lead organizations and participate in subsequent modernization efforts. The directive should: 1) mandate deconfliction and/or removal of policies and regulations that prevent rapid and effective data sharing across agencies; and 2) provide funding for contracting with commercial data engineering services.

Recommendation 6: DoD should incentivize deployment of commercial AI applications across the organization for knowledge management, business analytics, and robotic process automation.

In addition to the top-down business AI initiatives described above, DoD should create opportunities for bottom-up development of AI business use cases by incentivizing entities across the organization to deploy proven commercial applications tailored to their specific requirements. This bottom-up approach is useful for AI application areas in which the heterogeneity of defense agency, Service, and Component missions and workforces are likely to require bespoke software tools vice DoD-wide solutions. A mechanism for DoD to provide matching funds and technical support should be employed to incentivize and facilitate participation.

Promising categories of commercial AI include: 1) knowledge management applications such as intelligent search tools that index, retrieve, and display an agency's digital information, as well as collective intelligence and coaching tools that accumulate and exchange tacit knowledge across an agency's workforce; 2) AI-enabled tools that analyze business information to identify patterns, develop insights, and inform decision-making, and 3) robotic process automation tools including desktop assistants, bots, and other personal productivity applications that automate individual office functions.

Like the implementation challenges cited above—including, data governance, data labeling, and model training, optimization, and compilation—there are major regulatory and policy obstacles to acquiring and deploying new software. A top-down

[79] Dataset development should proceed in concert with the data cataloging process outlined in Tab 1, Recommendation 3 of this memorandum.

effort to modernize software governance should be pursued in parallel with data governance reform.

Proposed Executive Branch Action

DoD should launch a department-wide initiative to rapidly deploy commercial AI solutions for knowledge management, business analytics, and robotic process automation across the Department, defense agencies, Services, and Combatant Commands supported by matching funds from DoD. The Department should assign a lead organization to administer allocation of matching funds, monitor and assess results, and disseminate best practices and lessons learned. The initiative should include: 1) a DoD directive mandating deconfliction and/or removal of policies and regulations preventing rapid acquisition and deployment of commercial AI software, and;[80] and, 2) technical support to build, train, tune, and deploy new AI models; and 3) provision of matching incentive funds for agency contracting with commercial AI vendors.

---

[80] This should specifically include steps to further promote and recognize the authorization to operate process, as described in Tab 1, Recommendation 2 of this report.

TAB 3 — Improve the United States Government's Digital Workforce

"In a strategic competition," the Commission's Interim Report noted, "advantage will go to the competitor that can best attract, train, and retain a world-class, AI-ready workforce. Currently, there is a severe shortage of AI knowledge in DoD and other parts of government."[81] It has only become more apparent to us that the United States Government needs to become a digitally proficient enterprise. Current initiatives are helpful, but only work around the edges, and are insufficient to meet the government's needs. Bolder steps are needed. We must fundamentally re-imagine the way we recruit and build a digital workforce. Agency-specific models have proven inadequate and inefficient. The Commission envisions a government-wide effort to build a digital workforce.

Given the government's general shortage of digital talent, in this second quarter of 2020 the Commission recommends multiple avenues for addressing that need: reduce the challenge of part-time government service by creating a National Reserve Digital Corps (NRDC), train the next generation by building a United States Digital Service Academy (USDSA), and expand current scholarship-for-service programs.

Combined, the recommendations will increase the government's digital literacy by expanding and creating pathways for technical experts to serve in government as part-time or full-time employees. The NRDC, modeled after the military reserves, will create a mechanism for technical experts to serve in government part-time. The expansion of scholarship-for-service programs will increase the number of recent graduates with technical backgrounds that join the government as full-time employees. The USDSA will create a new source of civil servants with technical knowledge, and serve as a mechanism for government modernization.[82]

Issue 1: Providing AI Practitioners with Part-time Options for Government Service

The Commission's First Quarter Recommendations addressed issues related to hiring, establishing a baseline of AI knowledge among public servants, identifying existing talent within the government workforce, building recruitment pipelines from universities to the government, and creating temporary talent exchanges between the

[81] *Interim Report*, NSCAI at 35 (Nov. 2019), https://www.nscai.gov/reports.

[82] All three recommendations would produce scholarship recipients or academy graduates with a three- to five-year service obligation. They would begin their service as a GS-7, and advance to GS-11 by the end of their obligation with the potential to continue within the competitive service.

government and the academic and private sectors.[83] Those recommendations, while an important start, focused on people who are interested in becoming full-time government employees, and therefore will not affect a significant portion of the United States' overall digital talent pool. While there are digital experts who are willing to work for the government for a full career, and others who will serve full-time for several years, the United States Government needs a better way to tap into the expertise of those who would like to contribute to American national security but are unwilling or unable to become full-time government employees or military reservists. Our conversations with industry experts and academics have indicated that many would be interested in contributing to government missions, either because of a sense of civic responsibility or an interest in unique government missions, but do not want to leave their career field even temporarily to do so.

Recommendation 1: Create a National Reserve Digital Corps.

The United States Government should establish a civilian National Reserve Digital Corps (NRDC) modeled after the military reserves' service commitments and incentive structure. Members of the NRDC would become civilian special government employees (SGEs),[84] and work at least 38 days each year as short-term advisors, instructors, or developers across the government.[85] Longer-term positions would be established on an individual basis. While short-term volunteers are not a substitute for full-time employees, they can help improve AI education for both technologists and non-technical leaders, perform data triage and acquisition, help guide projects and frame technical solutions, build bridges between the public and private sector, and other important tasks.[86]

The government would benefit from access to a larger portion of the country's total digital workforce. Many government digital projects suffer from lack of access to digital expertise. Several AI practitioners within the United States Government have said during interviews with the NSCAI that their projects would benefit from the kind of reserve corps we propose here.

---

[83] *First Quarter Recommendations*, NSCAI at 21-43 (Mar. 2020), https://www.nscai.gov/reports [hereinafter First Quarter Recommendations].

[84] A special government employee is "an officer or employee of the executive or legislative branch of the United States Government . . . who is retained, designated, appointed, or employed to perform, with or without compensation, for not to exceed one hundred and thirty days during any period of three hundred and sixty-five consecutive days." 18 U.S.C. § 202.

[85] Members of the military reserves typically serve two to three days a month, and one 14-day obligation a year, averaging around 38 days a year.

[86] Organizations that employ full-time technical experts in temporary positions, such as the United States Digital Service or Defense Digital Service, already exist, and have proven successful. The NRDC is an alternative for experts that cannot or do not want to pursue a full-time route.

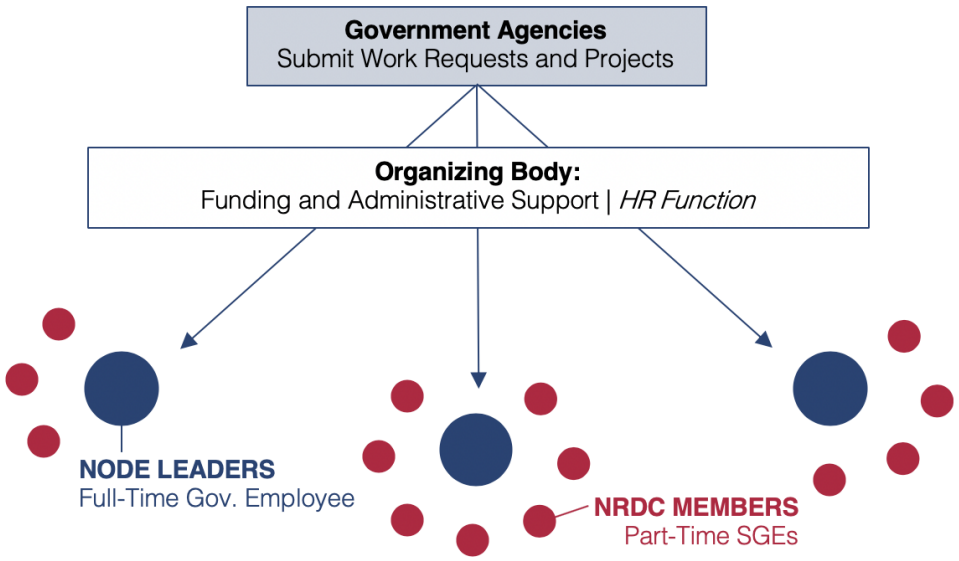

Figure 1: Illustrating the NRDC

General Structure. — We recommend establishing and managing the NRDC as a set of nodes that fall under the supervision of the Office of Management and Budget (OMB). Each node would be aligned with a full-time government employee leader selected by OMB rather than geography, digital applications, or government agency. In effect, OMB would select node leaders, who would then be responsible for recruiting and organizing their team. In addition to selecting node leaders, OMB would establish standards, ensure nodes meet government client requirements, provide funding and administrative support, maintain security clearances, establish access to an agile development environment and tools, and facilitate technical exchange meetings, when appropriate, to ensure stovepipes are not created.

Recruitment. — Each node would be responsible for recruiting and screening its digital experts. Notably, OMB would not be responsible for establishing qualification standards for volunteers. While volunteers would need to be able to pass a background check and would not be employees of a foreign government (though they might be foreign nationals), node leaders would be empowered to screen and select volunteers, and to recruit experts from within NRDC for specific tasks. OMB would provide administrative support, much like a human resources team in a private sector company.[87]

[87] Some administrative functions, such as background checks, security clearance processing, processing tax paperwork, and others, would place an unnecessary burden on local nodes and should be addressed by a central body such as OMB.

Project Selection. — Projects would be selected in three ways:

- Selection by a node after contact with a government client,
- OMB would direct a node to take on a project, and
- Node leadership would approve individual projects driven by a perceived need that is not tied to a request from a government client.

Government clients would directly contact node leaders or OMB. Nodes would be responsible for establishing relationships with government agencies and selecting projects, but OMB would be responsible for ensuring that agencies' requests are received and that nodes contribute to NRDC's mission and vision.[88] Individual projects that are not driven by a government client's request would be pursued at the node leadership's discretion.

Relationship with Government Agencies. — Members of the NRDC would work with agencies on a project-to-project basis— such as consulting for a specific project or teaching a specific course. They would not have a commitment to work with the same agency consistently. Government agencies would be responsible for paying for their projects, including the cost for reservist time.

Relationship with Civilian Employers. — Members of the NRDC and their civilian employers would be bound by the same rules as the military reserve under the Uniformed Services Employment and Reemployment Rights Act (USERRA).[89] Members would be responsible for identifying conflicts of interest and removing themselves as appropriate. Employers would not be able to discriminate against members of NRDC, fire them, or delay promotions as a consequence of spending time serving in NRDC.[90] Implementation could take the form of a legislative recommendation to modify USERRA or a proposal modeled after USERRA.

Incentivizing Reservist Participation. — Civilian reservists in this program would benefit in several ways. They would gain an opportunity to contribute to their communities, do exciting, meaningful work, and attain awareness of work and advances in a community that differs from their own. They may also benefit from the following incentives:

- The government should create an NRDC scholarship program modeled after the Reserve Officer Training Corps. Universities would select students through a competitive process to receive full tuition and study specific disciplines related to national security digital technology. In return for

---

[88] While agency requests should not be ignored, this does not assume that all agency requests will be reasonable, feasible, or accomplishable by NRDC personnel.
[89] *Uniformed Services Employment and Reemployment Rights Act of 1994*, Department of Justice (Aug. 6, 2015), https://www.justice.gov/crt-military/userra-statute.
[90] Frank Whitney, *Employment Rights of the National Guard & Reserve*, Department of Justice, https://www.justice.gov/sites/default/files/usao-ednc/legacy/2011/04/29/EmploymentRights.pdf.

accepting the scholarship, graduates would spend part of their summers during school in government internships. Between their freshman and sophomore years, students would spend six weeks becoming familiar with a range of United States Government departments and agencies. Between their sophomore and junior years, students would spend six weeks as an intern at a specific government agency or office. Between their junior and senior years, students would spend another six weeks interning at a specific agency or office. Upon graduation, scholarship recipients would spend five years serving in the NRDC, beginning as a GS-7 and advancing to a GS-11 over the course of five years. Students would also begin the security clearance process at least two years before graduating.[91]

- The NRDC should include a training and continuing education fund for all members. The NRDC would pay up to $50,000 to each reservist to attend training and educational opportunities related to AI or to pay for student loans. Educational opportunities would include conferences, seminars, degree and certificate granting programs, and other opportunities. An incentive explicitly tied to continuing education would increase the perceived and actual competency of AI reservists. It would also attract those with an active interest in continuing education, especially new practitioners seeking to establish themselves.

How NRDC Would Work: An Example. — The following is a hypothetical example of how the NRDC would function. In this example, OMB would begin creating a node by selecting a leader that would be trusted to establish and manage a team of reservists. OMB selects "Jennifer," a full-time government employee working within the NRDC division of OMB, to lead a new NRDC node. Jennifer decides to organize her node functionally rather than regionally. Using existing government tools and her professional contacts, she recruits people from across the country, most of whom have backgrounds in healthcare data management or recent graduates with degrees related to the field. She also recruits from within the NRDC by posting open positions on online job boards. During the recruitment process, OMB provides financial support for recruitment efforts, travel money, and processes new reservist administrative paperwork, including security clearance applications.

After the node is established and the team is in place, a government agency—in this example, the Centers for Disease Control and Prevention (CDC)—realizes it has two digital needs it cannot meet internally: improving a database and training their workforce in new data management practices at the National Center for Chronic Disease Prevention and Health Promotion. After reaching out to OMB, they determine that Jennifer's node is the best fit, and request assistance. After examining the request and her team's workload, Jennifer determines that she would support the

[91] All reservists would apply for security clearances, but this should not imply that reservists would work primarily on classified materials. A large part of the work needed to modernize the government is unclassified.

CDC's database improvement request with a four-person team and support workforce training with a two-person team. The four-person team spends 14 days examining the existing database and making updates to the database. The two-person team spends ten days on site at the National Center for Chronic Disease Prevention and Health Promotion speaking with leaders and employees about their data management needs and the current state of the workforce's skill level, developing curriculum, and teaching data management best practices.

The teams Jennifer selects to support the CDC include Michael. Michael received a four-year scholarship from NRDC to study computer science as an undergraduate. After graduating three years ago, he began working full-time as a data analyst at a healthcare company and working part-time on NRDC projects he coordinates with his node leader. He also used his education stipend to pay for an online course from MIT last year. This hypothetical shows that an NRDC can effectively increase the U.S. digital talent, connect private-sector workers with a government agency, and create a pathway for that connection to solve an actual problem.

Proposed Legislative Branch Action

Congress should pass legislation establishing the NRDC within OMB. In this legislation, OMB should be granted direct-hire authorities to hire node leaders and reservists.

The NRDC should offer full tuition scholarships to students studying specific disciplines related to national security digital technology for up to four years in exchange for five years of service as a member of the NRDC. This could be done by including service in the NRDC as an option for people with degrees in digital fields to pay off service obligations incurred as a result of education received in the Defense Civilian Training Corps.[92]

Legislation should authorize up to $50,000 in educational benefits for courses, seminars, conferences, and other educational opportunities that are approved by OMB. It should also ensure that members of the NRDC receive the same employment protections as military reservists under USERRA. This can be done by amending USERRA to cover "service in the uniformed services or the National Reserve Digital Corps."

Congress should use three metrics to evaluate NRDC's success: 1) the number of technologists who participate annually; 2) evaluations of results from government clients; and 3) evaluations of results from reservists. OMB should establish the central, organizing function for the NRDC within six months of the passage of legislation, and establish five nodes and a mechanism for distributing educational benefits within nine months of the passage of legislation.

---

[92] The Defense Civil Training Corps was created by the National Defense Authorization Act for Fiscal Year 2020. Pub. Law 116-92 §860 (2019).

Congress should make a two-year appropriation of $16 million to pay for initial administrative, scholarship, and education benefits.

Proposed Executive Branch Action

Immediately upon receiving authority from Congress, OMB should establish a National Reserve Digital Corps. OMB would be responsible for: selecting and hiring node leaders; ensuring government client needs are met by NRDC nodes; providing funding for education supplements and scholarship programs; providing administrative support (including for security clearances); establishing node access to development environments and tools; facilitating technical exchange meetings; and matching recipients of NRDC scholarships with node leaders.

At the outset, OMB should establish five NRDC nodes. Each node leader should be responsible for recruiting and hiring reservists, ensuring the quality of their work, and for partnering with government agencies. OMB should also encourage potential government clients to contact NRDC nodes, or OMB, with potential problems to resolve.

Issue 2: Scaling Digital Talent Across the Government Workforce

Digital proficiency requires greater expertise within the government across many disciplines, including cybersecurity, AI, network architecture, software engineering, data science, computer science, mathematics, robotics, and others.

A shortage of digital expertise impacts national security. The Deputy Assistant Director for Cybersecurity at the Department of Homeland Security said in November 2019 that the state of the cybersecurity workforce "is going to be a national security issue, if it isn't already."[93] The Commission's research has shown that many United States Government departments and agencies do not have the talent they need to modernize at the speed of technological change in the private sector and academia. Even when an agency has a modern digital system, it does not have the workforce needed to use the system effectively. This lack of talent is even more severe than it might seem at first glance. While the United States Government's digital workforce is already smaller than needed, its requirements are only going to increase as digital technology and data-driven systems become even more important.

[93] Maggie Miller, *Senior Official Describes Cyber Workforce Shortage as National Security Threat*, The Hill (Nov. 12, 2019), https://thehill.com/policy/cybersecurity/470117-senior-official-describes-cyber-workforce-shortage-as-national-security.

The talent deficit extends beyond the United States Government's workforce to the nation as a whole. As of January 2019, according to one estimate, the United States needed 314,000 additional cybersecurity professionals to meet the market's needs—a number that has grown more than 50 percent since 2015.[94] This deficit is even more severe in AI, and is projected to become worse over the next decade.[95] The resulting competitiveness of the job market will only exacerbate the United States Government's struggle to recruit and retain digital talent.

The United States Government has begun taking measures to address this issue. It has introduced a broad array of hiring authorities, internships, and scholarships. The CyberCorps: Scholarship-for-Service (C:SFS) program and the Science, Mathematics and Research for Transformation (SMART) Scholarship-for-Service program both have recruited digital talent. These programs and others like them, while beneficial, will not be sufficient for at least two reasons.

First, they do not produce a sufficient number of government employees. Between 2015 and 2019, C:SFS produced an average of 275 graduates a year.[96] Between 2016 and 2019, SMART produced an average of 315 graduates a year.[97] In 2019, approximately 51 percent of SMART scholarship awardees studied digital disciplines.[98] These programs are significant, but they do not produce enough graduates to achieve the enterprise change needed in the government.

Second, scholarship programs send new employees into the government without a common set of ideas or intent to help the government modernize. By contrast, military officers in each service have a common set of commissioning requirements, and within their specializations, complete the same training. The relationships and culture built into training helps those military officers shape their institutions. The lack of a single or even small number of institutions that produce a large number of

---

[94] William Crumpler & James Lewis, *The Cybersecurity Workforce Gap*, Center for Strategic & International Studies at 1 (Jan. 29, 2019), https://www.csis.org/analysis/cybersecurity-workforce-gap.
[95] Jacques Bughin, et al., *Skill Shift: Automation and the Future of the Workforce*, McKinsey & Company at 20 (May 2018), https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Future%20of%20Organizations/Skill%20shift%20Automation%20and%20the%20future%20of%20the%20workforce/MGI-Skill-Shift-Automation-and-future-of-the-workforce-May-2018.ashx. The report shows there is a present and growing mismatch between the market's needs and the workforce's skill set, particularly a deficit of specialized information technology workers and data scientists.
[96] Documents produced by CyberCorps: Scholarship for Service officials on March 9, 2020.
[97] *SMART Scholarship Program, Program Stats*, Department of Defense, (last accessed June 17, 2020), https://smartscholarshipprod.service-now.com/smart?id=kb_category&kb_category=6242a353dbbd0300b67330ca7c9619b9 [hereinafter SMART Program Stats].
[98] *SMART Scholarship Program, 2019 Award Statistics*, Department of Defense, (last accessed June 17, 2020), https://smartscholarshipprod.servicenowservices.com/smart?id=kb_article&sys_id=c5e2a163db6f3340e6e530dc7c9619c3. In this case, digital fields include computer and computational sciences and computer engineering, electrical engineering, information sciences, mathematics, and operations research.

graduates with shared experiences, professional culture, and a common mission to improve the government's digital technology is an impediment.

Recommendation 2: Expand Scholarship for Service Programs.

While today's scholarship-for-service programs do not produce a sufficient number of employees with digital expertise for the United States Government, they have been somewhat successful within their current mandates. C:SFS boasts a 92-95% placement rate, has over 70 active institutions participating, and has placed approximately 3,600 graduates in over 140 government institutions since 2001.[99] SMART has a similarly successful record, having awarded 1,262 scholarships from 2016 to 2019.[100] With more funding, scholarship-for-service programs could quickly increase the digital talent in government service.

However, expanding C:SFS and SMART will not be enough. C:SFS is focused on cyber skill sets, and does not address other important digital skills. SMART is broader but is focused on the DoD. Due to AI's increasing importance for cyber operations, the Commission previously recommended expanding C:SFS to include "digital engineering."[101]

Proposed Executive Branch Action

The Office of Personnel Management and the National Science Foundation (NSF) should expand CyberCorps: Scholarship for Service by an additional 85 scholarships per year. The DoD should expand SMART: Scholarship-for-Service to award an additional 100 scholarships per year. Both programs should increase their focus on AI. The NSF and DoD should create an opportunity for scholarship recipients to transition to the NRDC upon completing their service obligation.

Proposed Legislative Branch Action

Congress should appropriate an additional $6 million for CyberCorps: Scholarship for Service and an additional $7 million for SMART: Scholarship-for-Service.[102]

---

[99] *History/Overview*, CyberCorps: Scholarship for Service (last accessed June 17, 2020), https://www.sfs.opm.gov/Overview-History.aspx.

[100] See SMART Program Stats.

[101] First Quarter Recommendations at 39. We used the term "digital engineering" as it was defined by the National Defense Authorization Act for Fiscal Year 2020. Pub. Law 116-94, §230. The Commission also recommended an increase of $100 million in funding for fellowships managed by DoD, Department of Energy, National Aeronautics and Space Administration, and National Science Foundation, which includes C:SFS and SMART. See id. at 10.

[102] The two programs cost approximately $70,000 per student per year. Therefore, an additional 85 students for CyberCorps: Scholarship for Service and 100 students for SMART: Scholarship-for-Service would cost roughly the amount of the proposed appropriation.

Recommendation 3: Create a United States Digital Service Academy.

The United States needs a new academy to train future public servants in digital skills. Our proposed United States Digital Service Academy (USDSA) would be an accredited, degree-granting university that receives government funding,[103] be an independent entity within the Federal government, and have the mission to help meet the government's needs for digital expertise. It would be advised by an interagency board that would be assisted by a federal advisory committee composed of commercial and academic leaders in emerging technology.

Existing Models: The Military Service Academies. — The USDSA should be modeled off of the five U.S. military service academies but should produce trained government civilians not only to the military departments, but also to civilian departments and agencies beyond DoD.[104]

The five military service academies each produce commissioned officers for the armed forces.[105] The academies select cadets and midshipmen through a congressional and presidential nomination process, followed by a competitive admissions process. The cadets and midshipmen, who are government employees, exchange a commitment to serve after graduation for a tuition-free education. Many choose this path for the opportunity to serve; the free tuition and education often are considered a bonus. Those who depart prior to meeting the minimum requirements for graduation still incur either a service commitment or financial requirement to pay back education received upon their departure from the schools.

The academies contribute between 15 and 20 percent of the new junior officers to their respective services each year—the largest single commissioning source. Academy graduates also play an outsized role in the military services' senior leadership.[106] As a result, the academies help shape the identity and culture of their services, including their standards and ethical norms.

USDSA would be comparable to the other service academies in many ways. It would be a degree granting institution focused on producing leaders for the United States Government. USDSA students, like military service academy students, would not

---

[103] The USDSA should also have gift authority, particularly to help fund its establishment.
[104] The Council on Foreign Relations report, *Innovation and National Security: Keeping Our Edge*, recommends creating a digital military service academy. James Manyika & William McRaven, *Innovation and National Security: Keeping Our Edge*, Council on Foreign Relations (Sept. 2019), https://www.cfr.org/report/keeping-our-edge/. Our recommendation is for a civilian digital service academy that would not produce any uniformed military personnel.
[105] The five academies include the United States Military Academy, the United States Naval Academy, the United States Coast Guard Academy, the United States Merchant Marine Academy, and the United States Air Force Academy.
[106] Joseph Moreno & Robert Scales, *The Military Academies Strike Back*, The Chronicle of Higher Education (Nov. 12, 2012), https://www.chronicle.com/article/The-Military-Academies-Strike/135600. As an example, 5 Secretaries of the Navy, 29 Chiefs of Naval Operations, and nine Commandants of the Marine Corps graduated from the United States Naval Academy.

pay for tuition, or room and board, and would have a post-graduation service obligation. Americans should expect USDSA graduates to seek to serve, to lead the nation's digital workforce, and to ensure the United States sets an example of intelligent, responsible, and ethical high-tech leadership.

Key Differences Between USDSA and the Military Service Academies. The USDSA would differ in significant ways. First and foremost, USDSA students would enter the institution to become civil servants. They would know that their education would be repaid in the form of a five-year obligation to serve in government, which would begin upon graduation when they become a civil servant at a GS-7 level. Exclusively producing civil servants would eliminate the need for students to complete commissioning requirements, simplifying the school's curriculum and administrative burdens, and reduce the need for expansive campus real estate for training and parade grounds. It would also make USDSA less redundant, as the military service academies already produce hundreds of computer scientists, electrical engineers, and mathematicians every year.

USDSA students would also have a more STEM-focused education. While the core curriculum would ensure broad exposure to different fields, students would have a highly technical education. A wide variety of technical majors could include AI, software engineering, electrical science and engineering, computer science, molecular biology, computational biology, biological engineering, cybersecurity, data science, mathematics, physics, human-computer interaction, robotics, and design. Students could also blend those majors with humanities and social science disciplines such as political science, economics, ethics and philosophy, or history.

A third difference would be that USDSA graduates would serve across the Federal government. To avoid both perceived and real parochial bias from the organizations that administer service academies, USDSA would be administered as an independent Federal entity. The minimum and maximum number of graduates who would serve in each department or agency would be determined annually by an interagency board.[107]

Mission Statement of the USDSA. — We propose the following: "The United States Digital Service Academy's mission is to develop, educate, train, and inspire digital technology leaders and innovators and imbue them with the highest ideals of duty, honor, and service to the United States of America in order to prepare them to lead in service to our nation."

[107] Each military service academy has a maximum and minimum number of positions available for every available career field, causing some graduates to receive career fields other than their first choice. Similarly, USDSA graduating classes would have a minimum and maximum number of civilian graduates that join each military department or government agency.

The Student Experience. — During their first year, students would begin the Academy's core curriculum, explore some electives to help determine their major, and take a summer internship or fellowship. The core curriculum is envisioned to include, among other things, American history, government, and law, as well as composition, mathematics, computer science, and the physical and biological sciences. Once summer arrives, students would participate in summer internships with private sector companies.

Students would select their major early in their second year, begin concentrating on their technical field, and continue their core curriculum. They would also initiate their security clearance application process. The goal would be for all students to graduate with at least a secret clearance. After completing the classroom portion of their second year, students would complete internships in two government agencies, which would help them focus their goals for government service.

During their third year, USDSA students would increase the focus on their major, complete the majority of their core curriculum, and begin committing to a government agency. Similar to the military service academies, attendance of the first day of class at the start of their third year serves as a commitment to five years of government service upon graduation. After completing the classroom portion of the third year, students would participate in another private sector internship.

Students would commit to a particular government agency and career field during the first weeks of their fourth year and begin the job placement process. To select a department and career field, students would create a rank ordered list of career fields within departments, agencies, and services. The USDSA would then match student preferences to the government's needs as identified by an annual interagency process. After successfully completing all academic requirements, students would graduate as GS-7s, with the potential to progress rapidly to GS-11. After completing their service obligation, USDSA graduates would have the opportunity to transition to the NRDC.

Accreditation. — In order to receive federal funding, the USDSA would take the required steps to complete the accreditation process through a regional accreditation organization. The accreditation organization would be determined based on the physical location of the institution and recognized by the Department of Education and Council for Higher Education Accreditation.[108] Membership in such an organization ensures academic quality throughout the institution's lifespan, as accreditation requires ongoing assessment for improvement. Future employers are able to affirm the credentials of USDSA graduates, the academy is able to accept charitable donations, and post-graduate programs recognize the validity of undergraduate degrees through accreditation. Based on the location of USDSA, the

[108] The military service academies are accredited by different regional accreditation organizations recognized by the U.S. Secretary of Education and Council for Higher Education. Their engineering programs are generally accredited by the Accreditation Board for Engineering and Technology, Inc.

institution would also work with the hosting state to determine compliance with all core standards and processes.[109]

Proposed Implementation Plan for the USDSA:

Phase One (Years 1-2)

- Identify and secure an appropriate site for initial USDSA build-out with room for future expansion.
- Identify gaps in the government's current and envisioned digital workforce by an interagency task force under Office of Personnel Management leadership.
- Establish the USDSA administration as a new Executive Branch agency with an individual appropriation that will be responsible for the phased implementation plan and the management of the institution.
- Recruit tenure-track faculty.
- Recruit adjunct faculty, primarily from private-sector technology companies.[110]
- Grant the USDSA the authority to accept outside funds and gifts from individuals and corporations for startup, maintenance, and infrastructure costs.
- Appropriate $40 million to pay for administrative costs.
- Satisfy the necessary requirements set by the Department of Education as well as the state USDSA is in for degree-granting approval.
- Apply for degree program specific accreditation through Computing Accreditation Commission on Colleges of Accreditation Board for Engineering and Technology.[111]
- Apply for accreditation with a Regional Accrediting Organization approved by the Department of Education and Council for Higher Education Accreditation in order to be granted "Candidate" status.
- Construct initial physical infrastructure.
- Appropriate additional costs for the selection and purchase of the physical location and construction of infrastructure.

[109] State approval and accreditation are not the same, but both are required.
[110] Recruitment will rely on private-sector champions to recruit high-profile adjunct faculty that can serve as beacons that will attract additional faculty and high-quality students.
[111] A nonprofit, ISO 9001 certified organization that accredits college and university programs in applied and natural science, computing, engineering and engineering technology.

Phase Two (Years 3-5)

- Begin classes with an initial class of 500 students at the beginning of year three.[112]
- Demonstrate compliance with all requirements and standards of the regional accrediting organization in order to be granted Membership status.

Phase Three (Years 6-7)

- Graduate the first class.
- Ongoing improvement through accreditation assessments.
- Assess, and as appropriate, expand class sizes.

Proposed Legislative Branch Action

Congress should authorize the establishment of the USDSA as an independent entity with a mandate to establish the institution described above. Congress should also appropriate $40 million dollars over two years to pay for the USDSA's initial administrative costs.

Proposed Executive Branch Action

Immediately upon receiving authorization from Congress, the Executive Branch should act on authorization from the Congress to establish the USDSA as an independent Federal entity with a mandate to establish the institution described above. While the agency is being established, the Office of Personnel Management should begin an interagency process to identify skill and personnel gaps in the federal government's digital workforce.

---

[112] For comparison, since 2001, C:SFS has had 3,600 graduates, or about 189 graduates per year. *History/Overview*, CyberCorps: Scholarship for Service, (last accessed June 17, 2020), https://www.sfs.opm.gov/Overview-History.aspx.

TAB 4 – Improve Export Controls and Foreign Investment Screening

In its Interim Report, the National Security Commission on Artificial Intelligence (NSCAI or the Commission) noted that export controls and investment screening are key to protecting America's edge in defense and security-related technologies.[113] This remains the case—but these tools must be applied discriminately to be most effective and still allow collaborative work with researchers from around the world, as the Interim Report also highlights.[114] Here, the Commission presents a range of recommendations on these issues in order to improve U.S. technology protection. Our proposals are informed by three underlying realities:

- These are complex tools that create trade-offs between strategic impact, economic cost, geopolitical risk, and technical and political feasibility. Weighing these trade-offs is particularly difficult for Artificial Intelligence (AI), which is dual-use, widespread, and builds on a host of other technologies.

- Protection alone cannot sustain U.S. advantages and must remain focused on preventing the transfer of critical technologies that could create risk to U.S. national security. Technology protections must be integrated with a broader strategy for promoting U.S. innovation. This is also part of the Commission's ongoing work.

- Ongoing government efforts to strengthen protection tools have been slow and have created uncertainty, especially in the private sector. In 2018, the Congress enacted necessary legislative reforms to overhaul U.S. protection mechanisms through the Export Control Reform Act of 2018 (ECRA) and the Foreign Investment Risk Reduction Modernization Act of 2018 (FIRRMA). Yet nearly two years later, implementation of key aspects of both laws remains unfinished. This has left gaps in the U.S. approach to protecting AI.

The recommendations below are weighted toward Executive Branch action, primarily to assist with implementation of ECRA and FIRRMA and advise on

---

[113] *Interim Report*, NSCAI at 45 (Nov. 2019), https://www.nscai.gov/reports [hereinafter Interim Report].

[114] Another important element of preventing the illicit transfer of sensitive technologies which is not discussed in this memo is protecting talent. This can take the form of competitors attempting to lure individuals away from U.S. firms in order to gain access to sensitive intellectual property, or attempting to penetrate the U.S. academic ecosystem to obtain early-stage research under the fundamental research exemption to export controls. The Commission is examining this issue separately.

regulatory changes. They are separated into four categories. *First*, we outline broad principles guiding the Commission's approach to technology protection, which underpin our recommendations. *Next*, we present recommendations to enhance the United States Government's capacity to craft and implement technology protection policies. *Then*, we offer recommendations on applying export controls to AI. *Finally*, we propose measures to focus the Committee on Foreign Investment in the United States (CFIUS) on limiting foreign influence on sensitive technologies that are important for national security. The Commission also offers a draft executive order (E.O.) on applying export controls to AI (included in Appendix B), which would serve as an implementation vehicle for several of these recommendations, and proposes legislation to enhance CFIUS' ability to monitor investments in U.S. AI firms by Russia and China.

Part I: Principles for a Strategic Approach to Technology Protection

The Commission proposes four overarching principles to inform U.S. policy for protecting critical, dual-use technologies, including AI. We have found no similar framework within the government to guide such deliberation and action.[115]

Principle 1. Controls cannot supplant investment and innovation.

Export controls and investment screening will never eliminate the need for continued U.S. technical innovation. Technology protection policies are intended to slow U.S. competitors' pursuit and development of key strategic technologies, not stop them in their tracks. As the Commission has stated before, the United States must cultivate investment in these technologies through direct federal funding or changes to the regulatory environment in order to preserve existing U.S. advantages.[116] Toward that end, the Commission is encouraged by recent developments to revitalize domestic fabrication of state-of-the-art microelectronics, including Taiwan Semiconductor Manufacturing Corporation's (TSMC) decision to develop an advanced facility in the United States,[117] Intel's announcement of interest in working

[115] This memo outlines several actions pertaining to export controls which can be accomplished via Executive Order. These four principles, along with the recommendations pertaining to export controls in this memo which can be implemented via Executive Order are included in the draft Executive Order on Applying Export Controls to AI and Emerging Technologies, which is attached in Appendix B.

[116] Interim Report at 25; *First Quarter Recommendations*, NSCAI at 2-4 (Mar. 2020), https://www.nscai.gov/reports [hereinafter First Quarter Recommendations].

[117] See Michael Pompeo, *The United States Welcomes Taiwan Semiconductor Manufacturing Corporation's Intent to Invest $12 Billion to Bolster U.S. National Security and Economic Prosperity*, Department of State (May 14, 2020), https://www.state.gov/the-united-states-welcomes-taiwan-semiconductor-manufacturing-corporations-intent-to-invest-12-billion-to-bolster-u-s-national-security-and-economic-prosperity/.

with the United States government to develop a commercial U.S. foundry,[118] and the introduction of the "CHIPS for America Act" which would provide a substantial boost to U.S. semiconductor manufacturing.[119] Additionally, the United States should strongly consider when it is in its best interest to promote open-source development of AI rather than instituting controls on it. The United States leads in open-source AI software development, which is a key source of strength for developing technical standards, promoting platform adoption, and more.[120] Simply put, to ensure continued U.S. leadership in AI, the best defense is a good offense.

Principle 2. U.S. strategies to promote and protect must be integrated.

U.S. strategy to protect emerging technologies such as AI must be integrated with efforts to promote U.S. leadership in such technologies. Currently, most U.S. efforts to control technology flows are entirely divorced from efforts to promote growth in those same fields, resulting in inefficient outcomes. When choosing to implement controls the United States should simultaneously consider policies to spur domestic research and development (R&D) in key industries. This would help offset the resulting costs to U.S. firms, create alternative global markets, or encourage new investment to strengthen the U.S. industrial position. For instance, in its First Quarter Recommendations the Commission proposed several targeted steps the United States could take to boost funding and support for R&D in AI-related hardware, which should be implemented in conjunction with any AI-related hardware controls.[121] Doing so would magnify the impact of both actions, enhancing the compliance of U.S. firms with the controls while also offsetting their economic impact.

---

[118] *Letter from Intel Corporation CEO Bob Swan to Deputy Undersecretary of Defense Lisa Porter and Ms. Nicole Petta*, Wall Street Journal (Apr. 28, 2020), https://s.wsj.net/public/resources/documents/intel%20letter.pdf.

[119] NSCAI's Q1 Recommendations highlighted the need for the United States Government to pursue policies that encourage domestic facilities for advanced microelectronics packaging and testing to create an end-to-end domestic microelectronics industrial base. See First Quarter Recommendations. For background on the "CHIPS for America Act" see Senator Mark Warner & Senator John Cornyn, *Bipartisan, Bicameral Bill Will Help Bring Production of Semiconductors, Critical to National Security, Back to U.S.* United States Senate (June 10, 2020), https://www.warner.senate.gov/public/index.cfm/2020/6/bipartisan-bicameral-bill-will-help-bring-production-of-semiconductors-critical-to-national-security-back-to-u-s.

[120] Open source AI software development is also an area that the Chinese government has identified as a weakness within its AI ecosystem. See *Hearing On Technology, Trade, And Military-Civil Fusion: China's Pursuit Of Artificial Intelligence, New Materials, And New Energy*, United States-China Economic and Security Review Commission at 13 (June 7, 2019), https://www.uscc.gov/sites/default/files/2019-10/June%207,%202019%20Hearing%20Transcript.pdf.

[121] First Quarter Recommendations at 45.

Principle 3. Export controls must be targeted, strategic, and coordinated with allies.

In devising new export controls on technology that is as widespread and dual-use as AI, the United States must be careful and selective in the implementation of export controls. In order to ensure maximum effectiveness and minimize the adverse impact on U.S. industry, the Commission proposes that policymakers utilize the following three-part test in designing new export controls on emerging technologies, to include AI or any associated technologies:

1. Export controls must be targeted, clearly defined, discrete, and focused on choke points where they will have a strategic impact on the national security capabilities of competitors, but smaller repercussions on U.S. industry.[122]
2. Export controls must have a clear strategic objective, seeking to deter competitors from pursuing paths that endanger U.S. national security interests, and account for the projected cost and timeframe for competitors to create a domestic alternative.[123]
3. Export controls must be coordinated with key U.S. allies that are also capable of producing the given technology, in order to effectively restrict the supply to adversaries and also prevent circumstances where unilateral controls result in U.S. firms losing business to allied firms, without altering competitors' access.[124]

This test is particularly important when considering regulations on AI systems, which, as the Commission has previously noted, represent a constellation of interrelated technology blocks, including the hardware, algorithmic, and data subcomponents that feed each model.[125] Given the broad definition of AI and the inherently dual-use nature of the technology, any export controls on AI systems must be clear and precise, and focus on individual and specific subcomponents rather than AI systems writ large.

---

[122] The clarity of export controls, in particular, is critical for U.S. industry compliance. Firms are generally willing to shoulder a heavier regulatory burden if they have certainty in a regulation's intent; uncertainty creates extra costs and will result in lower compliance from industry, either due to ignorance or perceived legal gaps in regulations.

[123] For instance, in May 2019, the United States placed Huawei on the Entity List, which prevented Huawei phones from using the Android operating system (OS). This caused Huawei to expedite production of its own operating system, HarmonyOS, which it now views as the future of its phones. In January 2020, a Huawei official stated Huawei is committed to Harmony OS and will not return to the Android OS even if it is permitted to in the future. See *Huawei Exec Shocks By Saying It Will Forego Google Apps Even If The Us Trade Ban Is Lifted*, Pocket-lint (Jan. 30, 2020), https://www.pocket-lint.com/phones/news/huawei/150935-huawei-exec-says-it-will-forego-google-apps-even-if-the-us-trade-ban-is-lifted.

[124] Export controls will be most effective on items that are produced either only in the United States, or are limited to select, close U.S. allies. The more diffuse a given technology is, the larger the international coalition that will be necessary to ensure effectiveness.

[125] See Interim Report at 8.

Principle 4. Pursue a more discerning approach on export controls while broadening investment screening.

The Commission cautions against applying broad and sweeping export controls on AI and other dual-use emerging technologies due to the potential for significant blowback on U.S. industry, which would harm overall U.S. strategic competitiveness. By contrast investment screening—defined as the review of the national security aspects of foreign direct investment in the United States by CFIUS[126]— presents opportunities to take a more proactive regulatory approach while minimizing risk to U.S. firms. Screening provides the government with significantly more insight regarding transactions pertaining to specific sectors or countries. Screening also makes it easier to identify investments that seek to enable illicit technology transfer to competitors (e.g., through controlling stakes and access to source code). Expanding the number of transactions involving firms from competitor nations that require a CFIUS filing would increase costs to firms and the regulatory workload for the government. But creating more certainty in the investment screening process will offset some of those costs. If the United States can ensure that benign transactions continue to get approved expeditiously—including by applying a more risk-informed approach to CFIUS to decrease the burden for low-risk investors—enhancing investment screening can significantly blunt concerning transfers of technology. Under current law, however, only investments in export-controlled technologies require CFIUS filings, thus prohibiting this type of bifurcated approach.

Part II: Enhancing Capacity to Carry Out Effective Technology Protection Policies

Departments and agencies responsible for protecting U.S. technologies lack sufficient capacity to analyze the impact of their actions on emerging technologies such as AI. They lack both sufficient technical capacities to identify effective new policies and analytical capacity to enforce their policies efficiently, especially on dual-use goods. Filling these gaps in key elements of the Executive Branch—particularly in the Departments of State, the Treasury, and Commerce—will enhance the government's ability to craft targeted export controls that have the greatest strategic impact and the least harm on U.S. competitiveness.

Both the Departments of the Treasury and Commerce have delayed or scaled back actions aimed at preventing the transfer of sensitive technologies, because they do not have enough manpower, resources, and analytical capacity. Commerce officials have stated that lack of resources to manage an intense workload is one reason the

[126] See James Jackson, *The Committee on Foreign Investment in the United States (CFIUS)*, Congressional Research Service at 1 (Feb. 14, 2020), https://fas.org/sgp/crs/natsec/RL33388.pdf [hereinafter Jackson, The Committee on Foreign Investment in the United States (CFIUS)].

Department has been slow to implement critical aspects of ECRA.[127] Additionally, during the drafting of the FIRRMA legislation, the Department of the Treasury pushed back on proposals to require CFIUS filings for all relevant transactions involving Chinese investors. Even under the current program, CFIUS anticipates its workload expanding dramatically to over 1,000 cases per year, which requires increasing CFIUS staff by approximately 50 percent.[128] Similarly, the Department of Commerce's Bureau of Industry and Security (BIS) requested a budget increase of eight percent over last year's request for a total of $138 million and 448 positions.[129]

A dearth of technical talent inside the relevant departments and agencies exacerbates their already difficult task. As a dual-use technology, AI poses a particular challenge compared to military technologies such as missile systems or weapons of mass destruction, which have little civilian commercial value. When export controls primarily targeted items with only military applications, regulators could draw on individuals with military experience to fill technical needs. Given the current national security linkages of dual-use technologies, and the concentration of expertise for most dual-use emerging technologies in the private sector, this is no longer the case. BIS has a very limited bench of resident technical experts on emerging dual-use technologies and few other experts within government to consult.[130] While it is not realistic to expect agencies such as BIS to have a deep technical expert in every technology field, agencies need more people who can communicate effectively, and at a technical level, with industry and with the interagency in crafting new controls.

Agencies need to draw on people with academic or industry expertise in technologies such as AI, quantum computing, biotechnology, and advanced telecommunications to evaluate the impact of potential controls on these technologies. They must rely more heavily on advisory committees and input from external sources to help make policy. There are some existing mechanisms to serve this purpose: in June 2018, BIS renewed the charter of the existing Emerging Technology Research Advisory Committee (ETRAC), and renamed it the Emerging Technology Technical Advisory

---

[127] Ana Swanson, *Trump Officials Battle Over Plan to Keep Technology Out of Chinese Hands*, New York Times (Oct. 23, 2019), https://www.nytimes.com/2019/10/23/business/trump-technology-china-trade.html [hereinafter Swanson, Trump Officials Battle Over Plan to Keep Technology Out of Chinese Hands].

[128] *Committee on Foreign Investment in the United States* (last accessed June 18, 2020), https://home.treasury.gov/system/files/266/10.-CFIUS-FY-2021-BIB.pdf.

[129] *Fiscal Year 2021 Congressional Budget Submission*, Department of Commerce, Bureau of Industry and Security at 5 (last accessed June 18, 2020), https://www.commerce.gov/sites/default/files/2020-02/fy2021_bis_congressional_budget_justification.pdf [hereinafter BIS FY 21 Congressional Budget Submission].

[130] Recognizing its need for increased capacity, BIS' FY21 budget request includes funding for five new positions specifically to assist with "identifying and reviewing emerging and foundational technologies (as directed in ECRA Sec. 1758)." See BIS FY 21 Congressional Budget Submission at 5. It also requested eight new positions for "initiatives to address China and emerging technology." Id. at 9.

Committee (ETTAC).[131] The ETTAC contains roughly 20 leading technical experts from prominent U.S. technology and defense firms, universities, and think-tanks. However, following its redesignation, the ETTAC took nearly two years to hold a meeting, holding its first session on May 19, 2020.[132] Commerce must make greater use of outside experts as it formulates export control policies on emerging technologies.

To increase the capacity of the Departments of Commerce, the Treasury, and State to implement policies for protecting sensitive U.S. technologies, the Commission offers two recommendations.

Recommendation 1: Mandate that the Department of Commerce coordinate new rules with existing technical advisory groups that include outside experts.

The White House should issue an executive order[133] mandating that the Department of Commerce solicit and receive feedback on any proposed controls on emerging or foundational technologies, to include proposed rules and regulations, from the ETTAC and any other relevant technical advisory groups or technical special advisors, before putting them into effect or sharing them with the public. Such advisory groups and advisors—which should include deep subject matter experts from outside government serving on a temporary basis—can provide a wealth of expertise at minimal cost to the government. They can address whichever technologies are being considered for controls and develop important connections to industry and academia. While ECRA specifically states that Commerce should utilize information from the ETTAC in forming new rules,[134] there is no formal mechanism or statutory requirement for it to do so. To ensure that key regulatory agencies benefit from the committee's insight, Treasury and State should be granted non-voting observer seats in all ETTAC meetings.

Mandating that agencies consult with and receive feedback from technical advisory groups and consider seeking the input of technical special advisors would force agencies to better utilize these entities in the regulation drafting process. Although the ETTAC is permitted to advise Commerce on the potential impact of export control revisions, it is currently only obligated to do so via semi-annual reports to the

---

[131] *Emerging Technology Technical Advisory Committee Charter*, Department of Commerce (June 25, 2018), https://tac.bis.doc.gov/index.php/documents/pdfs/370-ettac-bis-charter/file [hereinafter ETTAC Charter].
[132] *Emerging Technology Technical Advisory Committee; Notice of Partially Closed Meeting*, 85 Fed. Reg. 13131 (Mar. 6, 2020), https://www.federalregister.gov/documents/2020/03/06/2020-04605/emerging-technology-technical-advisory-committee-notice-of-partially-closed-meeting.
[133] This memo outlines several actions pertaining to export controls which can be accomplished via Executive Order. These recommendations, along with the four principles outlined at the beginning of the memo, are included in the draft Executive Order on Applying Export Controls to AI and Emerging Technologies, which is attached in Appendix B.
[134] 50 U.S.C. § 4817(a)(2)(A)(iv).

Assistant Secretary of Commerce for Export Administration.[135] More frequent and effective use of such existing advisory committees would provide flexible technical expertise to key departments, and help prevent the implementation of controls that are counterproductive. It would also give industry a clearer view of Commerce's plans for export controls.

Recommendation 2: Designate a network of FFRDCs and UARCs to serve as a shared technical resource on export controls and help automate review processes.

The Department of Commerce should establish a network within existing federally funded research and development centers (FFRDCs) and university affiliated research centers (UARCs) to provide technical expertise to all departments and agencies for issues relating to export controls on emerging technologies.[136] This network would be coordinated by the Department of Commerce and ideally would encompass a regional distribution of FFRDCs and UARCs that are located in U.S. technology hubs or that have significant expertise in emerging technologies. It would provide deeper technical expertise than in-house experts are able to provide.[137] It could provide more tactical advice than the technical advisory committees and could give real-time technical input to policy discussions on export controls. This would inject a rigorous external voice into the policy process, without presenting the conflict of interest concerns raised by direct consultations with industry. Ultimately, the network would bring together experts from across the country with complementary technical backgrounds to offer Commerce and other agencies a range of informed perspectives regarding technology protection policies for emerging technologies on a case-by-case basis. As an initial step, the Department of Commerce should identify the FFRDCs and UARCs with existing expertise in emerging technologies under consideration for export controls. This could be followed by a request for funding in the Fiscal Year (FY) 2022 President's Budget to support and expand work of FFRDCs and UARCs in this area.

Additionally, the Departments of the Treasury and Commerce should work with FFRDCs, UARCs, and other contracted entities to construct AI-based systems that would enhance the United States Government's export control and investment screening processes. AI-based systems could reduce costs, increase efficiency, and free up time for staff to focus on strategic-level analysis of technology protection issues. For example, CFIUS cases more than doubled from 2010 to 2018, even before

---

[135] See ETTAC Charter at 2.

[136] The proposed Executive Order on Applying Export Controls to AI and Emerging Technologies, which is attached in Appendix B, contains implementing language for this recommendation.

[137] The Department of Commerce already sponsors the National Cybersecurity FFRDC, which is operated by the MITRE Corporation and is focused on providing technical advice to the government on issues pertaining to cyber security. This network would have a similar goal, although by leveraging a network of existing FFRDCs which already contain significant technical expertise, it obviates the need to create a new entity.

FIRRMA was implemented.[138] As the number of cases continues to increase, staff will need more powerful tools to process cases in a timely manner. Such a system could conduct a preliminary analysis of export license requests and CFIUS filings and attempt to determine their level of risk—scoring each new application based on the perceived risk of the technologies, countries, and individual actors involved. As an initial step, this could serve to bucket transactions into low, medium, and high-risk tranches, before a human conducts a more detailed review. As the system matures, it could conduct more granular levels of analysis, and potentially automatically approve or reject very low- or high-risk transactions without human involvement. Although such a system would be complicated, data-intensive, and likely err on the side of caution, in the long-run it could provide significant benefits. This system could be more accurate than human review and significantly less labor-intensive, allowing the government to more rapidly process benign requests and reject a greater share of malicious ones. Additionally, this would help integrate export control and investment screening data and strategies into a single risk framework, which would allow the government to conduct more precise risk analysis.

Part III: Applying Export Controls to AI

A. Prioritizing Feasible and Effective Export Controls Related to AI

Coupled with a more comprehensive approach to promoting innovation and technology leadership, export controls are central to protecting U.S. national security interests.[139] As described in NSCAI's Interim Report, when evaluating the effectiveness of export controls for AI, one must separately consider the effectiveness of controls for each element of the AI stack, specifically hardware, algorithms, and data.[140] As the Department of Commerce continues to apply ECRA, it should identify and prioritize elements of the AI stack where controls can have the greatest strategic impact.

---

[138] *Annual Report to Congress*, Committee on Foreign Investment in the United States (2018), https://home.treasury.gov/system/files/206/CFIUS-Public-Annual-Report-CY-2018.pdf.
[139] See also James Lewis, *Managing Semiconductor Exports to China*, Center for Strategic and International Studies (May 5, 2020), https://www.csis.org/analysis/managing-semiconductor-exports-china [hereinafter Lewis, Managing Semiconductor Exports to China].
[140] Talent is also part of the AI stack but is outside the scope of this memo. Immigration is an important issue that the Commission is examining separately.

Recommendation 3: Prioritize hardware controls to protect U.S. national security advantages in AI, and consider future controls surrounding data.

Overly broad export controls on general-purpose AI software run the risk of causing substantial harm to the U.S. AI innovation base, and ultimately are not practical to implement. If regulators heavily control the export of all AI software, it would likely compel U.S. firms to push all AI-related research and development overseas.[141] Additionally, given that many AI tools are widely available through open source, controlling the export of all items that utilize AI, or are critical to developing AI, is neither feasible nor economically viable.

Hardware—and to a lesser degree, data—present potential choke points where controls can be targeted, discrete, and effective in protecting U.S. national security interests. To support the Department of Commerce's efforts, the Commission offers the following assessment of which parts of the AI stack lend themselves to the most useful export controls. Controls on hardware—and specifically on semiconductor manufacturing equipment, rather than on individual chips—are most likely to have positive strategic effects, followed by potential future controls on key datasets. We offer a more detailed discussion below:

- **Hardware.** — Export controls on advanced hardware capabilities, particularly advanced semiconductor manufacturing equipment, are more likely to advance U.S. national security interests than controls on any other element of the AI stack. AI is compute-intensive, and some of the equipment necessary to manufacture advanced chips is extremely complicated and only manufactured by a select number of firms. This creates an opportunity to control the equipment that produces chips, which power high-end AI applications. China's concerted effort to grow its domestic semiconductor industry, which relies heavily on imports of advanced equipment necessary to manufacture high-end chips, threatens to upend U.S. and allied leadership in this field.[142] (Specific recommendations on controls on key types of semiconductor manufacturing equipment are detailed in Recommendation 5 of this memo.) Additionally, tighter controls on AI-specific chips, such as particular types of ASICs, GPUs, or FPGAs,[143] could be considered in the future, if the controls are sufficiently tailored, specific, and necessary beyond what is already controlled in existing regimes. However, controls on general-purpose semiconductors are unlikely to prove effective unless coordinated with all countries capable of producing such chips. If

[141] Cade Metz, *Curbs on A.I. Exports? Silicon Valley Fears Losing Its Edge*, New York Times (Jan. 1, 2019), https://www.nytimes.com/2019/01/01/technology/artificial-intelligence-export-restrictions.html.
[142] See Lewis, Managing Semiconductor Exports to China.
[143] ASICs are application-specific integrated circuits, GPUs are graphics processing units, and FPGAs are field programmable gate arrays.

implemented unilaterally, such controls could harm the U.S. semiconductor industry.

- **Data.** — Data represents an area for future potential AI-related export controls, although hardware-related controls should remain the priority. BIS and the Department of State should consider whether key datasets, AI enabling data (i.e. weights), and AI enriched data represent future potential opportunities for export controls, particularly as conversations about international data cooperation and standards continue to evolve.[144] Many sensitive datasets are already controlled through existing regimes, such as technical data controlled by the International Traffic in Arms Regulations (ITAR).[145] Outside of the ITAR regime, future definitions for controls on data could better account for personally identifiable information, genetic information, or other sensitive information about U.S. persons. Some of this information can be used to train AI algorithms, and its transfer outside of the country in bulk creates national security risks. Such transfers would require technical measures to securely anonymize and encrypt some data before export. It would also require additional guidelines for accessing and transferring sensitive data across international borders.[146] BIS and State should consider if there are bulk datasets that are not currently controlled but should be. There is also room to work with allies and partners to create standards for securely transferring key datasets, which would limit their distribution only to certain nations.[147]

---

[144] CFIUS is also playing an active role in restricting foreign access to sensitive data, as demonstrated by the Committee's decision to require divestment of the app Grindr by a Chinese firm last year over concerns regarding the app's collection of personal data. See Echo Wang, *China's Kunlun says U.S. Approves Sale of Grindr to Investor Group*, Reuters (May 29, 2020), https://www.reuters.com/article/us-grindr-m-a-sanvincente/chinas-kunlun-says-u-s-approves-sale-of-grindr-to-investor-group-idUSKBN2352PR.

[145] For a definition of "technical data," see 22 CFR 120.10. For example, ITAR restrictions on technical data controlled by USML Category XIII(i)(10) could apply to models used in machine learning: "Technical data for modifying visual, electro-optical, radiofrequency, electric, magnetic, electromagnetic, or wake signatures (e.g., low probability of intercept (LPI) techniques, methods or applications) of defense platforms or equipment through shaping, active, or passive techniques."

[146] Efforts in this area would build on the December 2019 U.S. Directorate of Defense Trade Controls interim final rule detailing encryption standards for ITAR data, including cloud transfer and storage of ITAR technical data. *International Traffic in Arms Regulations: Creation of Definition of Activities That Are Not Exports, Reexports, Retransfers, or Temporary Imports; Creation of Definition of Access Information; Revisions to Definitions of Export, Reexport, Retransfer, Temporary Import, and Release*, 84 Fed. Reg. 70887, 70892 (Dec. 26, 2019), https://www.govinfo.gov/content/pkg/FR-2019-12-26/pdf/2019-27438.pdf#page=6.

[147] For example, Japan has proposed a "data free flow with trust" approach. See Nigel Cory et al., *Principles and Policies for "Data Free Flow With Trust,"* Information Technology & Innovation Foundation (May 27, 2019), https://itif.org/publications/2019/05/27/principles-and-policies-data-free-flow-trust.

- **Algorithms.** — AI algorithms would be extremely difficult to control. Such algorithms often are dual-use and tend to originate in the commercial sector or academia. Many are available as open-source software.[148] Also, algorithms are iterative in nature and are constantly changing, which presents a definitional challenge for the export control regime. Some AI algorithms, including those meant for use in battlefield applications, are clear candidates for export controls, but such software is already controlled under the Commerce Control List.[149] While some specific applications may seem ripe for control—such as those used for censorship, disinformation, or deepfakes—the dual-use nature of these applications makes controls very hard to enforce. As a result, if BIS and the Department of State implement export controls on application-specific AI algorithms that are not otherwise controlled, they will need to shift away from the traditional item-based approach and focus more on the end uses and end users of such items.

- **End-Use and End-User Controls.** — End-use and end-user controls can be effective tools at preventing the involvement of U.S. firms in problematic uses of AI, but in isolation they will not be effective at preventing the transfer of key, strategic technologies to U.S. competitors. As the Commission's Interim Report highlights, end-use and end-user controls may prove more effective than list-based controls at preventing the transfer of specific U.S. AI technology to known human rights violators and other malicious actors.[150] For instance, prohibiting the export of facial recognition surveillance equipment to Chinese companies involved in mass surveillance of Uyghur populations in Xinjiang could prevent U.S. firms from wittingly or unwittingly facilitating human rights abuses.[151] Coupled with demonstrating U.S. commitment to ethical uses of AI, this approach would highlight Chinese disregard for ethical principles. Any end-use or end-user controls would have to be extremely specific and clear, in order to maximize compliance from U.S. industry and reduce unnecessary costs to firms associated with ambiguity. However, for high-end, critical components of the AI-stack which are key to technological breakthroughs and national security advantage, tailored end-use and end-user restrictions are unlikely to prevent the eventual transfer of that technology to restricted actors or governments, regardless of the compliance of U.S. firms.

---

[148] For example, Google released TensorFlow as an open source machine learning platform in 2015. *TensorFlow,* (last accessed June 18, 2020), https://www.tensorflow.org/.

[149] Carrick Flynn, *Recommendations on Export Controls for Artificial Intelligence*, Center for Security and Emerging Technology at 6 (Feb. 2020), https://cset.georgetown.edu/wp-content/uploads/Recommendations-on-Export-Controls-for-Artificial-Intelligence.pdf.

[150] Interim Report at 42.

[151] The United States has already imposed sanctions on Chinese surveillance and AI firms, such as Hikvision and Sensetime, for their roles in human rights abuses inside China. See Shawn Donnan & Jenny Leonard, *U.S. Blacklists Eight Chinese Tech Companies on Rights Violations*, Bloomberg (Oct. 7, 2019), https://www.bloomberg.com/news/articles/2019-10-07/u-s-blacklists-eight-chinese-companies-including-hikvision-k1gvpq77.

As AI adoption in national security applications expands, government and industry will have to adapt by working together to assess regularly whether existing controls are sufficient to aid in preserving U.S. technology advantages. At the same time, controls should not unduly hinder U.S. AI company competitiveness. Future controls should also be informed by case studies on the success and failure of prior efforts. This collaboration should help to inform and adjust the prioritization of controls over time. The notice-and-comment process will also be an important way for government and industry to develop standard definitions for compliance.[152]

B. Expediting Issuance of Key ECRA and FIRRMA Regulations

ECRA and FIRRMA sought to modernize the U.S. export control and investment screening regimes, respectively. The primary purpose of both laws is to address weaknesses in the existing regimes regarding the transfer of critical technologies to destinations of concern, particularly China.[153]

Specifically, the laws are intended to develop and integrate U.S. policies controlling "emerging and foundational technologies."[154] Under ECRA, BIS is responsible for developing a regular, formal interagency process to identify "emerging and foundational technologies that are essential to the national security of the United States," and are not otherwise controlled.[155] Additionally, FIRRMA stipulates that "emerging and foundational technologies" identified by BIS are treated as "critical technologies" under CFIUS, in addition to becoming subject to export licensing requirements. As a result, transactions involving foreign investors and a U.S. company that "designs, tests, develops, or produces" such technologies, regardless of whether the investment is for a controlling stake or not, must be reviewed under the CFIUS process.[156] This mechanism provides an important link between the export

---

[152] See e.g., Robert Atkinson & Stephen Ezell, *Information Technology & Innovation Foundation Comments on ANPRM on the Review of Controls for Certain Emerging Technologies* (December 13, 2018), http://www2.itif.org/2018-export-control-filing.pdf.

[153] *The Export Control Reform Act and Possible New Controls on Emerging and Foundational Technologies*, Akin Gump Strauss Hauer & Feld LLP (Sept. 12, 2018), https://www.akingump.com/en/news-insights/the-export-control-reform-act-of-2018-and-possible-new-controls.html.

[154] Although neither ECRA nor the Department of Commerce has a formal definition for what constitutes an "emerging" or "foundational" technology, "emerging" technologies are generally considered to be those which may pose over-the-horizon national security threats in the coming years, while "foundational" technologies are critical underlying technologies which can enable progress and advancement in a wide variety of domains.

[155] 50 U.S.C. § 4817(a)(1)(A).

[156] Harry Clark et al. *Some Foreign Investment Transactions Involving "Critical Technology" Soon Must be Notified to CFIUS*, Orrick Herrington & Sutcliffe LLP (Oct. 15, 2018), https://reaction.orrick.com/rs/vm.ashx?ct=24F76C1DD1E446A9CCDD89ACD42A911BD8F055B2DF8E0BD15EE5636069FFCB1CDB7A3A9C3.

control and investment screening regimes. Figure 2, below, illustrates the relationship between ECRA and FIRRMA:

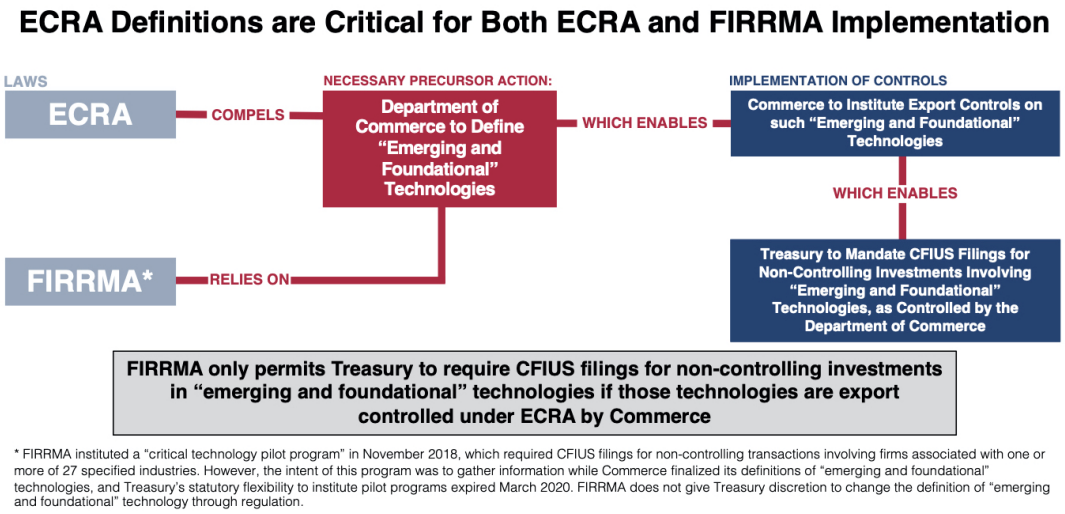

Figure 2: Illustrating the link between ECRA and FIRRMA

Commerce has yet to identify a single emerging or foundational technology as mandated by ECRA. This delay has slowed the implementation of both ECRA and FIRRMA. In November 2018, Commerce issued an Advanced Notice of Proposed Rulemaking (ANPRM) seeking industry comment on fourteen categories of technologies that could be considered "emerging," which included AI.[157] After receiving over 250 comments from prominent industry groups and stakeholders,[158] as of June 2020, Commerce has yet to release a final version of this list or identify any technology or classes of technology as "emerging" or "foundational."[159] While there

[157] *Review of Controls for Certain Emerging Technologies; A Proposed Rule by the Industry and Security Bureau*, 83 Fed. Reg. 58201 (Nov. 19, 2018), https://www.federalregister.gov/documents/2018/11/19/2018-25221/review-of-controls-for-certain-emerging-technologies.

[158] Jeffrey Cunard et al. *TMT Insights: What is on the Horizon for Export Controls on "Emerging Technologies"? Industry Comments May Hold A Clue*, Debevoise & Plimpton LLP (Sept. 23, 2019), https://www.debevoise.com/insights/publications/2019/09/tmt-insights-what-is-on-the-horizon-for-export.

[159] Commerce has implemented select, specific controls on particular individual technologies, such as its January 2020 addition of controls on AI software to facilitate geospatial imagery analysis. However, it implemented these regulations through a different process - the ECCN 0Y521 series procedures, designed for immediate action on tailored technologies - which ECRA sought to standardize and integrate into its process of creating and regularly updating lists of emerging and foundational technologies. See *Addition of Software Specially Designed To Automate the Analysis of Geospatial Imagery to the Export Control Classification Number 0Y521 Series; A Rule by the Industry and Security Bureau*, 85 Fed. Reg. 459 (Jan. 6, 2020), https://www.federalregister.gov/documents/2020/01/06/2019-27649/addition-of-software-specially-designed-to-automate-the-analysis-of-geospatial-imagery-to-the-export.

is reason to be judicious in developing this list, it is now time for action to meet the requirements of the law.

This delay has also limited CFIUS' insight into foreign investment in critical technologies, which FIRRMA intended to expand. FIRRMA defines "critical technologies" as items already controlled on the United States Munitions List or Commerce Control List, certain key nuclear or chemical equipment, and emerging and foundational technologies controlled under ECRA.[160] The Department of the Treasury cannot expand the scope of what constitutes a "critical technology" beyond what is listed in FIRRMA, and FIRRMA does not give Treasury the discretion to change the definition of "emerging and foundational" technologies through regulation. The Department of the Treasury smartly codified the "critical technology pilot program" into FIRRMA regulations that took effect in February 2020, in order to require CFIUS filings for certain key industries—such as space vehicles and semiconductors—independent of ECRA implementation. FIRRMA granted Treasury statutory flexibility to institute pilot programs but that authority expired in March 2020 and it was not intended to be an all-encompassing or a permanent solution for emerging technologies.[161]

Although Department of Commerce officials have stated multiple times that Commerce is close to releasing initial definitions, they have yet to emerge.[162] In November 2019, Senators Schumer and Cotton sent a joint letter to Secretary Ross noting Commerce's slow implementation of ECRA, asking for an explanation for the delay, and urging that Commerce conclude its review as quickly as possible.[163] The delay has caused significant uncertainty for firms working in fields that could be labeled as emerging or foundational technologies, while also delaying the government's ability to either control the export of, or more importantly gain insight into, transactions involving critical technologies that are not otherwise controlled.[164]

To expedite the issuance of key ECRA and FIRRMA regulations, diminish industry uncertainty, and increase the government's ability to regulate transactions involving critical technologies, it is important that Commerce release initial lists of technologies it considers to be "emerging" and "foundational" as soon as possible. Implementation of ECRA and FIRRMA rest on the completion of at least initial versions of these lists, and two years after both laws went into effect their implementation continues to languish.

---

[160] 50 U.S.C. § 4565(a)(6)(A).
[161] *Provisions Pertaining to Certain Investments in the United States by Foreign Persons*, 85 Fed. Reg. 3112 (Jan. 17, 2020), https://home.treasury.gov/system/files/206/Part-800-Final-Rule-Jan-17-2020.pdf [hereinafter 85 Fed. Reg. 3112].
[162] See Swanson, Trump Officials Battle Over Plan to Keep Technology Out of Chinese Hands.
[163] *Letter from Sen. Tom Cotton & Sen. Chuck Schumer to Wilbur Ross, Secretary of Commerce*, (Nov. 18, 2019), https://www.cotton.senate.gov/files/documents/191118_Cotton_Schumer_ECRA%20Letter%20to%20Sec.%20Ross%20copy.pdf.
[164] See 85 Fed. Reg. 3112.

Recommendation 4: Issue an executive order directing the Department of Commerce to finalize identification of emerging and foundational technologies under ECRA.

The White House should issue an E.O. laying out clear timelines for the Department of Commerce to develop its initial lists of "emerging" and "foundational" technologies.[165] Finalizing the initial version of these lists, if properly scoped and well-defined, would ensure critical technologies are controlled, provide clarity to industry regarding how Commerce intends to implement ECRA, and also ensure that such technologies are included within CFIUS. Developing these lists via a rigorous interagency process, rather than on an ad-hoc basis, should result in increased internal coordination, more refined export control policy proposals, and ensure that export controls are exclusively utilized to protect national security rather than as a tool of protectionism.

The E.O. would require that the proposed rules listing both "emerging" and "foundational" technologies be issued within 120 days of its implementation. Additionally, the order would explicitly make clear that the agreed upon and formal National Security Council decision making process for adjudicating and elevating disputes is responsible for resolving policy disagreements between agencies on specific key terms, and if necessary, escalating disputes to principals to ensure that all agencies' voices are fully reflected in the process. Finally, this approach would recognize that these lists, particularly the list of "emerging" technologies, are iterative in nature, and note that ECRA requires Commerce to continue to refine the list, and engage with industry, as technologies develop and mature.

C. Preventing the Flow of High-End Semiconductor Manufacturing Equipment to Competitors

The primary U.S. export control target to constrain competitors' AI capabilities should be the semiconductor manufacturing equipment (SME) necessary to manufacture high-end chips. Slowing the growth of China's high-end semiconductor manufacturing ability, coupled with continued U.S. investments in microelectronics R&D, will set back China's attempts to catch up to the United States and its allies, and force it to continue to rely on foreign firms to supply its high-end semiconductors. While constraining China's chip manufacturing capability does not inherently restrict China's ability to acquire high-end chips, it would force China to rely more on U.S. and allied firms for such production, which would provide the United States with significant leverage over China's future capabilities.

[165] The proposed Executive Order on Applying Export Controls to AI and Emerging Technologies, which is attached in Appendix B, contains implementing language for this recommendation.

There are four primary reasons SME makes an ideal target for export controls to limit China's future AI capabilities:

1. Compute is key to AI - AI is becoming increasingly reliant on compute over time,[166] even as its application becomes more widespread. These two forces demonstrate that high-end semiconductors will be essential to power many future AI applications.

2. China will likely remain reliant on high-end semiconductor imports - In 2016, semiconductors were China's largest import, totaling over $200 billion,[167] and it does not have significant domestic production capability for chips below 14nm.[168] However, it has invested heavily in the semiconductor field to grow its domestic supply chain and become an industry leader by 2030, with the ultimate goal of decreasing or eliminating its reliance on foreign hardware.[169] From 2014 to 2018, China was the world's largest importer of SME, accounting for 29 percent of global imports.[170]

3. High-end SME is very specialized - In particular, extreme ultraviolet (EUV) lithography tools, the most advanced photolithography technology, are necessary for producing chips at the 7nm node and below and cost $120 million, weigh 180 tons, and require 20 trucks or three fully loaded Boeing 747s to ship.[171] The complex nature, rarity, and size of this equipment makes it difficult to replicate or steal.

---

[166] OpenAI estimates that since 2012, the amount of compute used in the largest AI training runs is doubling every 3.4 months. See *AI and Compute*, OpenAI (May 16, 2018), https://openai.com/blog/ai-and-compute/.

[167] In 2016, semiconductors were China's largest import, totaling over $200 billion. See Cheng Ting-Fang, *China's Upstart Chip Companies Aim To Topple Samsung, Intel And TSMC*, Nikkei Asian Review (Apr. 25, 2018), https://asia.nikkei.com/Spotlight/The-Big-Story/China-s-upstart-chip-companies-aim-to-topple-Samsung-Intel-and-TSMC.

[168] Semiconductor Manufacturing International Corporation (SMIC), China's leading foundry, currently has limited production capability at the 14nm node. For anything more advanced, China is reliant on firms located in the United States, Taiwan, or South Korea. See Josh Horwitz, *Huawei Chip Unit Orders Up More Domestic Production As U.S. Restrictions Loom: Sources*, Reuters (Apr. 16, 2020), https://www.reuters.com/article/us-huawei-tech-tsmc/huawei-chip-unit-orders-up-more-domestic-production-as-u-s-restrictions-loom-sources-idUSKCN21Y1G5.

[169] *Made in China 2025: Global Ambitions Built on Local Protections*, U.S. Chamber of Commerce at 65 (Mar. 5, 2017), https://www.uschamber.com/sites/default/files/final_made_in_china_2025_report_full.pdf.

[170] John VerWey, *The Health and Competitiveness of the U.S. Semiconductor Manufacturing Equipment Industry*, U.S. International Trade Commission at 8 (Jul. 2019), https://www.usitc.gov/publications/332/working_papers/id_058_the_health_and_competitiveness_of_the_sme_industry_final_070219checked.pdf.

[171] Andreas Thoss, *EUV Lithography Revisited*, Laser Focus World (Aug. 29, 2019), https://www.laserfocusworld.com/blogs/article/14039015/how-does-the-laser-technology-in-euv-lithography-work.

4. U.S. Allies control the SME market - The manufacturers of SME are concentrated within a very small geographic group of allied nations. In 2017, the eight largest SME firms were located in the United States, Japan, and the Netherlands.[172] These three countries also contained over 90 percent of the global SME industry in 2015.[173]

Photolithography tools, the most complex and expensive type of SME, are even more concentrated than SME writ large, with one active Dutch company (ASML) and two active Japanese companies (Nikon and Canon).[174] Furthermore, ASML has a monopoly over EUV lithography tools. ASML also has a dominant 88 percent market share in ArF immersion photolithography tools, the next most advanced photolithography technology, necessary for chips from the 7nm to 45nm node. Nikon is the only other supplier of ArF immersion photolithography tools.[175]

Recommendation 5: The United States should work with the Netherlands and Japan to restrict the export of EUV and ArF immersion lithography equipment to China, and take steps to increase demand for such tools among U.S. firms.

The United States must work in cooperation with the Netherlands and Japan to prohibit the export of EUV and ArF Immersion lithography equipment to China, in order to restrict China's semiconductor production capability at the 45nm node and below, which the Commission assesses to be the chips most useful for advanced AI applications.[176] Although Chinese firms do have existing production capability down to 14nm at limited scale, China's ability to manufacture photolithography equipment capable of production below the 90nm node is significantly more limited. If these controls are effective, it would be very difficult for China to obtain any new high-end lithography equipment, and any repairs or maintenance on existing equipment would likely prove difficult. While chips 45nm and below currently present the most utility for advanced AI applications and are the most feasible to control, this standard

---

[172] David Manners, *Top Ten Foundries 2017*, Electronics Weekly (Dec. 1, 2017), https://www.electronicsweekly.com/news/business/top-ten-foundries-2017-2017-12/.
[173] Dorothea Blouin, *2016 Top Markets Report: Semiconductors and Related Equipment,* Department of Commerce, International Trade Administration at 5 (July 2016), https://legacy.trade.gov/topmarkets/pdf/Semiconductors_Top_Markets_Report.pdf.
[174] Peter Clarke, *ASML Increases Dominance of Lithography Market*. EE News Analog (Feb. 12, 2018) https://www.eenewsanalog.com/news/asml-increases-dominance-lithography-market.
[175] Robert Castellano, *ASML's Dominance of the Semiconductor Lithography Sector has Far-Reaching Implications*, Seeking Alpha (Jan. 23, 2018), https://seekingalpha.com/article/4139540-asmls-dominance-of-semiconductor-lithography-sector-far-reaching-implications.
[176] The Wassenaar Arrangement lists lithography equipment capable of making chips with features of 45nm or below as a controlled item. However, because the Wassenaar Arrangement is not binding, states parties are not obligated to comply with this as a legal restriction. See *List of Dual-Use Goods and Technologies and Munitions List,* Wassenaar Arrangement Secretariat (Dec. 2018), https://www.wassenaar.org/app/uploads/2019/consolidated/WA-DOC-18-PUB-001-Public-Docs-Vol-II-2018-List-of-DU-Goods-and-Technologies-and-Munitions-List-Dec-18-1.pdf.

will also have to be continuously reevaluated to ensure controls are capturing the proper equipment and not unnecessarily harming industry.

Given Dutch and Japanese companies are the sole suppliers of EUV and ArF Immersion lithography equipment, these two governments have the collective ability to significantly reduce China's ability to produce high-end semiconductors. In 2019, the United States reportedly put significant pressure on the Netherlands to block a sale of EUV lithography equipment from ASML to SMIC. These efforts proved successful, as ASML ultimately let the contract expire without delivering the equipment.[177] The United States should double down on such efforts, while also encouraging Japan to restrict China's access to ArF Immersion equipment.[178]

Furthermore, the United States should set a clear policy goal of remaining two generations ahead of China in state-of-the-art microelectronics fabrication capabilities by utilizing a combination of export controls and substantial commercial R&D investment.[179] In support of this goal, the United States should initiate a simultaneous effort to provide tax credits or subsidies to U.S. firms that purchase semiconductor manufacturing equipment, to include EUV and ArF Immersion lithography equipment from Dutch or Japanese firms, to support efforts to build advanced foundries in the United States. This program, which would require Congressional authorization, could partially assuage concerns from the governments of the Netherlands and Japan about the financial impact of controls on SME, while simultaneously working to revitalize the semiconductor manufacturing base in the United States. This credit could also be coupled with additional initiatives, such as a federal match program for existing state and local incentives, and tax credits for efforts to study and reduce the potential environmental impact of semiconductor facilities. Combined, these incentives would further efforts to grow diverse U.S. high-end microelectronics fabrication capabilities. This would complement the recent, encouraging announcements by TSMC that it intends to build a state-of-the-art

---

[177] Alexandra Alper et al., *Trump Administration Pressed Dutch Hard to Cancel Chip-Equipment Sale: Sources*, Reuters (Jan. 6, 2020), https://www.reuters.com/article/us-asml-holding-usa-china-insight/trump-administration-pressed-dutch-hard-to-cancel-china-chip-equipment-sale-sources-idUSKBN1Z50HN.

[178] The United and Japan have a dominant market share in many other SME chokepoints, meaning that there are additional export control opportunities only requiring collaboration between a small number of actors; most frequently the U.S., Japan and a third country. See Saif Khan & Carrick Flynn, *Maintaining China's Dependence on Democracies for Advanced Computer Chips*, Center for Security and Emerging Technology (Apr. 2020), https://cset.georgetown.edu/research/maintaining-chinas-dependence-on-democracies-for-advanced-computer-chips.

[179] This has been an informal U.S. policy goal in the past, but recent advancements in the Chinese semiconductor industry, combined with the offshoring of semiconductor production capabilities which used to reside in the United States, necessitate a more systematized approach to this challenge, See John VerWey, *Chinese Semiconductor Industrial Policy: Prospects for Future Success*, United States International Trade Commission Journal of International Commerce and Economics at 10 (Aug. 2019), https://www.usitc.gov/publications/332/journals/chinese_semiconductor_industrial_policy_prospects_for_success_jice_aug_2019.pdf.

fabrication facility in the United States,[180] and would also benefit other firms exploring similar proposals.[181]

The "CHIPS for America Act," a bipartisan, bicameral bill introduced by Senators John Cornyn (R-TX) and Mark Warner (D-VA), as well as by Representatives Michael McCaul (R-TX) and Doris Matsui (D-CA), includes such a tax credit for SME, along with several other investments which seek to revitalize the U.S. semiconductor manufacturing base.[182] This Commission supports this bill, which incorporates several of the Commission's first quarter recommendations focused on maintaining U.S. leadership in high-end microelectronics that are key to AI, including by funding research for next generation microelectronics technologies and creating a national laboratory and incubator dedicated to establishing U.S. leadership in microelectronics packaging and manufacturing. The Commission believes the CHIPS Act would create a more competitive U.S. market for semiconductors, revitalize the broader U.S. microelectronics industrial base, incentivize firms to bring additional elements of the manufacturing process back to the United States, and ensure the United States retains global leadership in advanced microelectronics research and development.

D. Increasing Export Control Capacity among U.S. Allies and Partners

As the United States attempts to modernize its own emerging technology export control regime, it will also be essential to work with allies and partners to ensure that they do the same. While unilateral controls can be an effective option when the United States has a monopoly on a given technology, usually this is not the case, and it is essential to work with allies and partners to ensure the global supply of a given technology is controlled. This will be particularly true for AI systems which, as previously discussed, have many different subcomponents, each of which has its own supply chain with a unique geographic dispersal. The technologies that power AI will continuously change, and therefore the United States and its allies will need

---

[180] *TSMC Announces Intention to Build and Operate an Advanced Semiconductor Fab in the United States,* TSMC (May 5, 2020), https://www.tsmc.com/tsmcdotcom/PRListingNewsArchivesAction.do?action=detail&newsid=THGOANPGTH&language=E.

[181] See e.g., Asa Fitch et al., *Trump and Chip Makers Including Intel Seek Semiconductor Self-Sufficiency,* Wall Street Journal, (May 11, 2020), https://www.wsj.com/articles/trump-and-chip-makers-including-intel-seek-semiconductor-self-sufficiency-11589103002; *Letter from Intel Corporation CEO Bob Swan to Deputy Undersecretary of Defense Lisa Porter and Ms. Nicole Petta*, Wall Street Journal (Apr. 28, 2020), https://s.wsj.net/public/resources/documents/intel%20letter.pdf.

[182] See Creating Helpful Incentives to Produce Semiconductors (CHIPS) for America Act, S. 3933, 116th Congress (2020), https://www.congress.gov/bill/116th-congress/senate-bill/3933/titles?r=1&s=1; see also Creating Helpful Incentives to Produce Semiconductors (CHIPS) for America Act, H.R. 7178, 116th Congress (2020), https://www.congress.gov/bill/116th-congress/house-bill/7178.

maximum flexibility to collectively and rapidly control given subcomponents should the need arise. Currently, many U.S. allies lack the domestic legal authority to implement unilateral controls, instead deferring all decisions about regulations to multilateral organizations such as the Wassenaar Arrangement and the European Union.[183]

The Wassenaar Arrangement, a multilateral body with 42 participating states,[184] is the primary international forum responsible for formulating potential controls on conventional and dual-use technologies. However, it has three structural deficiencies which make it ill-suited to be the sole venue through which the United States negotiates export control provisions on emerging technologies with other countries. First, the fact that it operates by consensus means it is slow to react to new technologies and developments, and when it has to revisit controls changes often take years to be implemented.[185] This deficiency is accentuated when dealing with fast-moving technology fields such as AI. Second, it is non-binding, so member states are not legally compelled to follow its guidance. Third, Russia is a member of Wassenaar, which could present challenges if the United States attempts to use the forum to restrict competitors' access to AI or related technologies, given Russia clearly views AI as important to its national security.[186]

Despite these flaws, Wassenaar remains an important body for multilateral coordination on export controls. Many states have linked their domestic export control regimes with Wassenaar, and states with limited regulatory capacity to analyze exports or formulate controls receive substantial benefit from adopting regulations approved by Wassenaar. However, Wassenaar's weaknesses in dealing with emerging technologies such as AI requires the United States to supplement these efforts with strong bilateral and plurilateral efforts in other fora.

---

[183] Norway, for instance, is unwilling to adopt unilateral export controls to particular countries. See Mark Bromley, *Norway's Controls on Arms Exports to China*, SIPRI at 2 (Jan. 2015), https://www.sipri.org/sites/default/files/files/misc/SIPRIBP1502.pdf.

[184] Wassenaar Arrangement member states include Australia, Argentina, Canada, India, Japan, Mexico, New Zealand, Russia, South Africa, South Korea, Turkey, Ukraine, the United Kingdom, the United States, and all EU members other than Cyprus.

[185] Most prominently, in 2013 the Wassenaar Arrangement approved new controls on cyber intrusion software, which were subsequently met with strong pushback from the U.S. cybersecurity community who feared the new controls would inadvertently weaken cybersecurity software. As a result, the United States never implemented the controls, but it took until 2017 for Wassenaar to pass an amendment to fix the problems identified by industry. See Garrett Hinck, *Wassenaar Export Controls on Surveillance Tools: New Exemptions for Vulnerability Research*, Lawfare (Jan. 5, 2018), https://www.lawfareblog.com/wassenaar-export-controls-surveillance-tools-new-exemptions-vulnerability-research.

[186] *'Whoever Leads in AI Will Rule the World': Putin to Russian Children on Knowledge Day*, Russia Today (Sep. 1, 2017), https://www.rt.com/news/401731-ai-rule-world-putin/.

Recommendation 6: State, Commerce, and Treasury should work with allies on legal reforms that would authorize them to implement unilateral export controls and enhance investment screening procedures.

In order to ensure allies can rapidly and most efficiently coordinate export control policies on emerging technologies, the Departments of State and Commerce must urge all allies which have not already done so to pass domestic legislation to overhaul their export control regimes, increasing their internal bureaucratic capacity and providing them with the authorities to implement unilateral export controls. As the United States seeks to overhaul how export controls apply to emerging technologies, it will be critical that allies have the legal authority to implement unilateral controls if necessary. If they do not, it will hinder the U.S. ability to coordinate any new controls among allies and partners, including on technologies key to AI.

This builds on existing work, as State and Commerce have been working closely with allies to grow their domestic export control regulatory bodies and determine alternative avenues of cooperation on export controls beyond Wassenaar, which has a limited capacity to respond quickly to emerging technologies given its size and consensus procedures. This work has been productive and should continue, with an immediate focus on urging all allies to have the proper domestic legal framework in place, particularly with countries that have a strong domestic emerging technology base.

Finally, the Departments of State and the Treasury have worked to enhance the investment screening capabilities of close allies and partners in recent years, an effort which has shown some successes but now takes on increased urgency.[187] It is critical that this effort proceed expeditiously, as U.S. allies must not represent a vulnerability in the overall U.S. investment screening regime, particularly as the Treasury moves to exempt some firms in allied nations from certain CFIUS requirements. Simultaneously, the Departments of State and the Treasury should diligently share data with allies about recent patterns in investment flows both in the United States and in allied countries, to the extent possible given gaps in U.S. and allied disclosure requirements. Doing so will both assist allied efforts to block predatory investments and help illustrate the nature of the threat.

Congress should ensure that efforts to build allied and partner regulatory capacity for export controls and investment screening within the Departments of State, the Treasury, and Commerce are sufficiently resourced. Additionally, to highlight its importance members of Congress should directly raise this issue in future engagements with political leadership from close allies and like-minded partners, as well as with legislative counterparts. Ultimately foreign legislatures are responsible for

[187] See Chris Darby, Gilman Louie, & Jason Matheny, *Mitigating Economic Impacts of the COVID-19 Pandemic and Preserving U.S. Strategic Competitiveness in Artificial Intelligence*, NSCAI at 16 (May 19. 2020), https://www.nscai.gov/reports.

implementing the necessary legal changes, so direct communication with legislators will play an important role.

Part IV: Focusing CFIUS on Preventing the Transfer of Technologies that Create National Security Risks

Export controls are a blunt instrument for preventing technology transfer, but investment screening by CFIUS can be applied more broadly and has power as a deterrent to adversarial capital through signaling. CFIUS screens primarily for controlling investments and investments that provide non-U.S. persons access to sensitive intellectual property. Passive investments, focused purely on financial return, have not traditionally required the same screening and mitigation efforts. FIRRMA broadens CFIUS' application to critical technologies[188] by increasing voluntary and mandatory filing requirements for foreign investors.[189]

However, CFIUS is not currently postured to address the range of threats that the United States faces from adversarial capital from strategic competitors such as China and Russia. This challenge is especially pronounced with respect to emerging technologies. In particular, FIRRMA's reliance on export control lists for identifying critical technologies which require CFIUS filings, as described in previous sections, rests on an incorrect assumption that export controls and investment screening require identical inputs to achieve their goals. This dynamic presents problems because the singular approach prevents CFIUS from being applied more broadly than export controls, which is necessary to mitigate threats from adversarial capital to early-stage companies involved in AI and other emerging technologies. In addition, Commerce's delay in defining and controlling emerging technologies under ECRA has constrained Treasury's ability to expand the scope of CFIUS to new, critical technologies. Finally, FIRRMA also offers CFIUS further opportunities to ease investment screening based on country of origin and investor risk profile.

---

[188] The FIRRMA strengthened and modernized the process through which CFIUS reviews the implications of foreign direct investment (FDI) on behalf of the President. CFIUS sets a legal standard for the President to suspend or block a transaction if, first, no other laws apply and, second, there is "credible evidence" that the transaction poses a national security risk. Any presidential determination must consider the results of the CFIUS national security review and investigation process, including the potential effects of the transaction on U.S. technological leadership in areas affecting U.S. national security (along with 11 other factors).

[189] David McCormick, et al., *Economic Might, National Security, and the Future of American Statecraft*, Texas National Security Review (Summer 2020), https://tnsr.org/2020/05/economic-might-national-security-future-american-statecraft/; Michael Brown & Pavneet Singh, *China's Technology Transfer Strategy*, Defense Innovation Unit Experimental (Jan. 2018), https://admin.govexec.com/media/diux_chinatechnologytransferstudy_jan_2018_(1).pdf [hereinafter Brown & Singh, China's Technology Transfer Strategy].

A. Tailoring CFIUS Requirements to Protect AI and Related Technologies from High-Risk Investors

To date, the U.S. investment screening processes have not imposed stricter requirements on foreign investors based on country of origin. Russian and Chinese investors are not subject to additional filing requirements compared to investors from non-competitor nations. CFIUS should differentiate among foreign investors by country of origin in reviewing investment transactions by identifying specific countries that pose heightened risks as "countries of special concern." This is especially important for transactions involving emerging technologies. China and Russia clearly meet the criteria for "special concern" based on their track records of attempting to acquire U.S. technology through both legal and illegal means.[190] In 2018, Congress and the Department of the Treasury considered a mandatory filing requirement specifically for Chinese investors, sometimes colloquially referred to as a "China deny list," but it was not explicitly included in the final FIRRMA legislation.[191]

Additionally, there are instances in which it may be appropriate to screen investments in an emerging technology prior to enacting export controls. For instance, for early-stage technology venture investments, particularly those which do not yet produce specific products, export controls have historically been ineffective.[192] However, China-based investors in particular have aggressively targeted early-stage U.S. artificial intelligence companies, concluding 81 deals worth over $1.3 billion between 2010 and 2017.[193] Unless the deal resulted in a controlling stake, these transactions would not generally prompt a mandatory CFIUS filing to screen for technology transfer risks, a challenge that continues post-FIRRMA. As highlighted previously in Principle #3, the Commission recommends a more discerning approach on export controls but increased focus on investment screening. To achieve this, Commerce should narrowly tailor export controls on AI in order to avoid unnecessary and substantial harm to U.S. industry, and simultaneously, Treasury should have the flexibility to compel increased disclosure of non-controlling Chinese investments into U.S. AI companies. Doing so would increase awareness regarding Chinese investments in critical technologies, deter state-sponsored IP theft, and preserve U.S. leadership in AI for national security purposes. As ECRA and FIRRMA are currently written, it forces regulators to choose between enacting overly expansive export controls or minimized investment screening for AI.

---

[190] *Information About the Department of Justice's China Initiative and a Compilation of China-Related Prosecutions Since 2018*, Department of Justice (last accessed Jun. 18, 2020), https://www.justice.gov/opa/page/file/1223496/download.
[191] Robert Atkinson, *How to Implement CFIUS to Support U.S Competitiveness*, Information Technology and Innovation Foundation (Jan. 2, 2020), https://itif.org/publications/2020/01/02/how-implement-cfius-support-us-competitiveness.
[192] See Brown & Singh, China's Technology Transfer Strategy.
[193] *Id.* at 29.

Finally, under the current process, CFIUS permits certain investors who are not foreign governments to submit a voluntary five-page short form declaration and receive a response within 30 days.[194] However, estimates suggest that only 10 percent of cases filed with CFIUS using the new short form declaration have been cleared in the expedited 30-day review period.[195] The majority of short form filers were required to submit the longer full disclosure form—also known as a voluntary notice filing—to inform a more extensive national security investigation and review. This is because CFIUS can still require transaction parties to file a full notice after finishing its review of a 30-day declaration, or can conclude its review of a declaration without clearing the investment, meaning that the parties must file a full CFIUS notice to obtain protection against post-closing review of the transaction.[196] Decisions based on mandatory disclosures through the national security investigation and review process generally take two to four months or longer.

Recommendation 7: Grant Treasury the authority to mandate CFIUS filings for non-controlling investments in AI and other sensitive technologies from China, Russia, and other competitor nations.

For investments in AI and other sensitive technologies, CFIUS should require a mandatory disclosure from countries of special concern. Doing so will require a legislative change to Section 721(a) of the Defense Production Act of 1950 (50 USC § 4565(a)) to grant Treasury new authorities for mandatory filing requirements. This change would enable Treasury to mandate CFIUS filings for investments in AI and other sensitive technologies from China, Russia, and other countries of special concern regardless of the technology's export control status. This change is necessary to increase Treasury's visibility into Chinese and Russian non-controlling investments in emerging technologies, as currently their investments in AI companies only require CFIUS filings if the company produces an export controlled good.

A separate list of "sensitive technologies" for the purposes of CFIUS would not be duplicative of existing lists, such as the Department of Commerce's to-be-released list of emerging and foundational technologies. The list of "emerging and foundational technologies" is still necessary, as previous recommendations and the Commission's

---

[194] *Voluntary Notice Filing Instructions (Part 800)*, Department of the Treasury (last accessed June 18, 2020) https://home.treasury.gov/policy-issues/international/the-committee-on-foreign-investment-in-the-united-states-cfius/voluntary-notice-filing-instructions-part-800; Timothy Keeler & Mickey Leibner, *Regulations Expanding Review of Foreign Investment in the US Are Now Effective*, Mayer Brown (Feb. 14, 2020), https://www.mayerbrown.com/en/perspectives-events/publications/2020/02/regulations-expanding-review-of-foreign-investment-in-the-us-are-now-effective.

[195] Judith Lee et al., *CFIUS Reform: Top Ten Takeaways from the Final FIRRMA Rules*, Gibson Dunn (Feb. 19, 2020), https://www.gibsondunn.com/cfius-reform-top-ten-takeaways-from-the-final-firrma-rules/ [hereinafter Lee, CFIUS Reform].

[196] James Barker et al., *Final CFIUS Regulations Implementing FIRRMA Take Effect in February 2020: 10 Key Questions Answered*, Latham & Watkins (Jan. 22, 2020), https://www.lw.com/thoughtLeadership/final-cfius-regs-take-effect-feb-2020-10-key-questions-answered.

Principle #3 for technology protection emphasize. This list should focus on clearly defined and discrete technologies which represent choke points for U.S. strategic competitors for which export controls are appropriate and necessary. The proposed list of "sensitive technologies" could be broader but, per the draft legislative text, mandatory filings would only be required for investments from "countries of special concern." This shift would enable CFIUS to focus on investments from U.S. strategic competitors in relevant sectors without first instituting export controls over the entire sector.

Accordingly, the Commission recommends that Congress amend 50 U.S.C. § 4565 to permit the Department of the Treasury to define a new set of "sensitive technologies," which are not currently subject to export controls but for which CFIUS filings should be mandatory for non-controlling investments involving select U.S. competitors. The Commission recommends that this provision include all non-controlling "sensitive technology" investments from states subject to export restrictions pursuant to section 744.21 of title 15 within the Code of Federal Regulations (China, Russia, and Venezuela), and any state that the Secretary of State designates as a state sponsor of terrorism (Iran, North Korea, Sudan, and Syria). The list of "sensitive technologies" should include any industries key to U.S. national security that face persistent threats from adversarial capital, specifically AI, semiconductors, telecommunications equipment, and quantum computing, as well as other products in the sectors identified in the Made in China 2025 strategic plan.[197] With this legislative change, U.S. competitors' investments in these technologies would be screened without forcing the Department of Commerce to implement broadly defined export controls on these entire fields. This revision should also offer greater transparency and predictability to the private sector, thereby reducing regulatory uncertainty and enhancing deterrence through clearer signaling.

Given this action will result in a significant increase in CFIUS filings, the Commission also recommends that the Department of the Treasury should introduce a shortened, "easy form" of one to two pages for such submissions, to limit the costs to firms and ease the regulatory burden. In its current format, the existing short-form filing, which is approximately five pages, remains burdensome for small companies, often requiring tens of thousands of dollars in legal fees. Making filings mandatory for an increased number of transactions would increase the costs to many tech firms, potentially harming innovation in the process. This shorter form would provide the basic information CFIUS needs to determine whether the transaction requires further review. Under FIRRMA, CFIUS could still require additional information by requesting a short-form or full filing if it is unable to make a determination based on the information provided in the mandatory one to two-page form. Treasury would detail these procedures in new regulations that, among other things, identify the

[197] Adam Hickey, *Deputy Assistant Attorney General Adam S. Hickey of the National Security Division Delivers Remarks at the Fifth National Conference on CFIUS and Team Telecom*, Department of Justice (Apr. 24, 2019), https://www.justice.gov/opa/speech/deputy-assistant-attorney-general-adam-s-hickey-national-security-division-delivers-0.

sectors within this new subset of critical technologies and outline the mandatory filing process for the aforementioned foreign investors. Input from industry, academia, and civil society during the rulemaking process would enable Treasury to craft final regulations that address critical policy goals while engaging in an open and transparent process.

Alternatively, should this recommendation be implemented without introducing a shorter one to two page "easy form" for such mandatory filings, Congress should consider methods for reimbursing firms' CFIUS legal fees, up to a designated limit. This could take the form of a tax incentive or a subsidy to assist with covering the legal fees paid to an attorney to complete the CFIUS short-form paperwork. This would reduce the financial cost of the filing process and incentivize its completion. It is important to note that full filings now require a filing fee, which ranges based on the transaction value up to $300,000 for transactions valued at $750 million and more.[198] Based on the proceeds from filing fees, which took effect on May 1, 2020, CFIUS should be able to begin scaling its review capacity to handle additional filings. However, in light of this additional fee, reducing the burden on mandatory filers either through a shorter form or tax rebate on legal fees would be an important counterbalance.

B. Applying a Risk-Informed Approach to CFIUS Exemptions

In addition to increasing scrutiny on possibly problematic foreign investors, it is also necessary to consider ways to reduce the burden on low-risk actors and allies to promote the free flow of capital. A risk-adjusted approach benefits companies and investors by reducing regulatory burden and allowing CFIUS to focus its time and resources on the transactions requiring the most scrutiny. There are several ways that CFIUS could better account for risk, including by taking into account investors' country of origin, investor type, ownership structure, and investment frequency.[199] In addition to reinvesting the newly added filing fees in screening capacity, Treasury will also be able to improve its capacity for targeted, risk-informed screening if it receives the resources identified in its FY 2021 budget request to scale its IT capabilities and staff.[200]

---

[198] *Fact Sheet: CFIUS Regulation Establishing Filing Fees for Notices*, Department of the Treasury, Office of Public Affairs at 2 (Apr. 27, 2020), https://home.treasury.gov/system/files/206/Fact-Sheet-for-Interim-Rule-on-CFIUS-Filing-Fees.pdf.

[199] Adam Szubin, *Combatting Kleptocracy: Beneficial Ownership, Money Laundering, and Other Reforms*, Testimony before the Senate Committee on the Judiciary (June 19, 2019), https://www.judiciary.senate.gov/imo/media/doc/Szubin%20Testimony.pdf.

[200] *FY 20201 Congressional Budget Justification and Annual Report and Plan*, Department of the Treasury, Committee on Foreign Investment in the United States (last accessed June 18 2020), https://home.treasury.gov/system/files/266/07.-CFIUS-FY-2021-CJ.pdf.

Recommendation 8: Expedite Treasury Department CFIUS exemption standards for allies and partners and create fast tracks for exempting trusted investors.

CFIUS should also adopt a more risk-adjusted approach to investors less likely to serve as channels for adversarial capital by fast-tracking their applications and reducing their filing burden. This requires, first, accelerating exemption standards for allied nations and second creating fast lanes for trusted investors based on track record and category. Combined with Recommendation #7, this would mean that investors from countries such as China and Russia would face the most stringent mandatory process; exempted countries and specific, trusted investors would be fast-tracked and face an expedited process compared to the current regime; and all other investors would be subject to the existing process.

First, the Department of the Treasury should issue clear guidance for which investment screening policies allied nations must implement in order to achieve CFIUS exempted status. CFIUS regulations released in January 2020 created an exception for non-controlling technology, infrastructure, and data (TID) investments for investors tied to "excepted foreign states," with Australia, Canada, and the United Kingdom forming the initial list.[201] CFIUS initially selected these nations due to aspects of their robust intelligence-sharing and defense industrial base integration with the United States.[202] The regulations require that excepted foreign states implement their own process to analyze foreign investments for national security risks and to facilitate coordination with the United States on investment screening by February 2022. In effect, the regulation grants Australia, Canada, and the United Kingdom a two-year grace period to finalize their approach to investment screening and coordination.

However, the Department of the Treasury has yet to publish the criteria CFIUS will use when determining whether additional countries can qualify as "excepted foreign states" in the future.[203] If Treasury can quickly and clearly define the standards for investment screening mechanisms in foreign nations, it will create a powerful incentive for nations to adopt stronger screening mechanisms for adversarial capital. Exemption standards should be tied to finalizing robust domestic investment screening regimes. For example, the European Union established a framework for foreign investment screening in March 2019 but it is still in the process of implementing the associated regulations.[204] The sooner the United States can set

---

[201] See Lee, CFIUS Reform.
[202] See 85 Fed. Reg. 3112.
[203] See Jackson, The Committee on Foreign Investment in the United States (CFIUS) at 18.
[204] *Regulation (EU) 2019/452 of the European Parliament and of the Council of 19 March 2019 establishing a framework for the screening of foreign direct investments into the Union*, Official Journal of the European Union (last accessed July 13, 2020), https://eur-lex.europa.eu/eli/reg/2019/452/oj; see also James Kirschenbaum et al., *EU Foreign Investment Screening - At Last, a Start*, German Marshall Fund (Sept. 24, 2019), https://securingdemocracy.gmfus.org/eu-foreign-investment-screening-at-last-a-start/.

standards for screening and coordination, the more likely it will have an impact on other nations' regulations.

After defining the standards, the next step should be proactively engaging with additional key allies and partners to bring them into the exempted list as quickly as possible. As a starting point, the United States should prioritize cooperation with New Zealand as a member of the Five Eyes intelligence sharing alliance. Japan, South Korea, India, Israel, Singapore, and Taiwan, and the European Union also have AI expertise, advanced industrial bases, and shared values. Working together on investment screening and promoting commercial ties with these like-minded nations benefits collective AI development and would help prevent sensitive technology from falling into the wrong hands.

Second, the Department of the Treasury should also take specific steps to allow CFIUS to differentiate between individual investors to facilitate investments from specific, trusted actors.[205] Beyond the exempted countries listed above, CFIUS has not made exemptions for individual investors from other nations and declined to endorse the concept of 'trusted investors' in its recent rulemaking actions.[206] In terms of filing requirements, CFIUS currently treats foreign investors that are submitting for the first time the same way as ones which have already submitted and been approved one hundred times. There is no certification for investors with a trusted track record. However, firms with a strong history of CFIUS approval could be treated as lower risk and, for example, permitted to file the short-form disclosure rather than the mandatory filing, which would also exempt them from filing fees. In addition to considering the investor's previous interactions with CFIUS, repeated compliance with mitigation agreements, sanctions, and export control regulations could serve as further evidence of trust and assist in fast-tracking specific investors.[207]

The Department of the Treasury has highlighted several submissions calling for a waiver for "trusted investors" in prior public comment periods. While the current

---

[205] For TID investments, CFIUS already differentiates between investment firms, publicly traded companies, and certain other investment vehicles. For example, private equity funds with foreign limited partners are not considered foreign and are therefore outside the jurisdiction of CFIUS if they meet certain conditions. See Richard Gilden & Abbe Dienstag, *The Impact of CFIUS on Private Equity and Hedge Fund Investors*, Kramer Levin (Feb. 27, 2020) https://www.kramerlevin.com/en/perspectives-search/the-impact-of-cfius-on-private-equity-and-hedge-fund-investors.html. Under another exemption, an investment fund is not considered foreign and subject to CFIUS if all of the following criteria are met: its principal place of business is in the U.S.; the general partner or equivalent is a U.S. person, and no foreign limited partners can exercise "control" or have non-controlling investment rights. See Chris Griner et al., *New Decade, New CFIUS: New Rules Expand CFIUS Reach Into Non-Controlling Investment and Real Estate*, Strook (Jan. 22, 2020,) https://www.stroock.com/news-and-insights/new-decade-new-cfius. CFIUS has also created exemptions for certain real estate investments, including commercial office space and housing units. See Jackson, The Committee on Foreign Investment in the United States (CFIUS) at 19.

[206] See 85 Fed. Reg. 3112.

[207] David Hanke, *CFIUS 2.0: Foreign Investors are Watching for CIFUS 'Good' or 'Bad' List*, Law360 (May 28, 2019), https://www.arentfox.com/perspectives/alerts/cfius-20-foreign-investors-are-watching-cfius-good-or-bad-list.

regulations have no such waiver program, the Department of the Treasury should institute one through new regulations.[208] Since some categories of investors, namely publicly traded companies, tend to exhibit lower risk for facilitating problematic technology transfer, the Department of the Treasury should also further refine its requirements according to investment category and relax filing requirements in some areas for lower risk classes of investors.

In totality, this recommendation creates multiple fast tracks, depending on country of origin, investor type, and track record. Lowest risk investors could voluntarily provide information in exchange for faster processing timelines. Investors with an established track record of approved or successfully mitigated deals could also qualify for faster review. These fast tracks would primarily apply to investors from countries that are not already exempted as allies. It could also be implemented more rapidly than the country-base exemptions, which may take time, especially if countries must revise their domestic statutes to improve their investment screening capacity.

---

[208] See e.g., Evan Kielar & Patrick McDonnell, *Treasury Department Implements New Investment Rules*, Lawfare (Mar. 5, 2020), https://www.lawfareblog.com/treasury-department-implements-new-foreign-investment-rules.

TAB 5 — Reorient the Department of State for Great Power Competition in the Digital Age

The intersection of great power competition and rapidly emerging technology is a profound national security challenge.[209] Artificial Intelligence (AI) is intensifying the broader geopolitical struggle between the United States and its competitors, and deepening the challenge democracies face from autocracies.[210] Increasingly, democracy is no longer perceived as the only viable path to economic prosperity and sustainable governance. Authoritarian models suggest that innovation can be planned, that scale matters most, and building technical capacity is more important than free and open thought. By working with a broad network of allies and partners, American diplomats can shape international AI policy to strengthen free societies and check the spread of digital authoritarianism.

In our First Quarter Recommendations, the Commission offered ways to improve AI cooperation among key allies and partners by establishing a National Security Policy Framework for AI Cooperation and pursuing AI-related military concept and capability development with allies and partners, beginning with a focus on the Five Eyes alliance.[211] In this quarter, we focus on recommendations that will empower the United States to play to its strengths and enable the Department of State to lead—and learn from—coalitions of free and open states and organizations to prevail on emerging technology issues in an era of great power competition.

Our recommendations seek to: 1) address China's reorientation of diplomacy for great power competition in a digital age, 2) provide the Department of State with

---

[209] In response to recent great power competition with China and Russia, some scholars have begun reusing the term "political warfare" (originally coined by George F. Kennan in late 1940s) to describe the "synchronized use of any aspect of national power short of overt conventional warfare—such as intelligence assets, alliance building, financial tools, diplomatic relations, to name a few—to achieve state objectives," particularly during times of peace. Kathleen McInnis & Martin Weiss, *Strategic Competition and Foreign Policy: What is "Political Warfare"?*, Congressional Research Services (Mar. 8, 2019), https://crsreports.congress.gov/product/pdf/IF/IF11127.

[210] China's AI strategy is nested within China's two centenaries: the centenary of the Chinese Communist Party in 2021 and centenary of the People's Republic of China in 2049. Xi Jinping noted this at the 19th Party Congress proclaiming that by the first centenary China would be a moderately prosperous society with increased significant economic and technological strength. Xi additionally remarked that China will become a global leader in innovation–which will inextricably be linked to its AI efforts. By the second centenary Xi notes that China will become a global leader in terms of composite national strength and international influence. It should not be understated that China intends on using AI to reach those national goals as a "leapfrog technology." Xi Jinping, Remarks at the 19th National Congress of the Communist Party of China, (Oct. 18, 2017); see also Rob Waugh, *How China is Leading the World in Tech Innovation (And What the West Can Learn From It)*, The Telegraph (Nov. 16, 2018), https://www.telegraph.co.uk/connect/better-business/business-solutions/china-technology-innovation/.

[211] *First Quarter Recommendations*, NSCAI at 64-67 (Mar. 2020), https://www.nscai.gov/reports.

appropriate organization, presence, and AI-related education and training to succeed in great power competition in a digital age, and 3) establish Congressional support to inform appropriate resourcing and policy direction.

The Department of State must be organized and staffed to exert influence in an environment of intensifying geopolitical competition, augmented by emerging technology. Department of State officers must acquire the knowledge and resources to advocate for American interests at the intersection of technology, security, economic interests, and democratic values. The Department should develop a corps of science and technology officers with the AI skills necessary to staff Washington offices and embassies. All officers should master AI fundamentals and receive the required training and tools to identify how emerging trends in AI will impact U.S. interests. They must marshal coalitions of like-minded partners to shape standards, norms, and commerce, while exploiting opportunities for selective and pragmatic cooperation with strategic competitors. A successful approach to "competitive diplomacy" on issues of international AI policy should be treated as a strategic imperative in an era of great power competition.[212]

Background: Reorienting Diplomacy for Great Power Competition in a Digital Age

Department of State officers—like all U.S. officials—are operating in an increasingly competitive global environment. The Commission's November 2019 Interim Report observed that "China, our most serious strategic competitor, has declared its intent to become the world leader in AI by 2030 as part of a broader strategy that will challenge America's military and economic position in Asia and beyond."[213] An expanding and newly assertive diplomacy is the leading edge of that strategy.

China recently surpassed the United States in total number of diplomatic missions, with 276, staffed by a diplomatic corps that has become more vociferous in recent years.[214] Beijing seeks to shape global norms and standards for technologies that will influence the direction of AI development in support of a broader geoeconomic agenda that will challenge the U.S. role and threaten the existing international

---

[212] On "competitive diplomacy," see *National Security Strategy of the United States of America*, The White House at 33 (Dec. 2017), https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf. See also Nadia Schadlow, *Competitive Engagement: Upgrading America's Influence*, Orbis (Fall 2013), https://www.sciencedirect.com/science/article/abs/pii/S0030438713000446.

[213] China's principal AI strategy, issued in 2017, envisions that "by 2030, China's AI theories, technologies, and applications should achieve world-leading levels, making China the world's primary AI innovation center." For a full translation, see Graham Webster et al., *Full Translation: China's New Generation Artificial Intelligence Development Plan*, New America (Aug. 1, 2017), https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.

[214] The United States has 273. See *Global Diplomacy Index*, Lowy Institute (2019), https://globaldiplomacyindex.lowyinstitute.org/; Chun Han Wong & Chao Deng, *China's 'Wolf Warrior' Diplomats are Ready to Fight*, Wall Street Journal (May 19, 2020), https://www.wsj.com/articles/chinas-wolf-warrior-diplomats-are-ready-to-fight-11589896722.

system.[215] China is also making technology agreements abroad to support its companies' AI advances and to enhance the technical capacities of repressive governments.[216] Meanwhile, Beijing has launched a broad effort to dominate the emerging global 5G network architecture.[217]

Beijing has embarked on a global campaign to project influence that incorporates competitive diplomacy with its notable economic, military, and technological maturation.[218] This approach is embodied in strategic initiatives such as the Belt and Road Initiative, Digital Silk Road, military modernization, Military-Civil Fusion, and Smart Cities. AI-related technologies are at the core of these efforts.

Chinese diplomacy is emerging as an impactful tool needed to realize its global aspirations. The increase in diplomatic posts paired with growing aggressive and negative messaging represents a departure from Beijing's previous strategy of Deng Xiaoping's "hiding and biding."[219] Chinese state media now describes a "Wolf Warrior" ethos guiding its diplomatic corps and foreign policy.[220] The COVID-19 crisis has revealed the extent to which coercion and antagonism now pervade Chinese foreign policy.[221] The Chinese Ministry of Foreign Affairs employs

---

[215] In 2018, the Chinese government issued a white paper on AI standards. For an English translation, see Jeffrey Ding & Paul Triolo, *Translation: Excerpts from China's 'White Paper on Artificial Intelligence Standardization'*, New America (June 20, 2018), https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-excerpts-chinas-white-paper-artificial-intelligence-standardization/. For an analysis, see Jeffrey Ding, et al., *Chinese Interests Take a Big Seat at the AI Governance Table*, New America (June 20, 2018), https://www.newamerica.org/cybersecurity-initiative/digichina/blog/chinese-interests-take-big-seat-ai-governance-table/. See also Emily de la Bruyere & Nathan Picarsic, *China Standards 2035: Beijing's Platform Geopolitics and 'Standardization Work in 2020'*, Horizon Advisory (Apr. 2020), https://www.horizonadvisory.org/china-standards-2035-first-report [hereinafter de le Bruyere & Picarsic, China Standards 2035].

[216] For example, some contracts have provided access to data on foreign citizens that can be used to train more sophisticated algorithms for facial recognition. See Amy Hawkins, *Beijing's Big Brother Tech Needs African Faces*, Foreign Policy (July 24, 2018), https://foreignpolicy.com/2018/07/24/beijings-big-brother-tech-needs-african-faces/; see also Steven Feldstein, *The Global Expansion of AI Surveillance*, Carnegie Endowment for International Peace (Sept. 2019), https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847.

[217] See, e.g., Emily Feng, *China's Tech Giant Huawei Spans Much of the Globe Despite U.S. Efforts To Ban It*, National Public Radio (Oct. 24, 2019), https://www.npr.org/2019/10/24/759902041/chinas-tech-giant-huawei-spans-much-of-the-globe-despite-u-s-efforts-to-ban-it.

[218] Martijn Rasser, *Countering China's Technonationalism: A New Approach is Needed if Today's Leaders are to Maintain Their Primacy in Cutting-edge Technology*, The Diplomat (Apr. 24, 2020). https://thediplomat.com/2020/04/countering-chinas-technonationalism/.

[219] The German Marshall Fund found there has been a 300% increase in official Chinese state Twitter accounts in the last year and a fourfold increase in posts. Jessica Brandt & Bret Schafer, *Five Things to Know About Beijing's Disinformation Approach*, German Marshall Fund (Mar. 19, 2020), https://securingdemocracy.gmfus.org/five-things-to-know-about-beijings-disinformation-approach/.

[220] Chun Han Wong & Chao Deng, *China's 'Wolf Warrior' Diplomats are Ready to Fight*, Wall Street Journal (May 19, 2020), https://www.wsj.com/articles/chinas-wolf-warrior-diplomats-are-ready-to-fight-11589896722; Hanna Barczyck, *China's "Wolf Warrior" Diplomacy Gamble*, The Economist (May 28, 2020), https://www.economist.com/china/2020/05/28/chinas-wolf-warrior-diplomacy-gamble.

[221] Kathrin Hille, *Wolf Warrior' Diplomats Reveal China's Ambitions*, Financial Times (May 11, 2020), https://www.ft.com/content/7d500105-4349-4721-b4f5-179de6a58f08.

increasingly aggressive rhetoric against the United States, even promoting conspiracy theories about U.S. involvement with the spread of COVID-19.[222] The People's Republic China's Ministry of Foreign Affairs seeks to rally support at home by criticizing and portraying Western governments' response to COVID-19 as ineffective. This has led to strained relations and interactions between China and many foreign governments.[223]

China is focused and aggressive. Beijing has coupled its modernized, coercive diplomacy with a campaign to establish its authoritarian technology-in-a-box approach as the global standard, with a backbone of cloud infrastructure and edge computing. As early as 2015, Chinese policymakers have prioritized influencing standard-setting bodies viewing them as an instrument of international power competition.[224] Later this year, Beijing intends to release a new plan called "China Standards 2035," which is expected to provide a blueprint for how the Chinese government and leading Chinese companies can lead on and set standards related to a collection of key emerging technologies such as AI, 5G, and the Internet of Things.[225] There is no equivalent, holistic effort inside the United States Government,[226] or a coordinated strategy to respond to this Chinese effort. These efforts complement and reinforce China's aggressive Belt and Road Initiative and Digital Silk Road[227] which enables China to pursue digital infrastructure agreements that reflect their desired technical standards.[228]

---

[222] James Landale, *Coronavirus: China's New Army of Tough Talking Diplomats*, BBC (May 13, 2020), https://www.bbc.com/news/world-asia-china-52562549.

[223] Steven Lee Myers, *China's Aggressive Diplomacy Weakens Xi Jinping's Global Standing*, New York Times (Apr. 20, 2020), https://www.nytimes.com/2020/04/17/world/asia/coronavirus-china-xi-jinping.html.

[224] "[A] popular saying in China posits that third-tier companies make products, second-tier companies make technology, first-tier companies make standards." See John Seaman, *China and the New Geopolitics of Technical Standardization*, French Institute of International Relations (Jan. 27, 2020), https://www.ifri.org/en/publications/notes-de-lifri/china-and-new-geopolitics-technical-standardization.

[225] See de le Bruyere & Picarsic, China Standards 2035.

[226] NIST did publish an AI-specific standards plan in August 2019, which contained recommendations for improving U.S. engagement in AI standards bodies. See *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools*, National Institute of Standards and Technology (Aug. 9, 2019), https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

[227] See J. Ray Bowen II, Testimony before the U.S.-China Economic and Security Review Commission, *A 'China Model?' Beijing's Promotion of Alternative Global Norms and Standards* (Mar. 13, 2020), https://www.uscc.gov/hearings/china-model-beijings-promotion-alternative-global-norms-and-standards ("The PRC makes diplomatic agreements—such as memorandums of understanding—incorporating PRC technical standards extensively within the BRI realm as a major policy component of its action plans.").

[228] Michael Kratsios warned in a speech to European allies that allowing China to set standards and influence technology systems risks "repeating the same mistakes our nations made nearly 20 years ago. . . . Chinese influence and control of technology will not only undermine the freedoms of their own citizens, but all citizens of the world. . . . Technological leadership from democratic nations has never been more of an imperative." *Remarks by U.S. Chief Technology Officer Michael Kratsios at the Web Summit in*

Additionally, the People's Liberation Army's (PLA) has recently increased its international engagements—working in conjunction with civilian diplomatic endeavors[229]—as President Xi Jinping views defense diplomacy as crucial to safeguarding China's "sovereignty, safety, and development interests."[230] Although the United States conducts more military diplomatic activities, China has increased its military diplomatic activities by 110 percent in comparison to its 2003 levels of similar activities (China had 277 military diplomatic activities in 2016 alone).[231] Furthermore, China has used U.N. Peacekeeping engagements and arms sales to increase its international stature, build PLA officers' skills, and develop partnerships, particularly with Asian partners.[232]

The United States must remain cognizant of diplomatic challenges beyond China. For example, Russia has taken steps to develop AI technologies towards applications

---

Lisbon, U.S. Embassy and Consulate in Portugal (Nov. 7, 2019), https://pt.usembassy.gov/u-s-chief-technology-officer-michael-kratsios-addresses-web-summit-2019/; Max Chafkin, U.S. Will Join G-7 AI Pact, Citing Threat from China, Bloomberg (May 27, 2020), https://www.bloomberg.com/news/articles/2020-05-28/g-7-ai-group-adds-u-s-citing-threat-from-china.

[229] Amy Ebitz, *The Use of Military Diplomacy in Great Power Competition: Lessons Learned from the Marshall Plan*, Brookings (Feb. 12, 2019), https://www.brookings.edu/blog/order-from-chaos/2019/02/12/the-use-of-military-diplomacy-in-great-power-competition/; *How is China Bolstering its Military Diplomatic Relations?*, China Power Project, Center for Strategic & International Studies (June 12, 2020), https://chinapower.csis.org/china-military-diplomacy/ [hereinafter How is China Bolstering its Military Diplomatic Relations?]; Kenneth Allen, et al., *Chinese Military Diplomacy, 2003-2016: Trends and Implications*, China Strategic Perspectives (July 17, 2017), https://ndupress.ndu.edu/Media/News/Article/1249864/chinese-military-diplomacy-20032016-trends-and-implications/ [hereinafter Allen, Chinese Military Diplomacy, 2003-2016].

[230] In a January 2015 speech at the All-Military Diplomatic Work Conference, President Xi Jinping stated that military diplomacy is critical in "protecting [China's] sovereignty, safety, and development interests." See How is China Bolstering its Military Diplomatic Relations?; see also Phillip Saunders and Jiunwei Shyy, *Chapter 13: China's Military Diplomacy, China's Global Influence: Perspectives and Recommendations*, Asia-Pacific Center for Security Studies (2019), https://apcss.org/wp-content/uploads/2019/10/13-Chinas-Military-Diplomacy-Saunders-Shyy-rev.pdf.

[231] China has overseen a dramatic increase in Chinese port calls, joint military exercises, senior-level meetings, personnel exchanges, and non-traditional security operations. See How is China Bolstering its Military Diplomatic Relations?; see also Allen, Chinese Military Diplomacy, 2003-2016.

[232] The U.S. proposed a 13% decrease in the State Department's Contributions for International Peacekeeping Account (CIPA) for FY2019, which provides contributions to U.N. peacekeeping operations, U.N. International criminal tribunals, and other mission monitoring funds. The U.S. remains the top financial contributor to U.N. Peacekeeping operations; however, China is the second-highest funder and contributes the most peacekeepers out of any permanent member of the U.N. Security Council. China established an 8,000 troop standalone peacekeeping force and developed a peacekeeping training center where 500 foreign military officials from 69 countries have already been trained. See Christoph Zürcher, *30 Years of Chinese Peacekeeping*, University of Ottawa Centre for International Policy Studies (Jan. 2019), https://www.cips-cepi.ca/wp-content/uploads/2019/01/30YearsofChinesePeacekeeping-FINAL-Jan23.pdf; Luisa Blanchfield, *United Nations Issues: U.S. Funding of U.N. Peacekeeping*, Congressional Research Service (March 23, 2020), https://fas.org/sgp/crs/row/IF10597.pdf; Luisa Blanchfield, *U.S. Funding to the United Nations System: Overview and Selected Policy Issues*, Congressional Research Service (Apr, 25, 2018), https://fas.org/sgp/crs/row/R45206.pdf.

in information operations and media manipulation.[233] During the COVID-19 crisis, Russia has used facial recognition-enabled cameras to identify citizens that violate quarantine orders[234] and spread disinformation, along with China, about the pandemic to undermine democracies.[235] Russia has also taken steps to collaborate with China on high-tech and AI research as well as 5G equipment to undermine America's competitive edge in AI and associated technologies.[236]

This is a pivotal moment for American diplomacy. The efforts of China, and the sustained disinformation campaigns by Russia, threaten to diminish the United States' role in global affairs and its influence on common cause for technology policy. The economic and military implications are recognized by the President, the National Security Council, and Congress.[237] Success requires a reorientation of

---

[233] Russia views AI and other technologies as vital to their security, using AI to create fake content and flood social media through AI-enabled "bots" that "grab information and send messages based on preset algorithmic principles, without human engagement." See Michael Mazarr, et al., *Hostile Social Manipulation: Present Realities and Emerging Trends*, RAND Corporation at 21 (2019), https://www.rand.org/pubs/research_reports/RR2713.html.

[234] Over the past five years, Russia has built a system of over 105,000 facial-recognition enabled cameras throughout Moscow. By mid-March, Russian police claimed these cameras had been used to arrest at least 200 people who had tested positive for COVID-19 and broke quarantine orders. Some Russian citizens worry this surveillance will continue after the pandemic. See Patrick Reevell, *How Russia is Using Facial Recognition to Police its Coronavirus Lockdown*, ABC News (April 30, 2020), https://abcnews.go.com/International/russia-facial-recognition-police-coronavirus-lockdown/story?id=70299736; *Coronavirus: Russia Uses Facial Recognition to Tackle Covid-19*, BBC (April 4, 2020), https://www.bbc.com/news/av/world-europe-52157131/coronavirus-russia-uses-facial-recognition-to-tackle-covid-19.

[235] Russia has spread disinformation linking 5G cell towers to various diseases like brain cancer and Alzheimer's disease, hoping to further disagreements between democracies about 5G. During COVID-19, information operations have linked 5G to COVID-19, leading to at least dozens of arsonist attacks of cell towers in Europe. See Stephanie Bodoni, *China, Russia Are Spreading Virus Misinformation, EU Says*, Bloomberg (June 10, 2020), https://www.bloomberg.com/news/articles/2020-06-10/eu-points-finger-at-china-russia-for-covid-19-disinformation; Travis Andrews, *Why Dangerous Conspiracy Theories About The Virus Spread So Fast - And How They Can Be Stopped*, The Washington Post (May 1, 2020), https://www.washingtonpost.com/technology/2020/05/01/5g-conspiracy-theory-coronavirus-misinformation/; William Broad, *Your 5G Phone Won't Hurt You. But Russia Wants You to Think Otherwise*, New York Times (May 12, 2019), https://www.nytimes.com/2019/05/12/science/5g-phone-safety-health-russia.html

[236] Examples of this collaboration include Huawei purchasing Russian facial recognition technology, building a 5G test zone in Moscow, and developing a joint investment fund for high-tech project with an initial $1 billion budget for AI research. Dimitri Simes, *Huawei Plays Star Role in New China-Russia AI Partnership*, Nikkei Asian Review (Feb. 4, 2020), https://asia.nikkei.com/Spotlight/Asia-Insight/Huawei-plays-star-role-in-new-China-Russia-AI-partnership. An alliance between Russia and Huawei may strengthen China's position in the 5G battle, particularly as Russia's use of Huawei equipment may lead other countries to follow suit. See Alexander Gabuev, *Huawei's Courtship of Moscow Leaves West in the Cold*, The Financial Times (June 21, 2020), https://www.ft.com/content/f36a558f-4e4d-4c00-8252-d8c4be45bde4.

[237] *National Security Strategy to Secure 5G of the United States of America*, The White House (Mar. 2020), https://www.whitehouse.gov/wp-content/uploads/2020/03/National-Strategy-5G-Final.pdf; *5G Supply Chain Security: Threats and Solutions*, U.S. Senate, Committee on Commerce, Science and Transportation (Mar. 4, 2020), https://www.commerce.senate.gov/2020/3/5g-supply-chain-security-

American diplomacy. We turn now to describe the key challenges that need to be addressed to reorient our Department of State.

Issue 1: Department of State's Strategy, Organization, and Expertise for AI Competition

The overseas response to China's strategy is led by the Department of State. The Department has several critical roles to play in AI policy and Great Power competition more broadly. It helps advance U.S. objectives on international AI principles and technology standards that are negotiated in multilateral fora.[238] The Department helps establish cooperative efforts in science and technology with partner nations. Officers posted to U.S. embassies and consulates analyze and report on trends in emerging technologies and their implications for U.S. economic prosperity. They coordinate and orchestrate all elements of U.S. power through the U.S. diplomatic missions overseas. They build coalitions of allies and partners to prevail in competitions with our adversaries and rivals.

Department leadership has made China, great power competition, and technology a central focus in recent months. The Secretary of State has emphasized that competition with China is the central priority for the Department, including in a major speech in November 2019 and another speech in January in Silicon Valley on the technology dimensions of the competition.[239] The Under Secretary of State for Economic Growth, Energy, and the Environment, a veteran of Silicon Valley, is leading the development of a strategy to maintain U.S. technological leadership and build a network of like-minded states, civil society organizations and companies.[240]

---

threats-and-solutions; *5G: National Security Concerns, Intellectual Property Issues, and the Impact on Competition and Innovation*, U.S. Senate, Committee on the Judiciary (May 14, 2019), https://www.judiciary.senate.gov/meetings/5g-national-security-concerns-intellectual-property-issues-and-the-impact-on-competition-and-innovation; *Department of Defense Spectrum Policy and the Impact of the Federal Communications Commission's Ligado Decision on National Security*, U.S. Senate, Committee on Armed Services (May 6, 2020), https://www.judiciary.senate.gov/meetings/5g-national-security-concerns-intellectual-property-issues-and-the-impact-on-competition-and-innovation.

[238] Key multilateral fora related to AI include the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapon systems within the Convention on Certain Conventional Weapons, and the OECD AI Policy Observatory. Additionally, Subcommittee 42 of the Joint Committee between the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC JTC 1/SC 42) is one of the most important multilateral technical standards bodies related to AI, although the State Department does not participate in its meetings.

[239] *The China Challenge*, *Speech*, *Michael R. Pompeo, Secretary of State*, Department of State (Oct. 30, 2019), https://www.state.gov/the-china-challenge/; https://www.state.gov/silicon-valley-and-national-security/.

[240] See *Special Briefing with Senior State and Commerce Department Officials*, Department of State (May 20, 2020), https://www.state.gov/special-briefing-with-keith-krach-under-secretary-of-state-for-economic-growth-energy-and-the-environment-cordell-hull-acting-under-secretary-of-commerce-for-industry-and-security-dr-christophe/.

The Department has aggressively pushed for new technology protection policies, including mandating a "clean path" for its own 5G network traffic and pushing allies to adopt similar standards.[241] It is working closely with like-minded partners both to raise awareness of threats posed by Chinese technological dominance and economic coercion, and to increase their regulatory capacity to protect against such threats.[242]

On AI specifically, the Department of State has made positive, if nascent, steps to invigorate technology diplomacy. Since 2016, the Department has had an AI Policy Small Group to enhance cross-bureau coordination. This group contributed to the successful negotiation of the Organisation for Economic Co-operation and Development (OECD) Principles on AI in May 2019, which were adopted by forty-two countries and represent the first set of internationally-agreed upon principles on AI, espouse the importance of using AI to support human rights and democratic values, demonstrate the benefits of AI, and seek to promote AI research and development as well as remove barriers to innovation.[243] The Department of State has subsequently led U.S. contributions in the OECD AI Policy Observatory.[244] The United States also agreed to support the Global Partnership on AI, an idea initially advanced by France and Canada to create a standing forum among like-minded countries to monitor and debate the policy implications of AI, leading to its formal launch in May 2020.[245] Finally, the Department leads United States participation in the Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapons systems, which was established in 2016 within the Convention on Certain Conventional Weapons.[246]

---

[241] *Secretary Michael R. Pompeo at a Press Availability*, Department of State (Apr. 29, 2020), https://www.state.gov/secretary-michael-r-pompeo-at-a-press-availability-4/. See also, Michael R. Pompeo, E*urope Must Put Security First with 5G*, Politico, (Dec. 2, 2019), https://www.politico.eu/article/europe-must-put-security-first-with-5g-mike-pompeo-eu-us-china/.

[242] The Commission addresses technology protection policies more fully in Tab 4 of this report on "Improve Export Controls and Investment Screening."

[243] *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, OECD (May 21, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. See also *White House OSTP's Michael Kratsios Keynote on AI Next Steps*, U.S. Mission to the OECD (May 21, 2019), https://usoecd.usmission.gov/white-house-ostps-michael-kratsios-keynote-on-ai-next-steps/.

[244] *OECD AI Policy Observatory*, OECD.AI (last accessed June 18, 2020), https://oecd.ai/; John Curran, *OECD Plans AI Policy 'Observatory' Following Standards Adoption*, MeriTalk (July 25, 2019), https://www.meritalk.com/articles/oecd-plans-ai-policy-observatory-following-standards-adoption/.

[245] See Michael Kratsios, *Artificial Intelligence Can Serve Democracy*, Wall Street Journal (May 27, 2020), https://www.wsj.com/articles/artificial-intelligence-can-serve-democracy-11590618319 [hereinafter Kratsios, Artificial Intelligence Can Serve Democracy]; see also *Accelerating American's Leadership in Artificial Intelligence*, Office of Science and Technology Policy (Feb. 11, 2019), https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/.

[246] See *2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, United Nations Office at Geneva (last accessed June 18, 2020), https://www.unog.ch/80256EE600585943/(httpPages)/7C335E71DFCB29D1C1258243003E8724?OpenDocument.

Taken together these actions signal an increasing appreciation of emerging technology as a strategic imperative more than a niche area.[247] However, they represent only the beginning of a necessary reorientation requiring focus, organizational reform, and resources.

Policy issues relating to emerging technology are currently spread across various offices and among several senior leaders including the Science and Technology Advisor to the Secretary.[248] Other important stakeholders for AI and emerging technology policy include elements of bureaus dedicated to specific regions, economic affairs, arms control and international security, human rights and democracy, as well as the Office of the Coordinator for Cyber Issues.[249]

The Department also works to build science and technology partnerships abroad, including through the Office of Science and Technology Cooperation and Science and Technology counselors in a limited number of embassies. Relevant programs for promoting cooperation include the Global Innovation through Science and Technology Initiative and the U.S. Science Envoy program.[250] These are good opportunities to enhance AI-related diplomacy.

The Department has several avenues to exchange views with the private sector. For example, the Bureau of Economic and Business Affairs holds innovation roundtables with the private sector on information and communication technologies, including AI.[251] The Lawrence Eagleburger Fellowship places Foreign Service Officers in one-year assignments with U.S. corporations.

Several fellowship mechanisms help the Department bring more outside scientific experts into temporary government assignments to advise on S&T policy issues. These include the American Association for the Advancement of Science and Technology Policy Fellowship program, the Jefferson Science Fellows program, the Professional Science and Engineering Fellows program, and the Embassy Science

---

[247] See Kratsios, Artificial Intelligence Can Serve Democracy.
[248] See *About Us - Office of the Science and Technology Advisor*, Department of State (Apr. 3, 2019), https://www.state.gov/about-us-office-of-the-science-and-technology-advisor/.
[249] The Office of the Under Secretary for Arms Control and International Security has made notable contributions to current debates surrounding AI policy, international competition, and related issues in a series of white papers, including on diplomatic aspects of "AI, Human-Machine Interaction, and Autonomous Weapons." See Christopher Floyd, *Arms Control and International Security Papers*, Department of State (June 18, 2020), https://www.state.gov/arms-control-and-international-security-papers/.
[250] See *Programs - Office of Science and Technology Cooperation*, Department of State (last accessed July 17, 2020), https://www.state.gov/programs-office-of-science-and-technology-cooperation/.
[251] See *Innovation Roundtables*, Department of State, Bureau of Economic and Business Affairs (last accessed July 17, 2020), https://www.state.gov/innovation-roundtables/.

Fellows program.[252] These are promising avenues for building AI expertise within the Department in the context of other necessary measures.[253]

To make use of AI applications within the Department, the recently established Center for Analytics holds promise.[254] The Center is the Department's first official enterprise-level data and analytics hub dedicated to transforming data into insights for better policy and management decisions. A machine learning practitioners group is making initial efforts to promote adoption of analytical tools and methods across the Department's operations.

Still, there is more to be done to expand and adapt science and technology work within the Department of State for great power competition. AI needs high-level champions among senior leadership, including the Deputy Secretary of State, as well as champions to drive organizational focus and training on technology issues, such as the Under Secretary for Management, the Director General of the Foreign Service, and the Director of the Foreign Service Institute. Bureaus and embassies will need to develop implementation plans, metrics, and talent in ways that prioritize and integrate AI issues within a broader prioritization of emerging technology on par with traditional areas of emphasis—such as regional expertise, foreign languages, and political and economic tradecraft.[255] One example of State moving in this direction is the Global Engagement Center Technology Engagement Team (TET) that runs an impressive tech-scouting process to vet and test tech applications to counter disinformation. The TET creates and runs a multi-stage interagency and international process to adapt these applications to agency use and all cases go into a repository called the Disinfo Cloud for later exploring, sharing, application, or vetting.[256]

We offer the following recommendations to improve the Department's ability to address issues surrounding AI and emerging technologies by increasing senior-level attention to these issues, enhancing the Department's organization and building its

---

[252] See *AAAS Science & Technology Fellowship Program*, Department of State (last accessed July 17, 2020), https://careers.state.gov/work/fellowships/aaas/; *Jefferson Science Fellows Program*, Department of State (last accessed July 17, 2020), https://careers.state.gov/work/fellowships/jefferson-science/; *Professional Science & Engineering Society Fellows Program*, Department of State (last accessed July 17, 2020), https://careers.state.gov/work/fellowships/science-engineering-society/; *Embassy Science Fellows Program*, Department of State (last accessed July 17, 2020), https://www.state.gov/programs-office-of-science-and-technology-cooperation/embassy-science-fellows-program/. The Department also can utilize the Intergovernmental Personnel Act to bring in outside experts.

[253] Manisha Singh, *Enabling the Future of Artificial Intelligence Innovation*, Department of State (Sept. 20, 2019), https://www.state.gov/enabling-the-future-of-artificial-intelligence-innovation/.

[254] See *Establishment of the Center for Analytics*, Department of State (Jan. 17, 2020), https://www.state.gov/establishment-of-the-center-for-analytics/.

[255] State Department plans consist of the Joint Strategic Plan, Joint Regional Strategies, Functional Bureau Strategies, Integrated Country Strategies, and Annual Performance Plans and Reports to the President, and Congress.

[256] See *Tackling Adversarial Propaganda and Disinformation*, Disinfo Cloud (last accessed June 18, 2020), https://disinfocloud.com/.

capacity both domestically and overseas, and improving training programs across the Department.

Recommendation 1: The Secretary of State should establish a senior-level Strategic Innovation and Technology Council within the Department.

The purpose of the Council should be to:

1) drive Departmental reorientation around great power and technology competition;
2) focus U.S. foreign policy, organization, and resources to lead coalitions;
3) foster interagency, international, and public-private partnerships that offer competitive alternatives to economic coercion, technology, and disinformation by foreign competitors;
4) build capacities that adapt the Foreign and Civil Service to be effective advocates in the context of AI-related developments and trends;
5) build a Digital Modernization and Readiness Partnership with Congress;
6) incorporate AI and analytics to improve foreign policy and management decisions; and
7) integrate AI-related objectives and metrics into State Department planning.

Recommendation 2: The Department of State and Congress should expedite efforts to establish the proposed Bureau of Cyberspace Security and Emerging Technology (CSET).

In its Interim Report, the Commission noted the need to rapidly establish a Bureau of Cyberspace Security and Emerging Technology (CSET) within the Department. The Commission agreed with the Department's proposal to designate the Under Secretary for Arms Control and International Security as the principal to oversee the new bureau.[257]

Department officials have identified key organizational gaps in effective diplomatic, intra-agency, and interagency engagement on the security elements of AI and other emerging technologies. The proposed CSET bureau would serve as the focal point and champion for the security challenges associated with emerging technologies, and provide a clear home for AI within the Department. The bureau would lead cooperative efforts abroad, and offer a diplomatic counterpart to existing DoD and IC efforts to promote cooperation on AI. The bureau would also signal to allies that the United States takes seriously the security implications of emerging technologies, and may encourage allies to undertake similar organizational reforms.

[257] *Interim Report*, NSCAI at 45 (Nov. 2019), https://www.nscai.gov/reports.

The CSET bureau would be instrumental to existing and future Department-wide efforts to drive high-level dialogues with allies and partners to further progress and cooperation in critical areas related to AI, such as promoting common data-sharing and test, evaluation, verification, and validation frameworks; enhancing interoperable capabilities and decision-making procedures; protecting intellectual property; establishing international norms and standards on responsible deployment and use; deepening foreign assistance cooperation related to emerging technologies; raising awareness, sharing best practices, and building capacity on export controls and foreign investment screening; and fostering and protecting joint research and development.

In June 2019, the Department submitted its proposal to Congress to establish the CSET bureau.[258] However, disagreements with Congress over where the bureau should be housed within the Department have stalled its implementation.[259] Given the urgency of enhancing allied cooperation on emerging technologies, the clear security implications, and aggressive Chinese efforts to drive wedges between the United States and its allies, the Commission recommends that the Department of State and Congress implement the proposal without further delay. The Department should move forward promptly to coordinate with key congressional committees for authorization to establish and seek funding for CSET personnel and responsibilities.

Recommendation 3: The Department of State should enhance its presence in major foreign and U.S. technology hubs and establish a cadre of dedicated technology officers at U.S. embassies and consulates to strengthen diplomatic advocacy, improve technology scouting, and inform policy and foreign assistance choices.

The Department of State's global network of embassies, consulates, and other outposts provides U.S. diplomats with a presence in major technology hubs around the world.[260] This presence gives reporting officers a front-row seat on emerging technology trends. Political and economic officers posted to these hubs also defend U.S. policy positions in engagements with foreign government counterparts. Many embassies include "environment, science, technology, and health" sections, some of which are staffed by Science and Technology counselors and the remainder of which are covered by Foreign Service Officers as a collateral duty. They serve as a focal

[258] The State Department proposed an 80-person bureau, led by a Senate-confirmed ambassador-at-large, with a projected budget of $20.8 million. See Sean Lyngaas, S*tate Department Proposes New $20.8 Million Cybersecurity Bureau*, Cyberscoop (June 5, 2019), https://www.cyberscoop.com/state-department-proposes-new-20-8-million-cybersecurity-bureau/.
[259] For additional background, see James Lewis, *End the Uncertainty about Cybersecurity at State*, Center for Strategic & International Studies (Oct. 16, 2019), https://www.csis.org/analysis/end-uncertainty-about-cybersecurity-state.
[260] See Foreign Affairs Manual, *Post Types of Diplomatic and Consular Posts* (2 FAM 131), Department of State, (May 8, 2020), https://fam.state.gov/FAM/02FAM/02FAM0130.html.

point for AI issues, report on technology developments in their countries, and coordinate U.S. technology initiatives with local officials and experts.

The Commission continues to examine recommendations that could enhance the ability of the Department of State to better support U.S. technology-related efforts at posts abroad. At this stage, the Commission offers some immediate steps to strengthen the Department of State's technology-related presence outside of Washington. For instance, when the proposed CSET bureau is established, the Department should enhance its posture abroad by enabling civil servants within the bureau to serve in rotational assignments as technology officers to relevant overseas posts. This would bolster the capacity to conduct effective diplomacy on technology issues, while also building the expertise of CSET bureau officers. In addition, as the Department reallocates Foreign Service billets from large missions, such as Afghanistan and Iraq, it should identify opportunities to assign more officers for training and assignment on emerging technology issues in important foreign technology hubs. The Department should also recruit highly skilled experts on AI under specialized hiring authorities to work directly with senior Department officials and foreign counterparts.

Within the United States, the Department of State posted its first representative to Silicon Valley in 2016-2017. The Commission urges the Department to consider reestablishing such a position, and also to examine further opportunities to engage across the United States with U.S. companies, universities, and others on AI policy. For example, the Department maintains a wide network of Diplomats in Residence, who often work at universities and who reach every region of the United States.[261] These officers are well positioned to enhance the Department's connections with America's AI community and bring that experience back to the Department.

Recommendation 4: The Department of State should incorporate AI-related technology modules into key Foreign Service Institute training courses, including the Ambassadorial Seminar, the Deputy Chiefs of Mission course, Political and Economic Tradecraft courses, and A-100 orientation training classes. FSI should also develop a stand-alone course on emerging technologies and foreign policy.

Embassy staff need a deeper understanding of technology to be effective advocates for democratic interests and values related to AI. Some training courses at the Foreign Service Institute (FSI) already include AI-related modules.[262] But coverage of AI issues is lacking in senior-level courses, including the Ambassadorial Seminar and

[261] See *Diplomats in Residence*, Department of State (last accessed July 17, 2020), https://careers.state.gov/connect/dir/.

[262] FSI courses that incorporate AI-related modules include: Environment, Science, Technology and Health Tradecraft (PE305; a two-week training); International Digital Economy Policy: Internet and Telecommunications Diplomacy (PE131; a two-day training); and the Digital Economy Officer Course (1 week).

the Deputy Chiefs of Mission course, as well as the basic Foreign Service orientation course (A-100). An additional, stand-alone AI course is also needed to make a deeper understanding of AI technology accessible to foreign affairs professionals.

Some Department of State cones (specialties), such as economic, public affairs, and public diplomacy officers, are making notable efforts to focus attention on AI-related issues.[263] Still, there is a pressing need to improve the skills of all Civil Service and Foreign Service officers to compete amid the information and influence operations waged by adversaries, which AI technologies will pervade.

Issue 2: Congressional Support and Resourcing for the State Department

Congress is a critical partner in accelerating the Department of State's reorientation toward great power competition in a digital age. In the past, successful Department of State reform initiatives have benefited from strong congressional support. For example, during Secretary of State Colin Powell's tenure, support in Congress for the Diplomatic Readiness Initiative enabled that effort to add to the Department's roster over one thousand Foreign Service Officers and specialists, and over two hundred Civil Service positions.[264] The initiative also improved professional development and training at the FSI, modernized the Department's information technology, and improved embassy security.

Today, Congress can provide the full-time equivalent authorities and funding to drive successful implementation of the Department of State's Strategic Framework for International Engagement on AI, adoption of AI and data analytics in the Department's decision making and operations, an enhanced diplomatic presence in foreign and domestic technology hubs, and improved training and education programs.

The Department's Legislative Affairs Bureau should lead an intra-Departmental effort to expand and deepen contacts with relevant congressional committees to reach agreement on needed authorities, staffing, and funding for Departmental AI initiatives.

---

[263] The 2019 Public Diplomacy Conference included sessions on great power competition, "AI for good," and countering disinformation. Other professional communities should make similar efforts, for example in connection with the Annual Chiefs of Mission Conference or individual bureau conferences.

[264] These efforts had their origin in the 2001 Independent Task Force on State Department Reform. See Frank Carlucci & Ian Brzezinski, *State Department Reform: Report of an Independent Task Force*, Council on Foreign Relations and the Center for Strategic and International Studies (2001), https://cdn.cfr.org/sites/default/files/pdf/2005/10/state_department.pdf; see also Shane Harris, *Powell's Army*, Government Executive (Nov. 1, 2003), https://www.govexec.com/magazine/2003/11/powells-army/15328/.

Recommendation 5: Congress should conduct hearings to assess the Department of State's posture and progress in reorienting to address emerging technology dimensions of great power competition.

Hearings should focus primarily on assessing the implementation of the Department's Strategic Framework for AI. Congress should invite senior Department officials as well as outside experts in technology and foreign policy, to provide independent assessments and recommendations and to ascertain needed funding and authorities. In addition, Congress should address and pass comprehensive foreign affairs reauthorization legislation, which it has not done since 2003.[265] Doing so would encourage productive debate over policy priorities, clarify the urgent need to establish a CSET bureau, and focus the minds of the American public on the importance of enhancing U.S. diplomacy and development around AI and other emerging technologies. Hearings may also inform broader resourcing questions, including whether the Department can effectively pivot to address the national security challenges associated with emerging technologies and great power competition while simultaneously cutting its budget by over 20 percent, as the Administration has proposed in each of the last four fiscal years but Congress has never implemented.[266]

---

[265] See Cory Gill & Emily Morgenstern, *Foreign Relations Reauthorization: Background and Issues*, Congressional Research Service (June 27, 2019), https://fas.org/sgp/crs/row/IF10293.pdf.
[266] See *International Affairs Budgets*, Department of State (last accessed June 18, 2020), https://www.state.gov/plans-performance-budget/international-affairs-budgets/.

TAB 6 — Implement Key Considerations as a Paradigm for Responsible Development and Fielding of Artificial Intelligence

We stated in our Interim Report that "defense and national security agencies must develop and deploy Artificial Intelligence (AI) in a responsible, trusted, and ethical manner to sustain public support, maximize operational effectiveness, maintain the integrity of the profession of arms, […]strengthen international alliances," and preserve democratic values in the U.S. and abroad.[267] Agencies need practical guidance for implementing commonly agreed upon AI principles and a more comprehensive strategy to develop and field AI responsibly.

Issues surrounding the responsible development and fielding of AI technologies for national security are wide-ranging, complex, and unique to the context of each use case. Debates are ongoing as the technology and its applications rapidly evolve, and the need for norms and best practices becomes more apparent. Entities in the government, civil society, and the private sector have undertaken critical steps to establish ethics guidance for AI, both domestically and globally.[268] Yet, while some agencies critical to national security have adopted[269] or are in the process of adopting AI principles,[270] others lack this guidance entirely. And even when guidance is available in the form of principles, it can be difficult to translate such high-level concepts into concrete actions. Agencies must not only articulate their aspirations with respect to ethics and responsible use of AI, but also operationalize them. Agencies would benefit from inter-agency consistency in prioritizing the recommended practices in the key categories that are detailed below, regardless of

---

[267] *Interim Report*, NSCAI at 16 (Nov. 2019), https://www.nscai.gov/reports.

[268] Examples of efforts to establish ethics guidelines are found within the U.S. government (*see* Memorandum from Russell T. Vought, Acting Director, Office of Management and Budget, to the Heads of Executive Departments and Agencies (2020), https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf); industry (*see* Jessica Fjeld, et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center (Jan. 15, 2020), https://cyber.harvard.edu/publication/2020/principled-ai); and internationally (*see Principles on Artificial Intelligence*, Organization for Economic Cooperation and Development (May 2019), https://www.oecd.org/going-digital/ai/; *Ethical Guidelines for Trustworthy AI*, High Level Expert Group, The European Commission (Apr. 8, 2019), https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top).

[269] The Department of Defense took the critical step of adopting high-level principles to guide its development and use of AI. See C. Todd Lopez, *DOD Adopts 5 Principles of Artificial Intelligence Ethics*, Department of Defense (Feb. 25, 2020), https://www.defense.gov/Explore/News/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/.

[270] Caroline Henry, *2020 Spring Symposium: Building an AI Powered IC Event Recap*, INSA (Mar. 9, 2020), https://www.insaonline.org/2020-spring-symposium-building-an-ai-powered-ic-event-recap/.

the department's respective AI principles. This is especially true given the need to have interoperability across a variety of AI systems.

Recommendation: Heads of departments and agencies should implement the Key Considerations as a paradigm for the responsible development and fielding of AI systems. This includes developing processes and programs aimed at adopting the paradigm's recommended practices, monitoring their implementation, and continually refining them as best practices evolve.

The Commission has created a paradigm for operationalizing ethical AI principles that we recommend national security agencies implement. Our implementation document entitled *Key Considerations for Responsible Development and Fielding of AI* offers a set of disciplinary perspectives for identifying challenges with responsibly developing and fielding AI systems unique to each use case, and concrete, readily accessible actions that help address these challenges. This document captures key categories from the AI lifecycle through which one can more readily identify context specific practices to achieve the ethical and responsible development and fielding of AI.

Our paradigm lays out high-level considerations and recommended practices in each of five categories that are broadly applicable across agencies: (1) **Values**, (2) **Engineering Practices**, (3) **System Performance**, (4) **Human-AI Interaction**, and (5) **Accountability and Governance**. Being high-level, they grant flexibility to departments and agencies to consider them within the broader context of their risk management processes, according to the department, application, and specific use case contexts. The recommended practices span ethical considerations (e.g. practices for aligning system development and use with American values) and broader considerations for responsible AI (e.g., engineering practices for reliability, robustness, and resilience to Machine Learning (ML) attacks).

Implementation of these Key Considerations and recommended practices implies that agencies will adopt certain AI systems requirements, which in turn will become an integral part of an agency's broader risk assessment process when deciding whether and how to develop and field AI. As part of this risk assessment, the agency could weigh the trade-offs of applying a particular recommended practice based on the specific use context. An agency could weigh whether to follow a certain recommended practice in a categorical area against the risk or cost it might cause in another area. This risk analysis could result in the agency deciding a recommended practice is less relevant to the specific application or context (e.g., the practice of testing for multi-agent interaction will not be applicable for some AI applications) or inform the degree to which to execute the recommended practice (e.g., an agency could decide that designing for interpretability is more appropriate than full explainability).

These recommended practices should apply both to systems that are developed by departments and agencies, as well as those that are acquired.[271]

For completeness, we offer an outline of these considerations and recommended practices below, but a more thorough explanation of each is found in the Key Considerations document. (Please see Appendix A-1 for a condensed Key Considerations document designed for government leaders and the public. For an extended version of the document with technical details for implementers, please see Appendix A-2.)

Outline

I. Aligning Systems and Uses with American Values and the Rule of Law

A. Developing uses and building systems that behave in accordance with American values and the rule of law
    1. Employing technologies and operational policies aligning with privacy preservation, fairness, inclusion, human rights, and law of armed conflict.

B. Representing objectives and trade-offs
    1. Consider and document value considerations in AI systems and components based on specifying how trade-offs with accuracy are handled.
    2. Consider and document value considerations in AI systems that rely on representations of objective or utility functions.
    3. Conduct documentation, reviews, and set limits on disallowed outcomes.

II. Engineering Practices

1. Concept of operations development, and design and requirements definition and analysis
2. Documentation of the AI lifecycle
3. Infrastructure to support traceability, including auditability and forensics
4. Security and robustness: addressing intentional and unintentional failures
5. Conduct red teaming

III. System Performance

A. Training and testing (including performance and performance metrics)
    1. Standards for metrics & reporting
        a. Consistency across testing/test reporting

[271] Systems acquired (commercial-off-the-shelf (COTS) or through contractors) should be subjected to the same rigorous standards and best practices—either in the acquisitions or acceptance processes.

b. Testing for blind spots
c. Testing for fairness
d. Articulation of performance standards and metrics

2. Representativeness of data and model for the specific context at hand
3. Evaluating an AI system's performance relative to current benchmarks
4. Evaluating aggregate performance of human-machine teams
5. Reliability and robustness
6. For systems of systems, testing machine-machine/multi-agent interaction

B. Maintenance and deployment

1. Specifying maintenance requirements
2. Continuously monitoring and evaluating AI system performance
3. Iterative and sustained testing and validation
4. Monitoring and mitigating emergent behavior

IV. Human-AI Interaction

A. Identification of functions of humans in design, engineering, and fielding of AI

1. Define functions and responsibilities of human operators and assign them to specific individuals.
2. Policies should define the tasks of humans across the AI lifecycle.
3. Enable feedback and oversight to ensure that systems operate as they should.

B. Explicit support of human-AI interaction and collaboration

1. Human-AI design guidelines
2. Algorithms and functions in support of interpretability and explanation
3. Designs that provide cues to the human operators about the level of confidence the system has in the results or behaviors of the system
4. Policies for machine-human handoff
5. Leveraging traceability to assist with system development and understanding
6. Training

V. Accountability and Governance

1. Identify responsible actors
2. Adopt technology to strengthen accountability processes and goals
3. Adopt policies to strengthen accountability
4. External oversight support

Proposed Executive Branch Action

Heads of departments and agencies critical to national security (at a minimum, the Department of Defense, Intelligence Community, Department of Homeland

Security, Federal Bureau of Investigation, Department of Energy, Department of State, and Department of Health and Human Services) should implement the Key Considerations as a paradigm for the responsible development and fielding of AI systems. This includes developing processes and programs aimed at adopting the paradigm's recommended practices, monitoring their implementation, and continually refining them as best practices evolve.

This approach would set the foundation for an intentional, government-wide, coordinated effort to incorporate recommended practices into current processes for AI development and fielding. However, our overarching aim is to allow agencies to continue to have the flexibility to craft policies and processes according to their specific needs. The Commission is mindful of the required flexibility that an agency needs when conducting the risk assessment and management of an AI system, as these tasks will largely depend on the context of the AI system.

Appendix A-1 — Key Considerations for Responsible Development & Fielding of AI (Abridged Version)

Introduction

The Commission acknowledges the efforts undertaken to date to establish ethics guidelines for AI systems.[1] While some national security agencies have adopted,[2] or are in the process of adopting, AI principles,[3] other agencies have not provided such guidance. In cases where principles are offered, it can be difficult to translate the high-level concepts into concrete actions. In addition, agencies would benefit from the establishment of greater consistency in policies to further the responsible development and fielding of AI technologies across government.

This Commission is identifying a set of challenges and making recommendations on directions with responsibly developing and fielding AI systems, and for pinpointing the concrete actions that should be adopted across the government to help overcome these challenges. Collectively, they form a paradigm for aligning AI system development and AI system behavior to goals and values. The first section, *Aligning Systems and Uses with American Values and the Rule of Law*, provides guidance specific to implementing systems that abide by American values, most of which are shared by democratic nations. The section also covers aligning the run-time behavior of systems to the related, more technical encodings of objectives, utilities, and trade-offs. The four following sections (on *Engineering Practices, System Performance, Human-AI Interaction*, and *Accountability & Governance*) serve in support of core American values and further outline practices needed to develop and field systems that are trustworthy, understandable, reliable, and robust.

Recommended practices span multiple phases of the *AI lifecycle*, and establish a baseline for the responsible development and fielding of AI technologies. The Commission uses "development" to refer to 'designing, building, and testing during development and prior to deployment' and "fielding" to refer to 'deployment, monitoring, and sustainment.'

The Commission recommends that heads of departments and agencies implement the Key Considerations as a paradigm for the responsible development and fielding of AI systems. This includes developing processes and programs aimed at adopting the paradigm's recommended practices, monitoring their implementation, and continually refining them as best practices evolve. These recommended practices should apply both to systems that are developed by departments and agencies, as well as those that are acquired. Systems acquired (whether commercial off-the-shelf systems or through contractors) should be subjected to the same rigorous standards and recommended practices—whether in the acquisitions or acceptance processes.

As such, the government organization overseeing the bidding process should require assertions of goals aligned with recommended practices for the Key Considerations in the process.

In each of the five categorical areas that follow, we first provide a conceptual overview of the scope and importance of the topic. We then illustrate examples of a current challenge relevant to national security departments that underscores the need to adopt recommended practices in this area. Then, we provide a list of recommended practices that agencies should adopt, acknowledging research, industry tools, and exemplary models within government that could support agencies in the adoption of recommended practices. (For more details on important aspects and implementation guidance for each of the recommended practices listed, see Appendix A-2 for the full Key Considerations document.) Finally, in areas where best practices do not exist or are especially challenging to implement, we note the need for future work as a priority; this includes, for example, R&D and standards development. We also identify potential areas in which collaboration with allies and partners would be beneficial for interoperability and trust, and note that the Key Considerations can inform potential future efforts to discuss military uses of AI with strategic competitors.

I. Aligning Systems and Uses with American Values and the Rule of Law

(1) Overview

Our values guide our decisions and our assessment of their outcomes. Our values shape our policies, our sensitivities, and how we balance trade-offs among competing interests. Our values, and our commitment to upholding them, are reflected in the U.S. Constitution, and our laws, regulations, programs, and processes.

One of the seven principles we set forth in our Interim Report (November 2019) is the following:

> The American way of AI must reflect American values—including having the rule of law at its core. For federal law enforcement agencies conducting national security investigations in the United States, that means using AI in ways that are consistent with constitutional principles of due process, individual privacy, equal protection, and non-discrimination. For American diplomacy, that means standing firm against uses of AI by authoritarian governments to repress individual freedom or violate the human rights of their citizens. And for the U.S. military, that means finding ways for AI to enhance its ability to uphold the laws of war and ensuring that current frameworks adequately cover AI.

Values established in the U.S. Constitution, and further operationalized in legislation, include freedoms of speech and assembly, the rights to due process, inclusion, fairness, non-discrimination (including equal protection), and privacy (including protection from unwarranted government interference in one's private affairs). These values are codified in the U.S. Constitution and the U.S. Code.[4] Our values also are found in international treaties that the United States has ratified that affirm our commitments to human rights and human dignity.[5] Within America's national security departments, our commitment to protecting and upholding privacy and civil liberties is further embedded in the policies and programs of the Intelligence Community,[6] the Department of Homeland Security,[7] the Department of Defense (DoD),[8] and oversight entities.[9] In the military context, core values such as distinction and proportionality are embodied in the nation's commitment to, and the DoD's policies to uphold, the Uniform Code of Military Justice and the Law of Armed Conflict.[10] Other values are reflected in treaties, rules, and policies such as the Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment;[11] the DoD's Rules of Engagement;[12] and the DoD's Directive 3000.09.[13] While not an exhaustive list of U.S. values, the paradigm of considerations and recommended practices for AI that we introduce resonate with these values as they have been acknowledged as critical by the U.S. government and national security departments and agencies. Further, many of these values are common to America's like-minded partners who share a commitment to democracy, human dignity, and human rights.

Our values demand that the development and use of AI respect these foundational values, and that they enable human empowerment as well as accountability. They require that the operation of AI systems and components be compliant with our laws and international legal commitments, and with our departmental policies. In short, American values must inform the way we develop and field AI systems, and the way our AI systems behave in the world.

In the more comprehensive document (Appendix A-2), we provide additional details and references for technical implementers and note where recommendations would support the fulfillment of the high-level AI principles that have been adopted by the Secretary of Defense.

(2) Examples of Current Challenges

Machine learning (ML) techniques can assist DoD agencies with large-scale data analyses to support and enhance decision making about personnel. As an example, the Proposed New Disability Construct (PNDC) seeks to leverage data analyses to identify service members on the verge of ineligibility due to concerns with their readiness. Other potential analyses can support personnel evaluations, including analyzing factors that lead to success or failure in promotion. Caution and proven practices are needed, however, to avoid pitfalls in fairness and inclusiveness, several of which have been highlighted in high-profile challenges in areas like criminal justice, recruiting and hiring, and face recognition.[14] Attention should be paid to

challenges with decision support systems to avoid harmful disparate impact.[15] Likewise, factors weighed in performance evaluations and promotions must be carefully considered to avoid inadvertently reinforcing existing biases through ML-assisted decisions.[16]

(3) Recommendations for Adoption

A. **Developing uses and building systems that behave in accordance with American values and the rule of law.** To implement core American values, it is important to:
   1. **Employ technologies and operational policies that align with privacy preservation, fairness, inclusion, human rights, and the law of armed conflict (LOAC).** Technologies and policies throughout the AI lifecycle should support achieving these goals; they should ensure that AI uses and systems are consistent with these values and mitigate the risk that AI system uses/outcomes will violate these values.

B. **Representing Objectives and Trade-offs**. Another important practice for aligning AI systems with values is to consider values as (1) embodied in choices about engineering trade-offs and (2) explicitly represented in the goals and utility functions of an AI system.[17] Recommended Practices for Representing Objectives and Trade-offs include the following:
   1. **Consider and document value considerations in AI systems and components based on specifying how trade-offs with accuracy are handled;** this includes operating thresholds that yield different true positive and false positive rates or different precision and recall.
   2. **Consider and document value considerations in AI systems that rely on representations of objective or utility functions,** including the handling of multi-attribute or multi-objective models.
   3. **Conduct documentation, reviews, and set limits on disallowed outcomes.**

(4) Recommendations for Future Action

Future R&D. R&D is needed to advance capabilities for preserving and ensuring that developed or acquired AI systems will act in accordance with American values and the rule of law. For instance, the Commission notes the need for R&D to assure that the personal privacy of individuals is protected in the acquisition and use of data for AI system development.[18] This includes advancing ethical practices with the use of personal data, including disclosure and consent about data collection and use models (including uses of data to build base models that are later retrained and fine-tuned for specific tasks), the use of anonymity techniques and privacy-preserving technologies, and uses of related technologies such as multiparty computation (to allow collaboration on the pooling of data from multiple organizations without sharing datasets). Additionally, we need to understand the compatibility of data usage policies and privacy preserving approaches with regulatory approaches such as the European Union's General Data Protection Regulation (GDPR).

II. Engineering Practices

(1) Overview

The government, and its partners (including vendors), should adopt recommended practices for creating and maintaining trustworthy and robust AI systems that are *auditable* (able to be interrogated and yield information at each stage of the AI lifecycle to determine compliance with policy, standards, or regulations[19]*; traceable* (to understand the technology, development processes, and operational methods applicable to AI capabilities, e.g., with transparent and auditable methodologies, data sources, and design procedure and documentation[20]); *interpretable* (to understand the value and accuracy of system output[21])*, and reliable* (to perform in the intended manner within the intended domain of use[22]). There are no broadly directed best practices or standards to guide organizations in the building of AI systems that are consistent with designated AI principles, but candidate approaches, minimal standards, and engineering proven practices are available.[23]

Additionally, several properties of the methods and models used in ML (e.g., data-centric methods) are associated with weaknesses that make the systems brittle and exploitable in specific ways—and vulnerable to failure modalities not seen in traditional software systems. Such failures can rise inadvertently or as the intended results of malicious attacks and manipulation.[24] Recent efforts integrate adversarial attacks[25] and unintended faults throughout the lifecycle[26] into a single framework that recognizes intentional and unintentional failure modes.[27]

Intentional failures are the result of malicious actors explicitly attacking some aspect of (AI) system behavior. Taxonomies on malicious attacks explain the rapidly developing Adversarial Machine Learning *(*AML*)* landscape. Attacks span ML training and testing and each have associated defenses.[28] Categories of intentional failures introduced by adversaries include training *data poisoning* attacks (contaminating training data), *model inversion* (recovering secret features used in the model through careful queries), and ML *supply chain attacks* (comprising the ML model as it is being downloaded for use).[29] National security uses of AI will be the subject of sustained adversarial efforts; AI developed for this community must remain current with a rapidly developing understanding of the nature of vulnerabilities to attacks as these attacks grow in sophistication. Technical and process advances that contribute to reducing vulnerability and to detecting and alerting about attacks must also be monitored routinely.

Unintentional failures can be introduced at any point in the AI development and deployment lifecycle. In addition to faults that can be inadvertently introduced into any software development effort, distinct additional failure modes can be introduced for machine learning systems.

Examples of unintentional AI failures include *reward hacking* (when AI systems act counter to the intent of the programmed rules because of a mismatch between stated

reward and real reward) and *distributional shifts* (when a system is tested in one kind of environment, but is unable to adapt to changes in other kinds of environment).[30] Another area of failure includes the inadequate specification of values per objectives represented in system utility functions (as described in Section 1 above on *Representing Objectives and Trade-offs*), leading to unexpected and costly behaviors and outcomes, akin to outcomes in the fable of the Sorcerer's Apprentice.[31] As AI systems that are separately developed and tested are composed and interact with other AI systems (within one's own services, forces, agencies, and between US systems and those of allies, adversaries, and potential adversaries), additional unintentional failures can occur.[32]

(2) Examples of Current Challenges

To make high-stakes decisions, and often in safety-critical contexts, the DoD and Intelligence Community (IC) must be able to depend on the integrity and security of the data that is used to train some kinds of ML systems. The challenges of doing so have been echoed by the leadership of the DoD and the IC,[33] including concerns with detecting adversarial attacks such as data poisoning.

(3) Recommendations for Adoption

Critical engineering practices needed to operationalize AI principles (such as 'traceable' and 'reliable'[34]) are described in the non-exhaustive list below. These practices span design, development, and deployment of AI systems.

1. **Concept of operations development and design and requirements definition and analysis**. Conduct systems analysis of operations, and identify mission success metrics and potential functions that can be performed by an AI technology. Assess general feasibility of specific candidate AI technologies, based on analyses of use cases and scenario development. This includes broad stakeholder engagement and hazard analysis with multi-disciplinary experts that ask key questions about potential disparate impact and document the process undertaken to ensure fairness and lack of unwanted bias in the ML application.[35] The feasibility of meeting these requirements may trigger a review of whether and where it is appropriate to use AI in the system being proposed.
    - **Risk assessment**. Trade-offs and risks, including a system's potential societal impact, should be discussed with a diverse, interdisciplinary group. Risk assessment questions should be asked about critical areas relevant to the national security context, including privacy and civil liberties, LOAC, human rights,[36] system security, and the risks of a new technology being leaked, stolen, or weaponized.[37]
2. **Documentation of the AI lifecycle:** Whether building and fielding an AI system or "infusing AI" into a preexisting system, require documentation in certain areas.[38] These include the data used in ML and origin of the data;[39] algorithm(s) used to build models, model characteristics, and intended uses of the AI capabilities; connections between and dependencies within systems,

and associated potential complications; the selected testing methodologies, performance indicators, and results for models used in the AI component; and required maintenance (including re-testing requirements) and technical refresh (including for when a system is used in a different scenario/setting or if the AI system is capable of online learning or adaptation).

3. **Infrastructure for traceability.** Invest resources and establish policies that support the traceability of AI systems. Traceability captures key information about the system development and deployment process for relevant personnel to adequately understand the technology.[40] Audits should support analyses of specific actions and characterizations of longer-term performance, and assure that performance on tests of the system and on real-world workloads meet requirements.
4. **Security and Robustness: Addressing Intentional and Unintentional Failures**
    - **Adversarial attacks, and use of robust ML methods**. Expand notions of adversarial attacks to include various "machine learning attacks,"[41] and seek latest technologies that demonstrate the ability to detect and notify operators of attacks, and also tolerate attacks.[42]
    - **Follow and incorporate advances in intentional and unintentional ML failures**. Given the rapid evolution of the field of study of intentional and unintentional ML failures, national security organizations must follow and adapt to the latest knowledge about failures and proven practices for monitoring, detection, and engineering and run-time protections. Related efforts and R&D focus on developing and deploying robust AI methods.[43]
    - **Adopt a security development lifecycle (SDL) for AI** systems focused on potential failure modes. This includes developing and regularly refining threat models to capture and characteristics of various attacks, establish a matrixed focus for developing and refining threat models, and ensuring SDL addresses ML development, deployment, and when ML systems are under attack.[44]
5. **Conduct red teaming** for both intentional and unintentional failure modalities. Bring together multiple perspectives to rigorously challenge AI systems, exploring the risks, limitations, and vulnerabilities in the context in which they'll be deployed (i.e., red teaming).
    - To mitigate intentional failure modes - Use methods to make systems more resistant to adversarial attacks, work with adversarial testing tools, and deploy teams dedicated to trying to brake systems and make them violate rules for appropriate behavior. [45]
    - To mitigate unintentional failure modes - test ML systems per a thorough list of realistic conditions they are expected to operate in. When selecting third-party components, consider the impact that a security vulnerability in them could have to the security of the larger system into which they are integrated. Have an accurate inventory of third-party components and a plan to respond when new vulnerabilities are discovered.[46]

- Organizations should consider establishing broader enterprise-wide communities of AI red teaming capabilities that could be applied to multiple AI developments (e.g., at a DoD service or IC element level, or higher).

(4) Recommendations for Future Action

- **Documentation strategy.** As noted in our First Quarter Recommendations, a common documentation strategy is needed to ensure sufficient documentation by all national security departments and agencies.[47] In the meantime, agencies should pilot documentation approaches across the AI lifecycle to help inform such a strategy.
- **Standards**. To improve traceability, future work is needed by standard setting bodies, alongside national security departments/agencies and the broader AI community, to develop audit trail requirements per mission needs for high-stakes AI systems including safety-critical applications.
- **Future R&D.** R&D is needed to advance capabilities for cultivating more robust methods that can overcome adverse conditions; to advance approaches that enable assessment of types and levels of vulnerability and immunity; and to enable systems to withstand or to degrade gracefully when targeted by a deliberate attack. R&D is also needed to advance capabilities to support risk assessment; to better understand the efficacy of interpretability tools and possible interfaces; and to develop benchmarks that assess the reliability of produced model explanations.

III. System Performance

(1) Overview

Fielding AI systems in a responsible manner includes establishing confidence that the technology will perform as intended. An AI system's performance must be assessed,[48] including assessing its capabilities and blind spots with data representative of real-world scenarios or with simulations of realistic contexts,[49] and its reliability, robustness (i.e., resilience in real-world settings—including adversarial attacks on AI components), and security during development and deployment.[50] System performance must also measure compliance with requirements derived from values such as fairness.

Testing protocols and requirements are essential for measuring and reporting on system performance. (Here, 'testing' broadly refers to what the DoD calls "Test, Evaluation, Verification, and Validation" (TEVV). This testing includes both what DOD refers to as Developmental Test and Evaluation and Operational Test and Evaluation.) AI systems present new challenges to established testing protocols and requirements as they increase in complexity, particularly for operational testing. However, existing methods like high-fidelity performance traces and means for sensing shifts, such as distributional shifts in targeted scenarios, allow for the

continuous monitoring of an AI system's performance.

When evaluating system performance, it is especially important to take into account holistic, end-to-end system behavior—the consequence of the interactions and relationships among system elements rather than the independent behavior of individual elements. While system engineering and national security communities have focused on system of systems engineering for years, specific attention must be paid to undesired interactions and emergent performance in AI systems. Multiple relatively independent AI systems can be viewed as distinct agents interacting in the environment of the system of systems, and some of these agents will be humans in and on the loop. Industry has encountered and documented problems in building 'systems of systems' out of multiple AI systems.[51] A related problem is encountered when the performance of one model in a pipeline changes, degrading the overall pipeline behavior.[52] As America's AI-intensive systems may increasingly be composed with allied AI-intensive systems, this becomes a topic for coordination with allies.

(2) Examples of Current Challenges

Unexpected interactions and errors commonly occur in integrated simulations and exercises, illustrating the challenges of predicting and managing behaviors of systems composed of multiple components. Intermittent failures can transpire after composing different systems; these failures are not the result of any one component having errors, but rather are due to the interactions of the composed systems.[53]

(3) Recommendations for Adoption

Critical practices to ensure optimal system performance are described in the following non-exhaustive list:

A. **Training and Testing procedures should cover key aspects of performance and appropriate performance metrics. These include:**
   1. **Standards for metrics and reporting needed to adequately achieve:**
      a. Consistency across testing and test reporting for critical areas.
      b. Testing for blindspots.[54]
      c. Testing for fairness. When testing for fairness, conduct sustained fairness assessments throughout development and deployment and document deliberations made on the appropriate fairness metrics to use. Agencies should conduct outcome and impact analysis to detect when subtle assumptions in the system show up as unexpected and undesired outcomes in the operational environment.[55]
      d. Articulation of performance standards and metrics. Clearly document system performance and communicate to the end user the meaning/significance of such performance metrics.
   2. **Representativeness of the data and model for the specific context at hand.** When using classification and prediction technologies, explicitly

consider and document challenges with representativeness of data used in analyses, and the fairness/accuracy of inferences and recommendations made with systems leveraging that data when applied in different populations/contexts.

3. **Evaluating an AI system's performance relative to current benchmarks** where possible. Benchmarks should assist in determining if an AI system's performance meets or exceeds current best performance.
4. **Evaluating aggregate performance of human-machine teams**. Consider that the current benchmark might be the current best performance of a human operator or the composed performance of the human-machine team. Where humans and machines interact, it is important to measure the aggregate performance of the team rather than the AI system alone.[56]
5. **Reliability and robustness:** Employ tools and techniques to carefully bound assumptions of robustness of the AI component in the larger system architecture. Provide sustained attention to characterizing the actual performance envelope (for nominal and off-nominal conditions) throughout development and deployment.[57]
6. **For systems of systems, testing machine-machine/multi-agent interaction**. Individual AI systems will be combined in various ways in an enterprise to accomplish broader missions beyond the scope of any single system, which can introduce its own problems.[58] As a priority during testing, challenge (or "stress test") interfaces and usage patterns with boundary conditions and assumptions about the operational environment and use.

B. **Maintenance and deployment**

Given the dynamic nature of AI systems, best practices for maintenance are also critically important. Recommended practices include:

1. **Specifying maintenance requirements** for datasets as well as for systems, given that their performance can degrade over time.[59]
2. **Continuously monitoring AI system performance**, including the use of high-fidelity traces to determine continuously if a system is going outside of acceptable parameters.[60]
3. **Iterative testing and validation**. Training and testing that provide characteristics on capabilities might not transfer or generalize to specific settings of usage; thus, testing and validation may need to be done recurrently, and at strategic intervention points, but especially for new deployments and classes of tasks.[61]
4. **Monitoring and mitigating emergent behavior**. There will be instances where systems are composed in ways not anticipated by the developers, thus, requiring monitoring the actual performance of the composed system and its components.

(4) Recommendations for Future Action

- **Future R&D.** R&D is needed to advance capabilities for TEVV of AI systems to better understand how to conduct TEVV and build checks and balances into an AI system. Improved methods are needed to explore,

predict, and control individual AI system behavior so that when AI systems are composed into systems-of-systems their interaction does not lead to unexpected negative outcomes.

- **Metrics**. Progress on a common understanding of TEVV concepts and requirements is critical for progress in widely used metrics for performance. Significant work is needed to establish what appropriate metrics should be to assess system performance across attributes for responsible AI and across profiles for particular applications/contexts.
- **International collaboration and cooperation.** Collaboration is needed to align on how to test and verify AI system reliability and performance, including along shared values (such as fairness and privacy). Such collaboration will be critical amongst allies and partners for interoperability and trust. Additionally, these efforts could potentially include dialogues between the U.S. and strategic competitors on establishing common standards of AI safety and reliability testing to reduce the chances of inadvertent escalation.

IV. Human-AI Interaction

(1) Overview

Responsible AI development and fielding requires striking the right balance of leveraging human and AI reasoning, recommendation, and decision-making processes. Ultimately, all AI systems will have some degree of human-AI interaction as they all will be developed to support humans.

(2) Examples of Current Challenges

There is an opportunity to develop AI systems to complement and augment human understanding, decision making, and capabilities. Decisions about developing and fielding AI systems for specific domains or scenarios should consider the relative strengths of AI capabilities and human intellect across expected distributions of tasks, considering AI system maturity or capability and how people and machine might coordinate.

Designs and methods for human-AI interaction can be employed to enhance human-AI teaming.[62] Methods in support of effective human-AI interaction can help AI systems understand when and how to engage humans for assistance, when AI systems should take initiative to assist human operators, and, more generally, how to support the creation of effective human-AI teams. In engaging with end users, it may be important for AI systems to infer and share with end users well-calibrated levels of confidence about their inferences, to provide human operators with an ability to weigh the importance of machine output or pause to consider details behind a recommendation more carefully. Methods, representations, and machinery can be employed to provide insight about AI inferences, including the use of interpretable machine learning.[63] Research directions include developing and fielding machinery

aimed at reasoning about human strengths and weaknesses, such as recognizing and responding to the potential for costly human biases of judgment and decision making in specific settings.[64] Other work centers on mechanisms to consider the ideal mix of initiatives, including when and how to rely on human expertise versus on AI inferences.[65] As part of effective teaming, AI systems can be endowed with the ability to detect the focus of attention, workload, and interruptability of human operators and consider these inferences in decisions about when and how to engage with operators.[66] Directions of effort include developing mechanisms for identifying the most relevant information or inferences to provide end users of different skills in different settings.[67] Consideration must be given to the prospect introducing bias, including potential biases that may arise because of the configuration and sequencing of rendered data. For example, IC research[68] shows that confirmation bias can be triggered by the order in which information is displayed, and this order can consequently impact or sway intel analyst decisions. Careful design and study can help to identify and mitigate such bias.

(3) Recommendations for Adoption

Critical practices to ensure optimal human-AI interaction are described in the non-exhaustive list below. These recommended practices span the entire AI lifecycle.

A. **Identification of functions of human in design, engineering, and fielding of AI**
   1. **Define functions, tasks, and responsibilities of human operators and assign them to specific individuals.** Functions will vary for each domain and project, and should be periodically revisited.
   2. **Policies should define the tasks of humans across the AI lifecycle**, given the nature of the mission and current competencies of AI.
   3. **Enable feedback and oversight to ensure that systems operate as they should**.

B. **Explicit support of human-AI interaction and collaboration**
   1. **Human-AI design guidelines**. AI systems designs should take into account the defined tasks of humans in human-AI collaborations in different scenarios; ensure the mix of human-machine actions in the aggregate is consistent with the intended behavior, and accounts for the ways that human and machine behavior can co-evolve;[69] and also avoid automation bias and unjustified reliance on humans in the loop as failsafe mechanisms. Practices should allow for auditing of the human-AI pair. And designs should be transparent to allow for an understanding of how the AI is working day-to-day, supported by an audit trail if things go wrong. Based on context and mission need, designs should ensure usability of AI systems by AI experts, domain experts, and novices, as appropriate.
   2. **Algorithms and functions in support of interpretability and explanation.** Algorithms and functions that provide individuals with task-relevant knowledge and understanding should take into account that key factors in an AI system's inferences and actions can be understood differently

by various audiences (e.g., real-time operators, engineers and data scientists, and oversight officials). Interpretability and explainability exists in degrees. In this regard, interpretability intersects with traceability, audit, and documentation practices.

3. **Designs that provide cues to the human operator(s) about the level of confidence the system has in the results or behaviors of the system**. AI system designs should appropriately convey uncertainty and error bounding. For instance, a user interface should convey system self-assessment of confidence alerts when the operational environment is significantly different from the environment the system was trained for, and indicate internal inconsistencies that call for caution.
4. **Policies for machine-human initiative and handoff.** Policies, and aspects of human computer interaction, system interface, and operational design, should define when and how information or tasks should be passed from a machine to a human operator and vice versa.
5. **Leveraging traceability to assist with system development and understanding**. Traceability processes must include audit logs or other traceability mechanisms to retroactively understand if something went wrong, and why, in order to improve systems and their use and for redress. Infrastructure and instrumentation[70] can also help assess humans, systems, and environments to gauge the impact of AI at all levels of system maturity; and to measure the effectiveness and performance for hybrid human-AI systems in a mission context.
6. **Training**. Train and educate individuals responsible for AI development and fielding, including human operators, decision makers, and procurement officers.[71]

(4) Recommendations for Future Action

- **Future R&D.** R&D is needed to advance capabilities of AI technologies to perceive and understand the meaning of human communication including spoken speech, written text, and gestures. This research should account for varying languages and cultures, with special attention to diversity given that AI typically performs worse in cases in gender and racial minorities. It is also needed to improve human-machine teaming, including disciplines and technologies centered on decision sciences, control theory, psychology, economics (human aspects and incentives), and human factors engineering. R&D for human-machine teaming should also focus on helping systems understand human blind spots and biases, and optimizing factors such as human attention, human workload, ideal mixing of human and machine initiatives, and passing control between the human and machine.
- **Training**. Ongoing work is needed to train the workforce that will interact with, collaborate with, and be supported by AI systems. In its First Quarter Recommendations, the Commission provided recommendations for such training. Operators should receive training on the specifics of the system and application, the fundamentals of AI and data science, and refresher trainings

(e.g., when systems are deployed in new settings and unfamiliar scenarios, and when predictive models are revised with new data as performance may shift with updates and introduce behaviors unfamiliar to operators).

V. Accountability and Governance

(1) Overview

National security departments and agencies must specify who will be held accountable for both specific system outcomes and general system maintenance and auditing, in what way, and for what purpose. Government must address the difficulties in preserving human accountability, including for end users, developers, testers, and the organizations employing AI systems. End users and those affected by the actions of an AI system should be offered the opportunity to appeal an AI system's determinations. And accountability and appellate processes must exist not only for AI decisions, but also for AI system inferences, recommendations, and actions.

(2) Examples of Current Challenges

If a contentious outcome occurs, overseeing entities need the technological capacity to understand what in the AI system caused this. For example, if a soldier uses an AI-enabled weapon and the result violates international law of war standards, an investigating body or military tribunal should be able to re-create what happened through auditing trails and other documentation. Without policies requiring such technology and the enforcement of those policies, proper accountability would be elusive, if not impossible. Moreover, auditing trails and documentation will prove critical as courts begin to grapple with whether AI system determinations reach the requisite standards to be admitted as evidence. Building the traceability infrastructure to permit auditing (as described in *Engineering Practices*) will increase the costs of building AI systems and take significant work -- a necessary investment given our commitment to accountability, discoverability, and legal compliance.

(3) Recommendations for Adoption

Critical accountability and governance practices are identified in the non-exhaustive list below.

1. **Identify responsible actors.** Determine and document the human beings accountable for a specific AI system or any given part of the system and the processes involved. This includes identifying who is responsible for the operation of the system (including its inferences, recommendations, and actions during usage) and who is responsible for enforcing system use policies. Determine and document the mechanism/structure for holding such actors accountable and to whom it should be disclosed for proper oversight.
2. **Adopt technology to strengthen accountability processes and goals**. Document the chains of custody and command involved in developing and fielding AI systems to know who was responsible at which

point in time. Improving traceability and auditability capabilities will allow agencies to better track a system's performance and outcomes.[72]

3. **Adopt policies to strengthen accountability.** Identify or, if lacking, establish policies that allow individuals to raise concerns about irresponsible AI development/use, e.g. via an ombudsman. Agencies should institute specific oversight and enforcement practices, including: auditing and reporting requirements; a mechanism that would allow thorough review of the most sensitive/high-risk AI systems to ensure auditability and compliance with responsible use and fielding requirements; an appealable process for those found at fault of developing or using AI irresponsibly; and grievance processes for those affected by the actions of AI systems. Agencies should leverage best practices from academia and industry for conducting internal audits and assessments,[73] while also acknowledging the benefits offered by external audits.[74]
4. **External oversight support**. Self-assessment alone may prove to be inadequate in all scenarios. Supporting traceability, specifically documentation to audit trails, will allow for external oversight.[75] Congress can provide a key oversight function throughout the AI lifecycle, asking critical questions of agency leaders and those responsible for AI systems.[76]

(4) Recommendations for Future Action

Currently no external oversight mechanism exists specific to AI in national security. Notwithstanding the important work of Inspectors General in conducting internal oversight, open questions remain as to how to complement current practices and structures.

Endnotes to Appendix A-1

1. Examples of efforts to establish ethics guidelines are found within the U.S. government, industry, and internationally. See, e.g., *Draft Memorandum for the Heads of Executive Departments and Agencies: Guidance for Regulation of Artificial Intelligence Applications*, Office of Management and Budget (Jan. 1, 2019), https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf; Jessica Fjeld & Adam Nagy, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center (Jan. 15, 2020), https://cyber.harvard.edu/publication/2020/principled-ai; *OECD Principles on AI*, OECD (last visited June 17, 2020), https://www.oecd.org/going-digital/ai/principles/; *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence, European Union at 26-31 (Apr. 8, 2019), https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.
2. C. Todd Lopez, *DOD Adopts 5 Principles of Artificial Intelligence Ethics*, Department of Defense (Feb. 5, 2020), https://www.defense.gov/Explore/News/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/ [hereinafter Lopez, DoD Adopts 5 Principles].
3. Ben Huebner, *Presentation: AI Principles*, Intelligence and National Security Alliance 2020 Spring Symposium, Building an AI Powered IC (Mar. 4, 2020), https://www.insaonline.org/2020-spring-symposium-building-an-ai-powered-ic-event-recap/.
4. See, e.g., U.S. Const. amendments I, IV, V, and XIV; Americans with Disability Act of 1990, 42 U.S.C. § 12101 et seq.; Title VII of the Consumer Credit Protection Act, 15 U.S.C. §§ 1691-1691f; Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq..
5. International Covenant on Civil and Political Rights, UN General Assembly, United Nations, Treaty Series, vol. 999, at 171 (December 16, 1966), https://www.refworld.org/docid/3ae6b3aa0.html. As noted in the Commission's Interim Report, America and its like-minded partners share a commitment to democracy, human dignity and human rights. *Interim Report*, NSCAI (Nov. 2019), https://www.nscai.gov/reports. Many, but not all nations, share commitments to these values. Even when values are shared, however, they can be culturally relative, for instance, across nations, owing to interpretative nuances.
6. See, e.g., Daniel Coats, *Intelligence Community Directive 107*, ODNI (Feb. 28, 2018), https://fas.org/irp/dni/icd/icd-107.pdf (on protecting civil liberties and privacy); *IC Framework for Protecting Civil Liberties and Privacy and Enhancing Transparency Section 702*, Intel.gov (Jan. 2020), https://www.intelligence.gov/index.php/ic-on-the-record/guide-to-posted-documents#SECTION_702-OVERVIEW (on privacy and civil liberties implication assessments and oversight); *Principles of Professional Ethics for the Intelligence Community*, ODNI (last accessed June 17, 2020), (https://www.dni.gov/index.php/who-we-are/organizations/clpt/clpt-related-menus/clpt-related-links/ic-principles-of-professional-ethics (on diversity and inclusion).
7. See, e.g., *Privacy Office*, Department of Homeland Security (last accessed June 3, 2020), https://www.dhs.gov/privacy-office#; *CRCL Compliance Branch*, Department of Homeland Security (last accessed May 15, 2020), https://www.dhs.gov/compliance-branch.

8. See Samuel Jenkins & Alexander Joel, *Balancing Privacy and Security: The Role of Privacy and Civil Liberties in the Information Sharing Environment*, IAPP Conference 2010 (2010), https://dpcld.defense.gov/Portals/49/Documents/Civil/IAPP.pdf.
9. See *Projects*, U.S. Privacy and Civil Liberties Oversight Board (last accessed June 17, 2020), https://www.pclob.gov/Projects.
10. See *Department of Defense Law of War Manual*, DoD Office of General Counsel (Dec. 2016), https://dod.defense.gov/Portals/1/Documents/pubs/DoD%20Law%20of%20War%20Manual%20-%20June%202015%20Updated%20Dec%202016.pdf?ver=2016-12-13-172036-190 [hereinafter DoD Law of War Manual]; see also *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense: Supporting Document*, Defense Innovation Board (Oct. 31, 2019), https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF ("More than 10,000 military and civilian lawyers within DoD advise on legal compliance with regard to the entire range of DoD activities, including the Law of War. Military lawyers train DoD personnel on Law of War requirements, for example, by providing additional Law of War instruction prior to a deployment of forces abroad. Lawyers for a Component DoD organization advise on the issuance of plans, policies, regulations, and procedures to ensure consistency with Law of War requirements. Lawyers review the acquisition or procurement of weapons. Lawyers help administer programs to report alleged violations of the Law of War through the chain of command and also advise on investigations into alleged incidents and on accountability actions, such as commanders' decisions to take action under the Uniform Code of Military Justice. Lawyers also advise commanders on Law of War issues during military operations.").
11. Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, United Nations General Assembly (Dec. 10, 1984), https://www.ohchr.org/en/professionalinterest/pages/cat.aspx.
12. See DoD Law of War Manual at 26 ("Rules of Engagement reflect legal, policy, and operational considerations, and are consistent with the international law obligations of the United States, including the law of war.").
13. See *Department of Defense Directive 3000.09 on Autonomy in Weapons Systems*, Department of Defense (Nov. 21, 2012), https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf ("Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.").
14. See, e.g., *Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System*, Partnership on AI, https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/; Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, Reuters (Oct. 9, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [hereinafter Dastin, Amazon Scraps Secret AI Recruiting Tool]; Andi Peng et al., *What You See Is What You Get? The Impact of Representation Criteria on Human Bias in Hiring*, Proceedings of the 7th AAAI Conference on Human Computation and Crowdsourcing (Oct. 2019), https://arxiv.org/pdf/1909.03567.pdf; Patrick Grother, et. al., *Face Recognition Vendor Test (FRVT) Part Three: Demographic Effects*, National Institute of Standards and Technology (Dec. 2019), https://doi.org/10.6028/NIST.IR.8280.
15. PNDC provides predictive analytics to improve military readiness; enable earlier identification of service members with potential unfitting, disabling, or career-ending

conditions; and offer opportunities for early medical intervention or referral into disability processing. To do so, PNDC provides recommendations at multiple points in the journey of the non-deployable service member through the Military Health System to make "better decisions" that improve medical outcomes and delivery of health services. This is very similar to the OPTUM decision support system that recommended which patients should get additional intervention to reduce costs. Analysis showed millions of US patients were processed by the system, with substantial disparate impact on black patients compared to white patients. Shaping development from the start to reflect bias issues (which can be subtle) would have produced a more equitable system and avoided scrutiny and suspension of system use when findings were disclosed. Heidi Ledford, *Millions of Black People Affected by Racial Bias in Health Care Algorithms*, Nature (October 26, 2019), https://www.nature.com/articles/d41586-019-03228-6.

16. See e.g., Dastin, Amazon Scraps Secret AI Recruiting Tool.
17. Mohsen Bayati, et. al., *Data-Driven Decisions for Reducing Readmissions for Heart Failure: General Methodology and Case Study*, PLOS One Medicine (Oct. 2014), https://doi.org/10.1371/journal.pone.0109264; Eric Horvitz & Adam Seiver, *Time-Critical Action: Representations and Application*, Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (Aug. 1997), https://arxiv.org/pdf/1302.1548.pdf.
18. The Commission is doing a fulsome assessment of where investment needs to be made; this document notes important R&D areas through the lens of ethics and responsible AI.
19. See Inioluwa Deborah Raji, et al., *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing,* ACM FAT (Jan. 3, 2020), https://arxiv.org/abs/2001.00973 [hereinafter, Raji, Closing the AI Accountability Gap].
20. See Lopez, DoD Adopts 5 Principles.
21. *Model Interpretability in Azure Machine Learning*, Microsoft (July 9, 2020), https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability.
22. Lopez, DoD Adopts 5 Principles.
23. Jessica Cussins Newman, *Decision Points in AI Governance: Three Case Studies Explore Efforts to Operationalize AI Principles* (May 5, 2020), Berkeley Center for Long-Term Cybersecurity, https://cltc.berkeley.edu/ai-decision-points/; Raji, *Closing the AI Accountability Gap*; Miles Brundage, et al., *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims* (Apr. 20, 2020), https://arxiv.org/abs/2004.07213 [hereinafter Brundage, Toward Trustworthy AI Development]; Saleema Amershi, et. al., *Software Engineering for Machine Learning: A Case Study*, Microsoft (Mar. 2019), https://www.microsoft.com/en-us/research/uploads/prod/2019/03/amershi-icse-2019_Software_Engineering_for_Machine_Learning.pdf.
24. Dario Amodei, et al. *Concrete problems in AI safety* (July 2016), https://arxiv.org/abs/1606.06565.
25. Guofu Li, et al., *Security Matters: A Survey on Adversarial Machine Learning*, (Oct. 2018), https://arxiv.org/abs/1810.07339; Elham Tabassi et al., *NISTIR 8269: A Taxonomy and Terminology of Adversarial Machine Learning (Draft)*, National Institute of Standards and Technology (Oct. 2019), https://csrc.nist.gov/publications/detail/nistir/8269/draft.
26. José Faria, *Non-Determinism and Failure Modes in Machine Learning*, 2017 IEEE 28th International Symposium on Software Reliability Engineering Workshops (Oct. 2017), https://ieeexplore.ieee.org/document/8109300.

27. Ram Shankar Siva Kumar et al. *Failure Modes in Machine Learning* (Nov. 2019), https://docs.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning [hereinafter Kumar, Failure Modes in Machine Learning"].
28. Id.
29. For 11 categories of attack, and associated overviews, see the Intentionally-Motivated Failures Summary in Kumar, Failure Modes in Machine Learning.
30. Unexpected performance represents emergent runtime output, behavior, or effects at the system level, e.g., through unanticipated feature interaction, … that was also not previously observed during model validation." See Colin Smith, et al., *Hazard Contribution Modes of Machine Learning Components*, AAAI-20 Workshop on Artificial Intelligence Safety (SafeAI 2020) (Feb. 7, 2020), https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20200001851.pdf.
31. Thomas Dietterich & Eric Horvitz, *Rise of Concerns about AI: Reflections and Directions*, Communications of the ACM, Vol. 58 No. 10, at 38-40 (Oct. 2015), http://erichorvitz.com/CACM_Oct_2015-VP.pdf.
32. Kumar, Failure Modes in Machine Learning.
33. For concerns about generative adversarial networks (GANS) voiced by Gen. Shanahan, JAIC, see Don Rassler, *A View from the CT Foxhole Lieutenant General John N.T. "Jack" Shanahan, Director, Joint Artificial Intelligence Center, Department of Defense*, Combating Terrorism Center at West Point (Dec. 2019) https://ctc.usma.edu/view-ct-foxhole-lieutenant-general-john-n-t-jack-shanahan-director-joint-artificial-intelligence-center-department-defense/. Concerns about GANS, information authenticity, and reliable and understandable systems were voiced by Dean Souleles, IC. See *Afternoon Keynote*, Intelligence and National Security Alliance 2020 Spring Symposium: Building an AI Powered IC (Mar. 4, 2020), https://www.insaonline.org/2020-spring-symposium-building-an-ai-powered-ic-event-recap/.
34. See Lopez, DOD Adopts 5 Principles.
35. There is no single definition of fairness. System developers and organizations fielding applications must work with stakeholders to define fairness, and provide transparency via disclosure of assumed definitions of fairness. Definitions or assumptions about fairness and metrics for identifying fair inferences and allocations should be explicitly documented. This should be accompanied by a discussion of alternate definitions and rationales for the current choice. These elements should be documented internally as machine-learning components and larger systems are developed. This is especially important as establishing alignment on the metrics to use for assessing fairness encounters an added challenge when different cultural and policy norms are involved when collaborating on development and use with allies.
36. For more on the importance of human rights impact assessments of AI systems, see *Report of the Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression,* UN Human Rights Office of the High Commissioner (2018), https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx. For an example of a human rights risk assessment for AI in categories such as nondiscrimination and equality, political participation, privacy, and freedom of expression, see Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, Data Society (Oct. 2018), https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.
37. For exemplary risk assessment questions that IARPA has used, see Richard Danzig, *Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority*, Center for a New American Security at 22 (June 28, 2018),

https://s3.amazonaws.com/files.cnas.org/documents/CNASReport-Technology-Roulette-DoSproof2v2.pdf?mtime=20180628072101.

38. Documentation recommendations build off of a legacy of robust documentation requirements. See *Department of Defense Standard Practice: Documentation of Verification, Validation, and Accreditation (VV&A) For Models and Simulations*, Department of Defense (Jan. 28, 2008), https://acqnotes.com/Attachments/MIL-STD-3022%20Documentation%20of%20VV&A%20for%20Modeling%20&%20Simulation%2028%20Jan%2008.pdf.
39. For an industry example, see Timnit Gebru, et al., *Datasheets for Datasets*, Microsoft (March 2018), https://www.microsoft.com/en-us/research/publication/datasheets-for-datasets/. For more on data, model and system documentation, see *Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles (ABOUT ML)*, an evolving body of work from the Partnership on AI about documentation practices at https://www.partnershiponai.org/about-ml/. Documenting caveats of re-use for both datasets and models is critical to avoid "off-label" use harms as one senior official notes. David Thornton, *Intelligence Community Laying Foundation for AI Data Analysis*, Federal News Network (Nov. 1, 2019), https://federalnewsnetwork.com/all-news/2019/11/intelligence-community-laying-the-foundation-for-ai-data-analysis/.
40. Jonathan Mace, et al., *Pivot Tracing: Dynamic Causal Monitoring for Distributed Systems*, Communications of the ACM, Vol. 63 No. 3, at 94-102 (March 2020), https://dl.acm.org/doi/10.1145/2815400.2815415 [hereinafter Mace, Pivot Tracing].
41. Aleksander Madry, et al., *Towards Deep Learning Models Resistant to Adversarial Attacks*, MIT (Sept 4, 2019), https://arxiv.org/abs/1706.06083 [hereinafter Madry, Towards Deep Learning Models Resistant to Adversarial Attacks].
42. See e.g., id.; Thomas Dietterich, *Steps Toward Robust Artificial Intelligence*, AI Magazine (2017), https://www.aaai.org/ojs/index.php/aimagazine/article/view/2756/2644; Eric Horvitz, *Reflections on Safety and Artificial Intelligence*, Safe AI: Exploratory Technical Workshop on Safety and Control for AI, White House OSTP and Carnegie Mellon University (June 27, 2016), http://erichorvitz.com/OSTP-CMU_AI_Safety_framing_talk.pdf.
43. On adversarial attacks on ML, see Kevin Eykholt, et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, IEEE Conference on Computer Vision and Pattern Recognition at 1625–1634 (2018), https://ieeexplore.ieee.org/document/8578273; On directions with robustness, see Madry, Towards deep learning models resistant to adversarial attacks. For a more exhaustive list of sources see the Commission's extended version of the Key Considerations in Appendix A-2.
44. Ram Shankar Siva Kumar, et al., *Adversarial Machine Learning--Industry Perspectives*, 2020 IEEE Symposium on Security and Privacy (SP) Deep Learning and Security Workshop (Feb. 2020), https://arxiv.org/pdf/2002.05646.pdf.
45. Dou Goodman, et al., *Advbox: A Toolbox to Generate Adversarial Examples that Fool Neural Networks* (2020), https://arxiv.org/abs/2001.05574.
46. See *What are the Microsoft SDL Practices?*, Microsoft (last accessed July 14, 2020), https://www.microsoft.com/en-us/securityengineering/sdl/practices .
47. See *First Quarter Recommendations*, NSCAI (Mar. 2020), https://www.nscai.gov/reports. Ongoing efforts to share best practices for documentation among government agencies through GSA's AI Community of Practice further indicate the ongoing need and desire for common guidance.
48. Ben Shneiderman, *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy*, International Journal of Human–Computer Interaction 2020, Vol. 36, No. 6, at 495–

504 (Mar. 23, 2020), https://doi.org/10.1080/10447318.2020.1741118 [hereinafter, Shneiderman, Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy].

49. However, test protocols must acknowledge test sets may not be fully representative of real-world usage.
50. Brundage, Toward Trustworthy AI Development; Ece Kamar, et al., *Combining Human and Machine Intelligence in Large-Scale Crowdsourcing*, Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (June 2012), https://dl.acm.org/doi/10.5555/2343576.2343643 [hereinafter Kamar, Combining Human and Machine Intelligence in Large-Scale Crowdsourcing].
51. One example is "Hidden Feedback Loops", where systems that learn from external world behavior may also shape the behavior they are monitoring. See D. Sculley, et al., *Machine Learning: The High Interest Credit Card of Technical Debt*, Google (2014), https://research.google/pubs/pub43146/.
52. Megha Srivastava, et al., *An Empirical Analysis of Backward Compatibility in Machine Learning Systems*, KDD'20 (forthcoming, Aug. 2020) [hereinafter Srivastava, An Empirical Analysis of Backward Compatibility in Machine Learning Systems].
53. David Sculley, et al., *Hidden Technical Debt in Machine Learning Systems*, Proceedings of the 28th International Conference on Neural Information Processing Systems (Dec. 2015), https://dl.acm.org/doi/10.5555/2969442.2969519.
54. Ramya Ramakrishnan, et al., *Blind Spot Detection for Safe Sim-to-Real Transfer*, Journal of Artificial Intelligence Research 67 at 191-234 (2020) https://www.jair.org/index.php/jair/article/view/11436.
55. See Microsoft's AI Fairness checklist as an example of an industry tool to support fairness assessments, Michael A. Madaio, et al., *Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI*, CHI 2020 (Apr. 25-30, 2020), http://www.jennwv.com/papers/checklists.pdf [hereinafter Madaio, Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI].
56. Kamar, Combining Human and Machine Intelligence in Large-scale Crowdsourcing.
57. See Shneiderman, Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy.
58. Cynthia Dwork, et al., *Individual Fairness in Pipelines*, https://arxiv.org/abs/2004.05167; Srivastava, An Empirical Analysis of Backward Compatibility in Machine Learning Systems.
59. *Artificial Intelligence (AI) Playbook for the U.S. Federal Government*, Artificial Intelligence Working Group, ACT-IAC Emerging Technology Community of Interest (Jan. 22, 2020), https://www.actiac.org/act-iac-white-paper-artificial-intelligence-playbook.
60. Ori Cohen, *Monitor! Stop Being A Blind Data-Scientist* (Oct. 8, 2019), https://towardsdatascience.com/monitor-stop-being-a-blind-data-scientist-ac915286075f; Mace, Pivot Tracing at 94-102.
61. Eric Breck, et al., *The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction*, 2017 IEEE International Conference on Big Data, (Dec. 11-14, 2017), https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8258038&tag=1.
62. Saleema Amershi, et al., *Guidelines for Human-AI Interaction*, Proceedings of the CHI Conference on Human Factors in Computing Systems (2019) https://dl.acm.org/doi/10.1145/3290605.3300233.
63. Rich Caruana, et al., *Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission,* Semantic Scholar (Aug. 2015),

https://www.semanticscholar.org/paper/Intelligible-Models-for-HealthCare%3A-Predicting-Risk-Caruana-Lou/cb030975a3dbcdf52a01cbd1c140711332313e13.

64. Eric Horvitz, *Reflections on Challenges and Promises of Mixed-Initiative Interaction*, AAAI Magazine 28 Special Issue on Mixed-Initiative Assistants (2007), http://erichorvitz.com/mixed_initiative_reflections.pdf.
65. Eric Horvitz, *Principles of Mixed-Initiative User Interfaces*, Proceedings of CHI '99 ACM SIGCHI Conference on Human Factors in Computing Systems (May 1999), https://dl.acm.org/doi/10.1145/302979.303030; Kamar, Combining Human and Machine Intelligence in Large-scale Crowdsourcing.
66. Eric Horvitz, et al., *Models of Attention in Computing and Communications: From Principles to Applications*, Communications of the ACM 46(3) at 52-59 (Mar. 2003), https://cacm.acm.org/magazines/2003/3/6879-models-of-attention-in-computing-and-communication/fulltext.
67. Eric Horvitz & Matthew Barry, *Display of Information for Time-Critical Decision Making*, Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (Aug. 1995), https://arxiv.org/pdf/1302.4959.pdf.
68. There has been considerable research in the IC on the challenges of confirmation bias for analysts. Some experiments demonstrated a strong effect that the sequence in which information is presented alone can shape analyst interpretations and hypotheses. Brant Cheikes, et al., *Confirmation Bias in Complex Analyses*, MITRE (Oct. 2004), https://www.mitre.org/sites/default/files/pdf/04_0985.pdf. This highlights the care that is required when designing the human machine teaming when complex, critical, and potentially ambiguous information is presented to analysts and decision makers.
69. Shneiderman, Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy at 495–504.
70. Infrastructure includes tools (hardware and software) in the test environment that support monitoring system performance (such as the timing of exchanges among systems, or the ability to generate test data). Instrumentation refers to the presence of monitoring and additional interfaces to provide insight into a specific system under test.
71. Gagan Bansal, et al., *Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff*, AAAI (Jul. 2019), https://www.aaai.org/ojs/index.php/AAAI/article/view/4087.
72. See Raji, Closing the AI Accountability Gap.
73. See id. ("In this paper, we present internal algorithmic audits as a mechanism to check that the engineering processes involved in AI system creation and deployment meet declared ethical expectations and standards, such as organizational AI principles"); see also Madaio, Co-Designing Checklists to Understand Organizational Challenges and Opportunities Around Fairness in AI.
74. For more on the benefits of external audits, see Brundage, Toward Trustworthy AI Development. For an agency example, see Aaron Boyd, *CBP Is Upgrading to a New Facial Recognition Algorithm in March*, Nextgov.com (Feb. 7, 2020), https://www.nextgov.com/emerging-tech/2020/02/cbp-upgrading-new-facial-recognition-algorithm-march/162959/ (highlighting a NIST algorithmic assessment on behalf of U.S. Customs and Border Protection).
75. Raji, Closing the AI Accountability Gap.
76. Maranke Wieringa, *What to Account for When Accounting for Algorithms*, Proceedings of the 2020 ACM FAT Conference, (Jan. 2020), https://doi.org/10.1145/3351095.3372833.

Appendix A-2 — Key Considerations for Responsible Development & Fielding of AI (Extended Version)

Outline:

Introduction

I. Aligning Systems and Uses with American Values and the Rule Of Law

(1) Overview
(2) Examples of Current Challenges
(3) Recommendations for Adoption
    A. Developing uses and building systems that behave in accordance with American values and the rule of law
        1. Employing technologies and operational policies aligning with privacy preservation, fairness, inclusion, human rights, and law of armed conflict.
    B. Representing objectives and trade-offs
        1. Consider and document value considerations in AI systems and components based on specifying how trade-offs with accuracy are handled.
        2. Consider and document value considerations in AI systems that rely on representations of objective or utility functions.
        3. Conduct documentation, reviews, and set limits on disallowed outcomes.
(4) Recommendations for Future Action

II. Engineering Practices

(1) Overview
(2) Examples of Current Challenges
(3) Recommendations for Adoption
        1. Concept of operations development, and design and requirements definition and analysis
        2. Documentation of the AI lifecycle
        3. Infrastructure to support traceability, including auditability and forensics
        4. Security and robustness: addressing intentional and unintentional failures
        5. Conduct red teaming
(4) Recommendations for Future Action

III. System Performance

(1) Overview
(2) Examples of Current Challenges
(3) Recommendations for Adoption
    A. Training and testing
    Performance and performance metrics
        1. Standards for metrics & reporting
            a. Consistency across testing/test reporting
            b. Testing for blind spots
            c. Testing for fairness
            d. Articulation of performance standards and metrics
        2. Representativeness of data and model for the specific context at hand
        3. Evaluating an AI system's performance relative to current benchmarks
        4. Evaluating aggregate performance of human-machine teams
        5. Reliability and robustness
        6. For systems of systems, testing machine-machine/multi-agent interaction
    B. Maintenance and deployment
        1. Specifying maintenance requirements
        2. Continuously monitoring and evaluating AI system performance
        3. Iterative and sustained testing and validation
        4. Monitoring and mitigating emergent behavior
(4) Recommendations for Future Action

IV. Human-AI Interaction

(1) Overview
(2) Examples of Current Challenges
(3) Recommendations for Adoption
    A. Identification of functions of humans in design, engineering, and fielding of AI
        1. Define functions and responsibilities of human operators and assign them to specific individuals.
        2. Policies should define the tasks of humans across the AI lifecycle
        3. Enable feedback and oversight to ensure that systems operate as they should.
    B. Explicit support of human-AI interaction and collaboration
        1. Human-AI design guidelines
        2. Algorithms and functions in support of interpretability and explanation.

Introduction

In the Commission's Interim Report, we stated that "defense and national security agencies must develop and deploy AI in a responsible, trusted, and ethical manner to sustain public support, maximize operational effectiveness, maintain the integrity of the profession of arms, and strengthen international alliances."[272]

As the Commission makes recommendations to advance ethical and responsible AI for national security, we are aware that this topic presents unique challenges. Concerns about the responsible development and fielding of AI technologies span a range of issues. Many debates are ongoing as the technology and its applications rapidly evolve, and the need for norms and best practices becomes more apparent.

The Commission acknowledges the efforts undertaken to date to establish ethics guidelines for AI by entities in government, in the private sector, and around the world.[273] The Department of Defense took the critical step of adopting a set of high-

---

[272] *Interim Report*, NSCAI at 16 (Nov. 2019), https://www.nscai.gov/reports [hereinafter Interim Report].

[273] Examples of efforts to establish ethics guidelines are found within the U.S. government, industry, and internationally. See e.g., *Draft Memorandum for the Heads of Executive Departments and Agencies: Guidance for Regulation of Artificial Intelligence Applications*, Office of Management and Budget (Jan. 1, 2019), https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-

level principles to guide its development and use of AI.[274] While some agencies critical to national security have adopted, or are in the process of adopting, AI principles,[275] others agencies have not provided such guidance. In cases where principles are offered, it can be difficult to translate the high-level concepts into concrete actions. There is often a gap between articulating high-level goals around responsible AI and operationalizing them.

In addition, agencies would benefit from the establishment of greater consistency in policies to further the responsible development and fielding of AI technologies across government. A unified approach would not only be more efficient, but it could also stimulate innovation and efficiencies through the sharing of models, data, and other information. Below the Commission is identifying a set of challenges and making recommendations on directions with responsibly developing and fielding AI systems, and for pinpointing the concrete actions that should be adopted across the government to help overcome these challenges.

This Commission has assessed a set of recommended practices in five categorical areas that are ripe for adoption. Collectively, they form a paradigm for aligning AI system development and AI system behavior to goals and values. The first section provides guidance specific to implementing systems that abide by American values and the rule of law. The section covers aligning the run-time behavior of systems to the related, more technical encodings of objectives, utilities, and trade-offs. The four following sections (on *Engineering Practices*, *System Performance*, *Human-AI Interaction*, and *Accountability & Governance*) serve in support of core American values and outline practices needed to develop and field systems that are trustworthy, understandable, reliable, and robust. Recommended practices span multiple phases of the *AI lifecycle*, from conception and early design, through development and testing, and maintenance and technical refresh. The Commission uses "development" to refer to 'designing, building, and testing during development and prior to deployment' and "fielding" to refer to 'deployment, monitoring, and sustainment.'

Though best practices will evolve (for instance, through future R&D), these recommended practices establish a baseline for the responsible development and fielding of AI technologies. They provide a floor, rather than a ceiling, for the

---

AI-1-7-19.pdf; Jessica Fjeld & Adam Nagy, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center (Jan. 15, 2020), https://cyber.harvard.edu/publication/2020/principled-ai; *OECD Principles on AI*, OECD (last accessed June 17, 2020), https://www.oecd.org/going-digital/ai/principles/; *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence, European Union at 26-31 (Apr. 8, 2019), https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.

[274] C. Todd Lopez, *DOD Adopts 5 Principles of Artificial Intelligence Ethics*, Department of Defense (Feb. 5, 2020), https://www.defense.gov/Explore/News/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/ [hereinafter, Lopez, DOD Adopts 5 Principles].

[275] See Ben Huebner, *Presentation: AI Principles*, Intelligence and National Security Alliance 2020 Spring Symposium, Building an AI Powered IC, (Mar. 4, 2020), https://www.insaonline.org/2020-spring-symposium-building-an-ai-powered-ic-event-recap/.

responsible development and fielding of AI technologies. The Commission recommends that heads of departments and agencies implement the Key Considerations as a paradigm for the responsible development and fielding of AI systems. This includes developing processes and programs aimed at adopting the paradigm's recommended practices, monitoring their implementation, and continually refining them as best practices evolve. These practices imply derived requirements for AI systems, requirements that in turn become an integral part of an agency's risk management process when deciding whether and how to develop and use AI for the context at hand. These recommended practices should apply both to systems that are developed by departments and agencies, as well as those that are acquired. Systems acquired (whether commercial, off-the-shelf systems or those acquired through contractors) should be subjected to the same rigorous standards and practices—whether in the acquisitions or acceptance processes. As such, the government organization overseeing the bidding process should require assertions of goals aligned with recommended practices for the Key Considerations in the process.

In each of the five categorical areas that follow, we first provide a conceptual overview of the scope and importance of the topic. We then illustrate an example of a current challenge relevant to national security departments that underscores the need to adopt recommended practices in this area. Then, we provide a list of recommended practices that agencies should adopt, acknowledging research, industry tools, and exemplary models within government that could support agencies in the adoption of recommended practices. Finally, in areas where recommended practices do not exist or they are especially challenging to implement, we note the need for future work as a priority; this includes, for example, R&D and standards development. We also identify potential areas in which collaboration with allies and partners would be beneficial for interoperability and trust, and note that the Key Considerations can inform potential future efforts to discuss military uses of AI with strategic competitors.

I. Aligning Systems and Uses with American Values and the Rule of Law

(1) Overview:

Our values guide our decisions and our assessment of their outcomes. Our values shape our policies, our sensitivities, and how we balance trade-offs among competing interests. Our values, and our commitment to upholding them, are reflected in the U.S. Constitution, and our laws, regulations, programs, and processes.

One of the seven principles we set forth in our Interim Report (November 2019) is the following:

> The American way of AI must reflect American values—including having the rule of law at its core. For federal law enforcement agencies conducting national security investigations in the United States, that means using AI in ways that are consistent with constitutional principles of due process, individual privacy, equal protection, and non-discrimination. For American diplomacy, that means standing firm against uses of AI by authoritarian governments to repress individual freedom or violate the human rights of their citizens. And for the U.S. military, that means finding ways for AI to enhance its ability to uphold the laws of war and ensuring that current frameworks adequately cover AI.[276]

Values established in the U.S. Constitution, and further operationalized in legislation, include freedoms of speech and assembly, the rights to due process, inclusion, fairness, non-discrimination (including equal protection), and privacy (including protection from unwarranted government interference in one's private affairs).[277] Beyond the values codified in the U.S. Constitution and the U.S. Code, our values also are expressed via international treaties that the United States has ratified that affirm our commitments to human rights and human dignity, including the International Convention of Civil and Political Rights.[278] Within America's national security departments, our commitment to protecting and upholding privacy and civil liberties is further embedded in the policies and programs of the

---

[276] *Interim Report* at 17.

[277] See e.g., U.S. Const. amendments I, IV, V, and XIV; Americans with Disability Act of 1990, 42 U.S.C. § 12101 et seq.; Title VII of the Consumer Credit Protection Act, 15 U.S.C. §§ 1691-1691f; Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq..

[278] International Covenant on Civil and Political Rights, UN General Assembly, United Nations, Treaty Series, vol. 999 at 171 (Dec. 16, 1966), https://www.refworld.org/docid/3ae6b3aa0.html. As noted in the Commission's Interim Report, America and its like-minded partners share a commitment to democracy, human dignity and human rights. See Interim Report at 48. Many, but not all nations, share commitments to these values. Even when values are shared, however, they can be culturally relative, for instance, across nations, owing to interpretative nuances.

Intelligence Community,[279] the Department of Homeland Security,[280] the Department of Defense (DoD),[281] and oversight entities.[282] This is not an exhaustive set of values that U.S. citizens would identify as core principles of the United States. However, the paradigm of considerations and recommended practices for AI that we introduce resonate with these highlighted values as they have been acknowledged and elevated as critical by the U.S. government and national security departments and agencies. Further, many of these values are common to America's like-minded partners who share a commitment to democracy, human dignity, and human rights.

In the military context, core values such as distinction and proportionality are embodied in the nation's commitment to, and the DoD's policies to uphold, the Uniform Code of Military Justice and the Law of Armed Conflict (LOAC).[283] Other values are reflected in treaties, rules, and policies such as the Convention Against

---

[279] See e.g., Daniel Coats, *Intelligence Community Directive 107*, ODNI (Feb. 28, 2018), https://fas.org/irp/dni/icd/icd-107.pdf (on protecting civil liberties and privacy); *IC Framework for Protecting Civil Liberties and Privacy and Enhancing Transparency Section 702*, Intel.gov (Jan. 2020), https://www.intelligence.gov/index.php/ic-on-the-record/guide-to-posted-documents#SECTION_702-OVERVIEW (on privacy and civil liberties implication assessments and oversight); *Principles of Professional Ethics for the Intelligence Community*, ODNI, (https://www.dni.gov/index.php/who-we-are/organizations/clpt/clpt-related-menus/clpt-related-links/ic-principles-of-professional-ethics (last visited June 17, 2020) (on diversity and inclusion).

[280] See e.g., *Privacy Office*, Department of Homeland Security (June 3, 2020), https://www.dhs.gov/privacy-office#; *CRCL Compliance Branch*, Department of Homeland Security (May 15, 2020), https://www.dhs.gov/compliance-branch.

[281] See Samuel Jenkins & Alexander Joel, *Balancing Privacy and Security: The Role of Privacy and Civil Liberties in the Information Sharing Environment*, IAPP Conference 2010 (2010), https://dpcld.defense.gov/Portals/49/Documents/Civil/IAPP.pdf.

[282] See *Projects*, U.S. Privacy and Civil Liberties Oversight Board, (last visited June 17, 2020), https://www.pclob.gov/Projects.

[283] See *Department of Defense Law of War Manual*, DoD Office of General Counsel (Dec. 2016), https://dod.defense.gov/Portals/1/Documents/pubs/DoD%20Law%20of%20War%20Manual%20-%20June%202015%20Updated%20Dec%202016.pdf?ver=2016-12-13-172036-190 [hereinafter DoD Law of War Manual]. See also *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense: Supporting Document*, Defense Innovation Board (Oct. 31, 2019), https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF ("More than 10,000 military and civilian lawyers within DoD advise on legal compliance with regard to the entire range of DoD activities, including the Law of War. Military lawyers train DoD personnel on Law of War requirements, for example, by providing additional Law of War instruction prior to a deployment of forces abroad. Lawyers for a Component DoD organization advise on the issuance of plans, policies, regulations, and procedures to ensure consistency with Law of War requirements. Lawyers review the acquisition or procurement of weapons. Lawyers help administer programs to report alleged violations of the Law of War through the chain of command and also advise on investigations into alleged incidents and on accountability actions, such as commanders' decisions to take action under the Uniform Code of Military Justice. Lawyers also advise commanders on Law of War issues during military operations.").

Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment;[284] the DoD's Rules of Engagement;[285] and the DoD's Directive 3000.09.[286]

U.S. values demand that the development and use of AI respect these foundational values, and that they enable human empowerment as well as accountability. They require that the operation of AI systems and components be compliant with our laws and international legal commitments, and with departmental policies. In short, core American values must inform the way we develop and field AI systems, and the way our AI systems behave in the world.

To date, AI Principles adopted and endorsed by the Executive Branch, including by national security department and agencies, have focused on aligning AI with many of the values discussed in this section, including fairness and non-discrimination,[287] privacy and civil liberties,[288] and accountability.[289] Taking the DoD Principles as one example, fairness is evoked by the "Equitable" principle that the department will "take deliberate steps to minimize unintended bias in AI capabilities."[290] Accountability is evoked by the "Responsible" principle that "DoD personnel will exercise appropriate levels of judgment and care while remaining responsible for the development, deployment and use of AI capabilities."[291] The work on establishing principles reiterates the importance of developing and deploying AI systems in accordance with these values. They form the foundation that the Commission's recommendations build upon.

(2) Examples of Current Challenges

Machine learning techniques can assist DoD agencies with conducting large-scale data analyses to support and enhance decision-making about personnel. As an example, the JAIC Warfighter Health Mission Initiative Integrated Disability Evaluation System model seeks to leverage data analyses to identify service members

---

[284] Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, United Nations General Assembly (Dec. 10, 1984), https://www.ohchr.org/en/professionalinterest/pages/cat.aspx.

[285] See DoD Law of War Manual at 26 (("Rules of Engagement reflect legal, policy, and operational considerations, and are consistent with the international law obligations of the United States, including the law of war.").

[286] See *Department of Defense Directive 3000.09 on Autonomy in Weapons Systems*, Department of Defense (Nov. 21 2012), https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf ("Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.").

[287] See e.g., Lopez, DOD Adopts 5 Principles; *Draft Memorandum for the Heads of Executive Departments and Agencies: Guidance for Regulation of Artificial Intelligence Applications*, Office of Management and Budget (Jan. 1, 2019), https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf.

[288] See Ben Huebner, *Presentation: AI Principles*, Intelligence and National Security Alliance 2020 Spring Symposium, Building an AI Powered IC (Mar. 4, 2020), https://www.insaonline.org/2020-spring-symposium-building-an-ai-powered-ic-event-recap/.

[289] Id.

[290] See Lopez, DOD Adopts 5 Principles.

[291] Id.

on the verge of ineligibility due to concerns with their readiness[292]. Other potential analyses can support personnel evaluations, including analyzing various factors that lead to success or failure in promotion. Caution and proven practices are needed however to avoid pitfalls in fairness and inclusiveness, several of which have been highlighted in high-profile challenges in such areas as criminal justice,[293] recruiting and hiring,[294] and face recognition.[295] Attention should be paid to challenges with decision support systems to avoid harmful disparate impact.[296] Likewise, factors chosen to weigh in performance evaluations and promotions must be carefully considered to avoid inadvertently reinforcing existing biases through ML-assisted decisions.

(3) Recommendations for Adoption

Recommended Practices to Implement American Values

A. Developing uses and building systems that behave in accordance with American values and the rule of law.
   1. **Employ technologies and operational policies that align with privacy preservation, fairness, inclusion, human rights, and law of armed conflict.** Technologies and policies throughout the AI lifecycle should support achieving the goals that AI uses and systems are consistent with these values—and should mitigate the risk that AI system uses/outcomes

---

[292] See *JAIC Mission Initiative in the Spotlight: Warfighter Health*, JAIC (Apr. 15, 2020), https://www.ai.mil/blog_04_15_20-jaic_mi_warfighter_health.html.

[293] *Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System*, Partnership on AI, (last accessed July 14, 2020), https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/.

[294] Andi Peng et al., *What You See Is What You Get? The Impact of Representation Criteria on Human Bias in Hiring*, Proceedings of the 7th AAAI Conference on Human Computation and Crowdsourcing (Oct. 2019), https://arxiv.org/pdf/1909.03567.pdf; Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, Reuters (Oct. 9, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [hereinafter Dastin. Amazon Scraps Secret AI Recruiting Tool].

[295] Patrick Grother, et. al., *Face Recognition Vendor Test (FRVT) Part Three: Demographic Effects*, National Institute of Standards and Technology (Dec. 2019), https://doi.org/10.6028/NIST.IR.8280 [hereinafter Grother, Face Recognition Vendor Test (FRVT) Part Three: Demographic Effects].

[296] PNDC provides predictive analytics to improve military readiness; enable earlier identification of service members with potential unfitting, disabling, or career-ending conditions; and offer opportunities for early medical intervention or referral into disability processing. To do so, PNDC provides recommendations at multiple points in the journey of the non-deployable service member through the Military Health System to make "better decisions" that improve medical outcomes and delivery of health services. This is very similar to the OPTUM decision support system that recommended which patients should get additional intervention to reduce costs. Analysis showed millions of US patients were processed by the system, with substantial disparate impact on black patients compared to white patients. Shaping development from the start to reflect bias issues (which can be subtle) would have produced a more equitable system and avoided scrutiny and suspension of system use when findings were disclosed. See Heidi Ledford, *Millions of Black People Affected by Racial Bias in Health Care Algorithms*, Nature (Oct. 26, 2019), https://www.nature.com/articles/d41586-019-03228-6.

will violate these values. While not an exhaustive list, we offer the following examples based upon core values discussed above:

- For ensuring *privacy*, employ privacy protections, privacy-sensitive analyses, eyes-off ML, ML with encrypted data and models, and multi-party computation methods.
- For *fairness and to mitigate unwanted bias*, use tools to probe for unwanted bias in data, inferences, and recommendations. [297]
- For *inclusion*, ensure usability of systems, accessible design, appropriate ease of use, learnability, and training availability.
- For commitment to *human rights,* place limitations and constraints on applications that would put commitment to human rights at risk, for example, limits on storing observational data beyond its specific use or using data for purposes other than its primary, intended focus.
- For compliance with the *Law of Armed Conflict*, tools for interpretability and to provide cues to the human operator should enable context-specific judgments to ensure, for instance, distinction between active combatants, those who have surrendered, and civilians.[298]

B. **Representing Objectives and Trade-offs**

Above, we described the goals of developing systems that align with key values through employing technologies and operational policies. Another important practice for aligning AI systems with values is to consider values as (1) embodied in choices about engineering trade-offs and (2) explicitly represented in the goals and utility functions of an AI system.[299]

On (1), multiple trade-offs may be encountered with the engineering of an AI system. With AI, trade-offs need to be made based on what is most valued (and the benefits and risks to those values)[300] including for high-stakes, high-risk

---

[297] Data should be appropriately biased (in a statistical sense) for what it's needed to do in order to have accurate predictions. However, beyond this, diverse concerns with unwanted bias exist, including factors that could make a system's outcomes morally or legally unfair. See Ninaresh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning*, USC Information Sciences Institute (Sept. 17, 2019) https://arxiv.org/pdf/1908.09635.pdf. For an illustration of ways fairness can be assessed across the AI lifecycle, see Sara Robinson, *Building Machine Learning Models for Everyone: Understanding Fairness in Machine Learning*, Google (Sept. 25, 2019) https://cloud.google.com/blog/products/ai-machine-learning/building-ml-models-for-everyone-understanding-fairness-in-machine-learning.

[298] For more examples on the law of armed conflict, see *Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach*, International Committee of the Red Cross (June 6, 2019), https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach.

[299] Mohsen Bayati, et al., *Data-Driven Decisions for Reducing Readmissions for Heart Failure: General Methodology and Case Study*, PLOS One Medicine (Oct. 2014), https://doi.org/10.1371/journal.pone.0109264; Eric Horvitz & Adam Seiver, *Time-Critical Action: Representations and Application*, Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (Aug. 1997), https://arxiv.org/pdf/1302.1548.pdf.

[300] Jessica Cussins Newman, *Decision Points in AI Governance: Three Case Studies Explore Efforts to Operationalize AI Principles*, Berkeley Center for Long-Term Cybersecurity (May 5, 2020), https://cltc.berkeley.edu/ai-decision-points/ [hereinafter Newman, Decision Points in AI Governance].

pattern recognition, recommendation, and decision making under uncertainty. Trade-off decisions for AI systems must be made about internal representations, policies of usage and controls, run-time execution monitoring, and thresholds. These include a number of well-known, inescapable engineering trade-offs when it comes to building and using machine-learning to develop models for prediction, classification, and perception. For example, systems that perform recognition or prediction tasks can be set to work at different operating thresholds or settings (along a well-characterized curve) where different settings change the trade-off between precision and recall or the rates of true positives and false positives. By changing the settings, the ratio of true positives to false positives is changed. Often, one can raise the rate of true positives but will also raise the false negatives.[301] In high-stakes applications, different kinds of inaccuracies (e.g., missing a recognition and falsely recognizing) are associated with different outcomes and costs. Thus, the setting of thresholds and understanding the influences of different settings on the behavior of a system *entail making value judgments*. As with all engineering trade-offs, making choices about trade-offs explicitly and deliberately provides more transparency, accountability, and confidence in the process than making decisions implicitly and ad hoc as they arise.

On (2), systems may be guided by optimization processes that seek to maximize an objective function or *utility model*. Such objectives can represent a set of independent goals, as in *multi-objective optimization*. A multi-attribute utility function may be employed to guide a system's actions based on an objective that is constructed by weighing several individual factors, where explicit weights are assigned to capture the asserted importance of each of the different factors. Different weightings on factors can be viewed as embedding different values into a system. Here too trade-offs are made when using multi-attribute utility or objective functions within applications.[302] Even when tuning a model for fairness, when multiple metrics of fairness are relevant, optimizing for one metric can cause a trade-off in performance across the second metric.[303] As a result, it is important to acknowledge inherent trade-offs and the need for setting or encoding "preferences" - which requires *someone* or *some organization* to make a call

[301] For more on the trade-offs between false positive and false negative rates, and the implications of chosen thresholds, see Grother, Face Recognition Vendor Test (FRVT) Part Three: Demographic Effects.

[302] Optimal decisions may require making a decision when trade-offs exist between two or more conflicting objectives. For example, a predictive maintenance system for aircraft will have objectives that are in tension including: minimizing false positives, minimizing false negatives, minimizing the need for instrumentation on the aircraft, maximizing the specificity of the recommended maintenance action, and adapting to new operational profiles the aircraft perform in over time.

[303] It is sometimes impossible to simultaneously satisfy different fairness criteria. See Yungfeng Zhang, et al., *Joint Optimization of AI Fairness and Utility: A Human-Centered Approach, Association for Computing Machinery*, AIES '20 (Feb. 7-8, 2020), https://dl.acm.org/doi/10.1145/3375627.3375862.

about the right trade.[304]

Recommended Practices for Representing Objectives and Trade-offs

1. **Consider and document value considerations in AI systems and components based on specifying how trade-offs with accuracy are handled;** this includes operating thresholds that yield different true positive and false positive rates or different precision and recall.
2. **Consider and document value considerations in AI systems that rely on representations of objective or utility functions,** including the handling of multi-attribute or multi-objective models.
3. **Conduct documentation, reviews, and set limits on disallowed outcomes.** It is important to:
    - Be transparent and keep documentation on assertions about the trade-offs made, optimization justifications, and acceptable thresholds for false positives and false negatives.
    - During system development and testing, consider the potential need for context-specific changes in goals or objectives that would require a revision of parameters on settings or weightings on factors.
    - Establish explicit controls in specific use cases and have the capability to change or set controls, potentially by context or by policy, per organization.
    - Review documentation and run-time execution trade-offs, potentially on a recurrent basis, by appropriate experts/authorities.
    - Acknowledge that performance characteristics are statistics over multiple cases, and that different settings and workloads have different performance.
    - Set logical limits on disallowed outcomes, where needed, to put additional constraints on allowed performance.

(4) Recommendations for Future Action

Future R&D is needed to advance capabilities for preserving and ensuring that developed or acquired AI systems will act in accordance with American values and the rule of law. For instance, the Commission notes the need for R&D to assure that the personal privacy of individuals is protected in the acquisition and use of data for AI system development.[305] This includes advancing ethical practices with the use of personal data, including disclosure and consent about data collection and use models (including uses of data to build base models that are later retrained and fine-tuned for specific tasks). R&D should also advance development of anonymity techniques and privacy-preserving technologies including homomorphic encryption and differential

[304] See *Analyses of Alternatives, Systems Engineering Guide*, MITRE (May 2014), https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/acquisition-program-planning/performing-analyses-of-alternatives.

[305] The Commission is doing a fulsome assessment of where investment needs to be made; this document notes important R&D areas through the lens of ethics and responsible AI.

privacy techniques and identify optimal approaches for specific use cases. Research should focus upon advancing multi-party compute capabilities (to allow collaboration on the pooling of data from multiple organizations without sharing datasets), and developing a better understanding of the compatibility of the promising privacy preserving approaches with regulatory approaches such as the European Union's General Data Protection Regulation (GDPR), as both areas are important for allied cooperation.

II. Engineering Practices

(1) Overview

The government, and its partners (including vendors), should adopt recommended practices for creating and maintaining trustworthy and robust AI systems that are *auditable* (able to be interrogated and yield information at each stage of the AI lifecycle to determine compliance with policy, standards, or regulations[306]); *traceable* (to understand the technology, development processes, and operational methods applicable to AI capabilities, e.g., with transparent and auditable methodologies, data sources, and design procedure and documentation[307]); *interpretable* (to understand the value and accuracy of system output[308]), *and reliable* (to perform in the intended manner within the intended domain of use[309]).

There are no broadly directed best practices or standards (e.g., endorsed by the Secretary of Defense or Director of National Intelligence) in place to define how organizations should build AI systems that are consistent with designated AI principles. But efforts in commercial, scientific, research, and policy communities are generating candidate approaches, minimal standards, and engineering proven practices to ensure the responsible design, development, and deployment of AI systems.[310]

While AI refers to a constellation of technologies, including logic-based systems, the rise in capabilities in AI systems over the last decade is largely attributable to

---

[306] See Inioluwa Deborah Raji, et al., *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing,* ACM FAT (Jan. 3, 2020), https://arxiv.org/abs/2001.00973 [hereinafter Raji, Closing the AI Accountability Gap].

[307] Lopez, DOD Adopts 5 Principles.

[308] *Model Interpretability in Azure Machine Learning (preview)*, Microsoft (July 2020), https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability.

[309] Lopez, DOD Adopts 5 Principles.

[310] See Newman, Decision Points in AI Governance; Raji, Closing the AI Accountability Gap; Miles Brundage, et al., *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims* (Apr. 20, 2020), https://arxiv.org/abs/2004.07213 [hereinafter Brundage, Toward Trustworthy AI Development]; Saleema Amershi, et. al., *Software Engineering for Machine Learning: A Case Study*, Microsoft (Mar. 2019), https://www.microsoft.com/en-us/research/uploads/prod/2019/03/amershi-icse-2019_Software_Engineering_for_Machine_Learning.pdf [hereinafter Amershi, Software Engineering for Machine Learning].

capabilities provided by data-centric machine learning (ML) methods. New security and robustness challenges are linked to different phases of ML system construction and operations.[311] Several properties of the methods and models used in ML are associated with weaknesses that make the systems brittle and exploitable in specific ways—and vulnerable to failure modalities not seen in traditional software systems. Such failures can rise inadvertently or as the intended results of malicious attacks and manipulation. Attributes of machine learning training procedures and run-times linked to intentional and unintentional failures include: (1) the critical reliance on data for training, (2) the common use of such algorithmic procedures as differentiation and gradient descent to construct and optimize the performance of models, (3) the ability to probe models with multiple tasks or queries, and (4) the possibility of gaining access to information about models and their parameters.

Given the increasing consequences of failure in AI systems as they are integrated into critical uses, the various failure modes of AI systems have received significant attention. The exploration of AI failure modes has been divided into adversarial attacks[312] or unintended faults introduced throughout the lifecycle.[313] The pursuit of security and robustness of AI systems requires awareness, attention, and proven practices around intentional and unintentional failure modes.[314]

Intentional failures are the result of malicious actors explicitly attacking some aspect of (AI) system training or run-time behavior. Researchers and practitioners in the evolving area of Adversarial Machine Learning *(*AML) have created taxonomies of malicious attacks on machine learning training procedures and run-times. Attacks span ML training and testing and each has associated defenses.[315] Categories of intentional failures introduced by adversaries include training *data poisoning* attacks, *model inversion*, and ML *supply chain attacks*.[316] National security uses of AI are likely targets of sustained adversarial efforts; awareness of sets of potential vulnerabilities and proven practices for detecting attacks and protecting systems is critical. AI developed for this community must remain current with a rapidly developing

---

[311]Elham Tabassi, et al., *A Taxonomy and Terminology of 4 Adversarial Machine Learning (Draft NISTIR 8269)*, National Institute of Standards and Technology (Oct. 2019), https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf [hereinafter Tabassi, A Taxonomy and Terminology of 4 Adversarial Machine Learning (Draft NISTIR 8269)].

[312]See Guofu Li, et al., *Security Matters: A Survey on Adversarial Machine Learning*, (Oct. 2018), https://arxiv.org/abs/1810.07339; Tabassi, A Taxonomy and Terminology of 4 Adversarial Machine Learning (Draft NISTIR 8269).

[313]See José Faria, *Non-Determinism and Failure Modes in Machine Learning*. 2017 IEEE 28th International Symposium on Software Reliability Engineering Workshops (Oct. 2017), https://ieeexplore.ieee.org/document/8109300; Dario Amodei et al., *Concrete Problems in AI Safety*, (Jun. 2016), https://arxiv.org/abs/1606.06565.

[314] Ram Shankar Siva Kumar et al., *Failure Modes in Machine Learning*, (Nov. 2019), https://docs.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning [hereinafter, Kumar, Failure Modes in Machine Learning].

[315]See Tabassi, A Taxonomy and Terminology of 4 Adversarial Machine Learning (Draft NISTIR 8269).

[316] For 11 categories of attack, and associated overviews, see the "Intentionally-Motivated Failures Summary" in Kumar, Failure Modes in Machine Learning.

understanding of the nature of vulnerabilities to attacks as these attacks grow in sophistication. Advances in new attack methods and vectors must be followed with care and recommended practices implemented around technical and process methods for mitigating vulnerabilities and detecting, alerting, and responding to attacks.

Unintentional failures can be introduced at multiple points in the AI development and deployment lifecycle. In addition to faults that can be inadvertently introduced into any software development effort (e.g., requirements ambiguity, coding errors, inadequate TEVV, flaws in tools used to develop and evaluate the system), distinct additional failure modes can be introduced for machine learning systems. Examples of unintentional AI failures (with particular relevance to deep learning and reinforcement learning) include *reward hacking, side-effects, distributional shifts*, and *natural adversarial examples*.[317] Another area of failure includes the inadequate specification of values per objectives represented in system utility functions (as described in Section 1 above on *Representing Objectives and Trade-offs*), leading to unexpected and costly behaviors and outcomes, akin to outcomes in the fable of the Sorcerer's Apprentice[318]. Additional classes of unintentional failures can arise as unexpected and potentially costly behaviors generated via the interactions of multiple distinct AI systems that are each developed and tested in isolation. The explicit or inadvertent composition of sets of AI systems within one's own services, forces, agencies, and between US systems and those of allies, adversaries, and potential adversaries, can lead to complex multi-agent situations with unexpected and poorly-characterized behaviors.[319]

(2) Examples of Current Challenges

To make high-stakes decisions, and often in safety-critical contexts, DoD and the IC must be able to depend on the integrity and security of the data that is used to train some kinds of ML systems. The challenges of doing so have been echoed by the leadership of the DoD and the Intelligence Community,[320] including concerns with

---

[317] Id.

[318] Thomas Dietterich & Eric Horvitz, *Rise of Concerns about AI: Reflections and Directions*, Communications of the ACM, Vol. 58 No. 10 at 38-40 (Oct. 2015), http://erichorvitz.com/CACM_Oct_2015-VP.pdf .

[319] Unexpected performance represents emergent runtime output, behavior, or effects at the system level, e.g., through unanticipated feature interaction, … that was also not previously observed during model validation." See Colin Smith et al., *Hazard Contribution Modes of Machine Learning Components*, AAAI-20 Workshop on Artificial Intelligence Safety (SafeAI 2020) (Feb. 7, 2020), https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20200001851.pdf.

[320] See Don Rassler, *A View from the CT Foxhole Lieutenant General John N.T. "Jack" Shanahan, Director, Joint Artificial Intelligence Center, Department of Defense*, Combating Terrorism Center at West Point (Dec. 2019), https://ctc.usma.edu/view-ct-foxhole-lieutenant-general-john-n-t-jack-shanahan-director-joint-artificial-intelligence-center-department-defense/ ("I am very well aware of the power of information, for good and for bad. The profusion of relatively low-cost, leading-edge information-related capabilities and advancement of AI-enabled technologies such as generative adversarial networks or GANs, has made it possible for almost anyone—from a state actor to a lone wolf terrorist—to use information as a precision weapon. What was viewed largely as an annoyance a few years ago has now

detecting adversarial attacks such as data poisoning, sensor spoofing, and "enchanting attacks" (when the adversary lures a reinforcement learning agent to a designated target state that benefits the adversary).[321]

(3) Recommendations for Adoption

Engineering Recommended Practices

Critical engineering practices needed to operationalize AI principles (such as 'traceable' and 'reliable'[322]) are described in the non-exhaustive list below. These practices span design, development, and deployment of AI systems.

1. **Concept of operations development and design and requirements definition and analysis**. Conduct systems analysis of operations and identify mission success metrics. Identify potential functions that can be performed by an AI technology. Incorporate early analyses of use cases and scenario development, assess general feasibility, and make a critical assessment of the reproducibility and demonstrated maturity of specific candidate AI technologies. This includes broad stakeholder engagement and hazard analysis, including domain experts and individuals with expertise and/or training in the responsible development and fielding of AI technologies. This includes for example asking key questions about potential disparate impact early in the development process and documenting deliberations, actions, and approaches used to ensure fairness and lack of unwanted bias in the machine learning application.[323] The feasibility of

---

become a serious threat to national security. Even more alarming, it's almost impossible to predict the exponential growth of these information-as-a-weapon capabilities over the next few years."); see also Dean Souleles, *2020 Spring Symposium: Building an AI Powered IC*, Intelligence and National Security Alliance (Mar. 9, 2020), https://www.insaonline.org/2020-spring-symposium-building-an-ai-powered-ic-event-recap/ ("We need to be thinking of authenticity of information and provenance of information.....How do you know that the news you are reading is authentic news? How do you know that its source has provenance? How can you trust the information of the world? And in this era of deep fakes and generative artificial neural networks scans that can produce images and texts and videos and audio that are increasingly indistinguishable from authentic, where then is the role of the intelligence officer? If you can no longer meaningfully distinguish truth from falsehood, how do you write an intelligence report? How do you tell national leadership with confidence you believe something to be true or not to be true. That is a big challenge. . . . We need systems that are reliable and understandable. We need to be investing in the gaps.").

[321] Naveed Akhtar & Ajmal Mian, *Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey* (Feb. 2018), https://arxiv.org/abs/1801.00553.

[322] See *DOD Adopts Ethical Principles for Artificial Intelligence*, Department of Defense (Feb. 24, 2020) https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

[323] There is no single definition of fairness. System developers and organizations fielding applications must work with stakeholders to define fairness, and provide transparency via disclosure of assumed definitions of fairness. Definitions or assumptions about fairness and metrics for identifying fair inferences and allocations should be explicitly documented. This should be accompanied by a discussion of alternate definitions and rationales for the current choice. These elements should be

meeting these requirements may trigger a review of whether and where it is appropriate to use AI in the system being proposed. Opportunities exist to use experimentation, modeling/simulation, and rapid prototyping of AI systems to validate operational requirements and assess feasibility.[324]

- **Risk assessment**. In conducting stakeholder engagement and hazard analysis, it is important to assess risks and trade-offs with a diverse interdisciplinary group. This includes a discussion of a system's potential societal impact. Prior to developing or acquiring a system, or conducting AI R&D in a novel area, risk assessment questions should be asked relevant to the national security context in critical areas, including questions about privacy and civil liberties, the law of armed conflict, human rights,[325] system security, and the risks of a new technology being leaked, stolen, or weaponized. [326]

2. **Documentation of the AI lifecycle:** Whether building and fielding an AI system or "infusing AI" into a preexisting system, require documentation[327] on:

---

documented internally as machine-learning components and larger systems are developed. This is especially important as establishing alignment on the metrics to use for assessing fairness encounters an added challenge when different cultural and policy norms are involved when collaborating on development and use with allies.

[324] Design reviews take place at multiple stages in the Defense Acquisition process. Recent reforms to the Defense Acquisition System efforts, include the release of a new DoD 5000.02, which issues the "Adaptive Acquisition Framework" and an interim policy for a software acquisition pathway; this reflects efforts to further adapt the system to support agile and iterative approaches to software-intensive system development. See *Software Acquisition,* Defense Acquisition University (last visited June 18, 2020), https://aaf.dau.edu/aaf/software/; *DoD Instruction 5000.02: Operation Of The Adaptive Acquisition Framework*, Office of the Under Secretary of Defense for Acquisition and Sustainment (Jan. 23, 2020) https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf?ver=2020-01-23-144114-093.

[325] For more on the importance of human rights impact assessments of AI systems, see *Report of the Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression,* UN Human Rights Office of the High Commissioner (2018), https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx. For an example of a human rights risk assessment for AI in categories such as nondiscrimination and equality, political participation, privacy, and freedom of expression, see Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, Data Society (Oct. 2018),. https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

[326] For exemplary risk assessment questions that IARPA has used, see Richard Danzig, *Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority*, Center for a New American Security at 22 (June 28, 2018), https://s3.amazonaws.com/files.cnas.org/documents/CNASReport-Technology-Roulette-DoSproof2v2.pdf?mtime=20180628072101.

[327] Documentation recommendations build off of a legacy of robust documentation requirements. See *Department of Defense Standard Practice: Documentation of Verification, Validation, and Accreditation (VV&A) For Models and Simulations*, Department of Defense (Jan. 28, 2008), https://acqnotes.com/Attachments/MIL-STD-3022%20Documentation%20of%20VV&A%20for%20Modeling%20&%20Simulation%2028%20Jan%2008.pdf.

- If ML is used, the data used for training and testing, including clear and consistent annotation of data, the origin of the data (e.g., why, how, and from whom), provenance, intended uses, and any caveats with re-uses;[328]
- The algorithm(s) used to build models, characteristics about the model (e.g, training), and the intended uses of the AI capabilities separately or as part of another system;
- Connections between and dependencies within systems, and associated potential complications;
- The selected testing methodologies and performance indicators and results for models used in the AI component (e.g., confusion matrix and thresholds for true and false positives and true and false negatives area under the curve (AUC) as metrics for performance/error); this includes how tests were done, and the simulated or real-world data used in the tests--including caveats about the assumptions of the training and testing, per type of scenarios, per the data used in testing and training;
- Required maintenance, including re-testing requirements, and technical refresh. This includes requirements for re-testing, retraining, and tuning when a system is used in a different scenario or setting (including details about definitions of scenarios and settings) or if the AI system is capable of online learning or adaptation.

3. **Infrastructure to support traceability.** Invest resources and establish policies that support the traceability of AI systems. Traceability, critical for high-stakes systems, captures key information about the system development and deployment process for relevant personnel to adequately understand the technology.[329] It includes selecting, designing, and implementing measurement tools, logging, and monitoring and applies to (1) development and testing of AI systems and components,[330] (2) operation of AI systems,[331] (3) users and their behaviors in engaging with AI systems or components,[332]

---

[328] For an industry example, see Timnit Gebru et al., *Datasheets for Datasets*, Microsoft (March 2018), https://www.microsoft.com/en-us/research/publication/datasheets-for-datasets/. For more on data, model and system documentation, see *Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles (ABOUT ML)*, an evolving body of work from the Partnership on AI about documentation practices at https://www.partnershiponai.org/about-ml/. See also David Thornton, *Intelligence Community Laying Foundation for AI Data Analysis*, Federal News Network (Nov. 1, 2019), https://federalnewsnetwork.com/all-news/2019/11/intelligence-community-laying-the-foundation-for-ai-data-analysis/ (documenting any caveats of re-use for both datasets and models is critical to avoid "off-label" use harms).

[329] Jonathan Mace et al., *Pivot Tracing: Dynamic Causal Monitoring for Distributed Systems*, Communications of the ACM (March 2020), https://www.cs.purdue.edu/homes/bb/cs542-20Spr/readings/others/pivot-tracing-cacm-202003.pdf [hereinafter, Mace, Pivot Tracing].

[330] Examples include logs of steps taking in problem and purpose definition, design, training and development. See e.g., Brundage, Toward Trustworthy AI Development.

[331] This includes logs of steps taken in operation which can support retrospective accident analysis. Id.

[332] Examples include logs of access and use of the system by operators, per understanding human access, oversight; nonrepudiation (e.g., cryptographic controls on access).

and (4) auditing.[333] Audits should support analyses of specific actions as well as characterizations of longer-term performance. Audits should also be done to assure that performance on tests of the system and on real-world workloads meet requirements, such as fairness asserted at specification of the system and/or established by stakeholders.[334] When a criminal investigation requires it, forensic analyses of the AI system must be supported. A recommended practice is to carefully consider how you expose APIs for audit trails and traceability infrastructure in light of the potential vulnerability to an adversary detecting how an algorithm works and conducting an attack using counter AI exploitation.

4. **Security and Robustness: Addressing Intentional and Unintentional Failures**
    - **Adversarial attacks, and use of robust ML methods**. Expand notions of adversarial attacks to include various "machine learning attacks," which may take the form of an attack through supply chain, online access, adversarial training data, or model inference attacks, including through Generative Adversarial Networks (GANS).[335]

---

[333] Auditing examples include real-time system health and behavior monitoring, longer-term reporting, via logging of system recommendations, classifications, or actions and why they were taken per input, internal states of the system that were important in the chain of inferences and ultimate actions, and the actions taken, and logs to assure maintenance of accountability for decision systems (e.g. signoff for a specific piece of business logic).

[334] All of the above are consistent with, and support the fulfillment of, the DOD's AI Principle, Traceable.

Documentation practices that support traceability (e.g. data sources and design procedures and documentation) are expanded upon in additional bullets throughout the Engineering Practices section. See Lopez, DOD Adopts 5 Principles ("Traceable: - The department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources and design procedures and documentation.").

[335] The approach is to simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. As the generator gets better (producing ever more credible samples) the discriminator also improves (getting ever better at discerning real samples from the generated "fake" samples). This is useful for improving discriminator performance. Given the vulnerability of deep learning models to adversarial examples (slight changes in an input that produce significantly different results in output and can be used to confound a classifier), there has been interest in using adversarial inputs in a GAN framework to train the discriminator to better distinguish adversarial inputs. There is also considerable theoretical work being done on fundamental approaches to making DL more robust to adversarial examples. This remains an important focus of research. For more on adversarial attacks, see e.g., Ian Goodfellow et al., *Generative Adversarial Networks*, Universite de Montreal (June 10, 2014), https://arxiv.org/abs/1406.2661; Ian Goodfellow et. al., *Explaining And Harnessing Adversarial Examples*, Google (Mar. 20, 2015), https://arxiv.org/pdf/1412.6572.pdf; Kevin Eykholt, et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition at 1625–1634 (2018), https://arxiv.org/abs/1707.08945; Anish Athalye, et al., *Synthesizing Robust Adversarial Examples*, International conference on machine learning (2018), https://arxiv.org/pdf/1707.07397.pdf; Kevin Eykholt, et al., *Physical Adversarial Examples for Object Detectors*, USENIX Workshop on Offensive Technologies (2018), https://arxiv.org/abs/1807.07769; Yulong Cao, et al., *Adversarial Sensor Attack on LiDAR-based Perception*

Agencies should seek latest technologies that demonstrate the ability to detect and notify operators of attacks, and also tolerate attacks. [336]

- **Follow and incorporate advances in intentional and unintentional ML failures**. Given the rapid evolution of the field of study of intentional and unintentional ML failures, national security organizations must follow and adapt to the latest knowledge about failures and proven practices for monitoring, detection, and engineering and run-time protections. Related efforts and R&D focus on developing and deploying robust AI methods.[337]
- **Adopt a security development lifecycle (SDL) for AI** systems to include a focus on potential failure modes. This includes developing and regularly refining threat models to capture and consolidate the characteristics of various attacks in a way that can shape system development to mitigate vulnerabilities.[338] A matrixed focus for developing and refining threat models is valuable. SDL should address ML development, deployment, and when ML systems are under attack.[339]

5. **Conduct red teaming** for both intentional and unintentional failure modalities. Bring together multiple perspectives to rigorously challenge AI systems, exploring the risks, limitations, and vulnerabilities in the context in which they'll be deployed.
    - To mitigate intentional failure modes – Employ methods that can make systems more resistant to adversarial attacks, work with

---

in Autonomous Driving, Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (2019), https://dl.acm.org/doi/10.1145/3319535.3339815; Mahmood Sharif, et al., *Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition*, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (2016) https://dl.acm.org/doi/10.1145/2976749.2978392; Stepan Komkov & Aleksandr Petiushko, *Advhat: Real-World Adversarial Attack on Arcface Face ID System* (Aug. 23, 2019), https://arxiv.org/pdf/1908.08705.pdf. On directions with robustness, see e.g., Aleksander Madry, et al., *Towards Deep Learning Models Resistant to Adversarial Attacks*. MIT (Sep. 4, 2019), https://arxiv.org/abs/1706.06083 [hereinafter Madry, Toward Deep Learning Models Resistant to Adversarial Attacks]; Mathias Lecuyer, et al., *Certified Robustness to Adversarial Examples with Differential Privacy*, IEEE Symposium on Security and Privacy (2019), https://arxiv.org/abs/1802.03471; Eric Wong & J. Zico Kolter, *Provable Defenses Against Adversarial Examples via the Convex Outer Adversarial Polytope*, International Conference on Machine Learning (2018), https://arxiv.org/abs/1711.00851.

[336] Madry, Towards Deep Learning Models Resistant to Adversarial Attacks.

[337]See e.g., *Id.*; Thomas Dietterich, *Steps Toward Robust Artificial Intelligence*, AI Magazine at 3-24 (Fall 2017), https://www.aaai.org/ojs/index.php/aimagazine/article/view/2756/2644; Eric Horvitz, *Reflections on Safety and Artificial Intelligence*, Safe AI: Exploratory Technical Workshop on Safety and Control for AI, White House OSTP and Carnegie Mellon University, Pittsburgh, PA (June 27, 2016), http://erichorvitz.com/OSTP-CMU_AI_Safety_framing_talk.pdf.

[338] See Andrew Marshall et al, *Threat Modeling AI/ML Systems and Dependencies* (Nov. 2010), https://docs.microsoft.com/en-us/security/engineering/threat-modeling-aiml.

[339] Ram Shankar Siva Kumar et al., *Adversarial Machine Learning—Industry Perspectives*, 2020 IEEE Symposium on Security and Privacy (SP) Deep Learning and Security Workshop, (May 2020), https://arxiv.org/pdf/2002.05646.pdf.

adversarial testing tools, and deploy teams dedicated to trying to brake systems and make them violate rules for appropriate behavior.

- To mitigate unintentional failure modes - test ML systems per a thorough list of realistic conditions they are expected to operate in. When selecting third-party components, consider the impact that a security vulnerability in them could have to the security of the larger system into which they are integrated. Have an accurate inventory of third-party components and a plan to respond when new vulnerabilities are discovered.[340]
- Because of the scarcity of required expertise and experience for AI red teams, organizations should consider establishing broader enterprise-wide communities of AI red teaming capabilities that could be applied to multiple AI developments (e.g., at a DoD service or IC element level, or higher).

(4) Recommendations for Future Action

- *For documentation:* The Commission noted the urgency of a documentation strategy in its First Quarter Recommendations.[341] Future work is needed to ensure sufficient documentation by all national security departments and agencies, including the precisions noted above in this section. In the meantime, national security departments and agencies should pilot documentation approaches across the AI lifecycle to help inform such a strategy.
- *To improve traceability:* While recommended practices exist for audit trails, standards have yet to be developed.[342] Future work is needed by standard setting bodies, alongside national security departments/agencies and the broader AI community (including industry), to develop audit trail requirements per mission needs for high-stakes AI systems including safety-critical applications.
- Future R&D is needed to advance capabilities for:
  - AI security and robustness - to cultivate more robust methods that can overcome adverse conditions; advance approaches that enable assessment of types and levels of vulnerability and immunity; and to enable systems to withstand or to degrade gracefully when targeted by a deliberate attack.
  - Interpretability - to support risk assessment and better understand the efficacy of interpretability tools and possible interfaces. (Complementary

[340] See *What are the Microsoft SDL Practices?*, Microsoft (last accessed July 14, 2020), https://www.microsoft.com/en-us/securityengineering/sdl/practices.

[341] See *First Quarter Recommendations*, NSCAI (Mar. 2020) https://www.nscai.gov/reports. Ongoing efforts to share best practices for documentation among government agencies through GSA's AI Community of Practice further indicate the ongoing need and desire for common guidance.

[342] For more on current gaps in audit trail standards for AI systems, see Brundage, Toward Trustworthy AI Development at 25 ("Existing standards often define in detail the required audit trails for specific applications. For example, IEC 61508 is a basic functional safety standard required by many industries, including nuclear power. Such standards are not yet established for AI systems.").

to this R&D, standards work is needed to develop benchmarks that assess the reliability of produced model explanations.)

III. System Performance

(1) Overview

Fielding AI systems in a responsible manner includes establishing confidence that the technology will perform as intended, especially in high-stakes scenarios.[343] An AI system's performance must be assessed,[344] including assessing its capabilities and blind spots with data representative of real-world scenarios or with simulations of realistic contexts,[345] and its reliability and robustness (i.e., resilience in real-world settings—including adversarial attacks on AI components) during development and in deployment.[346] For example, a system's performance on recognition tasks can be characterized by its false positives and false negatives on a test set representative of the environment in which a system will be deployed, and test sets can be varied in realistic ways to estimate robustness. Testing protocols and requirements are essential for measuring and reporting on system performance, including reliability, during the test phase (pre-deployment) and in operational settings. (The Commission uses industry terminology 'testing' to broadly refer to what the DoD calls "Test, Evaluation, Verification, and Validation" (TEVV). This testing includes both what DoD refers to as Developmental Test and Evaluation and Operational Test and Evaluation). AI systems present new challenges to established testing protocols and requirements as they increase in complexity, particularly for operational testing. However, there are some existing methods to continuously monitor AI system performance. For example, high-fidelity performance traces and means for sensing shifts, such as distributional shifts in targeted scenarios, permit ongoing monitoring to ensure system performance does not stray outside of acceptable parameters; if inadequate performance is detected, they provide insight needed to improve and update systems.[347]

---

[343] This includes, for example, safety-critical scenarios or those where AI-assisted decision making would impact an individual's life or liberty.

[344] Ben Shneiderman, *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy*, International Journal of Human–Computer Interaction (Mar. 23, 2020), https://doi.org/10.1080/10447318.2020.1741118 [hereinafter Shneiderman, Human Centered Artificial Intelligence: Reliable, Safe & Trustworthy].

[345] However, test protocols must acknowledge test sets may not be fully representative of real-world usage.

[346] See Brundage, Toward Trustworthy AI Development; Ece Kamar, et al., *Combining Human and Machine Intelligence in Large-Scale Crowdsourcing*, Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (June 2012), https://dl.acm.org/doi/10.5555/2343576.2343643 [hereinafter Kamar, Combining Human and Machine Intelligence in Large-Scale Crowdsourcing].

[347] For a technical paper that puts monitoring in development lifecycle context, see Amershi, Software Engineering for Machine Learning. For a good example of open source frameworks to support, see *Overview*, Prometheus, (last accessed June 18, 2020), https://prometheus.io/docs/introduction/overview/.

System performance characterization also includes assessing robustness. As noted above, this entails determining how resilient the system is in real-world settings where there may be blocking and handling of attacks and where natural real-world variation exists.[348] In addition to reliability, robustness, and security, system performance must also measure compliance with requirements derived from values such as fairness.

When evaluating system performance, it is especially important to take into account holistic, end-to-end system behavior. Emergence is the principle that entities exhibit properties which are meaningful only when attributed to the whole, not to its parts. Emergent system behavior can be viewed as a consequence of the interactions and relationships among system elements rather than the independent behavior of individual elements. It emerges from a combination of the behavior and properties of the system elements and the systems structure or allowable interactions between the elements, and may be triggered or influenced by a stimulus from the systems environment. [349]

The System Engineering Community and the National Security Community have focused on system of systems engineering for years,[350] but AI-intensive systems introduce additional opportunities and challenges for emergent performance. Given the requirement to establish and preserve justified confidence in the performance of AI systems, attention must be paid to the potential for undesired interactions and emergent performance as AI systems are composed. This composition may include pipelines where the output of one system is part of the input for another in a potentially complex and distributed ad hoc pipeline.[351] As a recent study of the software engineering challenges introduced by developing and deploying AI systems at scale notes, "AI components are more difficult to handle as distinct modules than traditional software components — models may be 'entangled' in complex ways."[352] These challenges are pronounced when the entanglement is the result of system composition and integration.

---

[348] Joel Lehman, *Evolutionary Computation and AI Safety: Research Problems Impeding Routine and Safe Real-world Application of Evolution* (Oct. 4, 2019), https://arxiv.org/abs/1906.10189 [hereinafter Lehman, Evolutionary Computation and AI Safety].

[349] Greg Zacharias, *Autonomous Horizons: The Way Forward*, Air University Press at 61 (Mar. 2019), https://www.airuniversity.af.edu/Portals/10/AUPress/Books/b_0155_zacharias_autonomous_horizons.pdf.

[350] Judith Dahmann & Kristen Baldwin, *Understanding the Current State of US Defense Systems of Systems and the Implications for Systems Engineering*, Presented at IEEE Systems Conference (Apr. 7-10, 2008), https://ieeexplore.ieee.org/document/4518994.

[351] D. Sculley, et al., *Machine Learning: The High Interest Credit Card of Technical Debt*, Google (2014), https://research.google/pubs/pub43146/ [hereinafter Sculley, Machine Learning: The High Interest Credit Card of Technical Debt].

[352] Amershi, Software Engineering for Machine Learning (illustrating non-monotonic error as a possible complexity result from model entanglement).

As America's AI-intensive systems may increasingly be composed (including through ad hoc opportunities to integrate systems) with allied AI-intensive systems, this becomes a topic for coordination with allies as well. Multi-agent systems are being explored and adopted in multiple domains,[353] as are swarms, fleets, and teams of autonomous systems.[354]

(2) Examples of Current Challenges

Unexpected interactions and errors commonly occur in integrated simulations and exercises as an illustration of the challenges of predicting and managing behaviors of systems composed of multiple components. Intermittent failures can transpire after composing different systems; these failures are not the result of any one component having errors, but rather are due to the interactions of the composed systems.[355]

(3) Recommendations for Adoption

Critical practices for ensuring optimal system performance are described in the following non-exhaustive list:

System Performance Recommended Practices

A. **Training and Testing: Procedures should cover key aspects of performance and appropriate performance metrics. These include:**
   1. **Standards for metrics and reporting needed to adequately achieve:**
      a. Consistency across testing and test reporting for critical areas.
      b. Testing for blinds pots as a specific failure mode of importance to some ML implementations.[356]
      c. Testing for fairness. When testing for fairness, sustained fairness assessments are needed throughout development and deployment, including assessing a system's accuracy and errors relative to one or more agreed to statistical definitions of fairness[357] and documenting

---

[353] Ali Dorri, et al., *Multi-Agent Systems: A Survey*, IEEE Access at 28573-28593 (Apr. 20, 2018), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8352646.

[354] Andrew Ilachinski, *AI, Robots, and Swarms: Issues, Questions, and Recommended Studies*, CNA (Jan. 2017), https://www.cna.org/CNA_files/PDF/DRM-2017-U-014796-Final.pdf.

[355] David Sculley et al., *Hidden Technical Debt in Machine Learning Systems*, NIPS '15: Proceedings of the 28th International Conference on Neural Information Processing Systems (Dec. 2015), https://dl.acm.org/doi/10.5555/2969442.2969519.

[356] Ramya Ramakrishnan et al., *Blind Spot Detection for Safe Sim-to-Real Transfer*, Journal of Artificial Intelligence Research 67 at 191-234 (2020), https://www.jair.org/index.php/jair/article/view/11436.

[357] There is no single definition of fairness. System developers and organizations fielding applications must work with stakeholders to define fairness, and provide transparency via disclosure of assumed definitions of fairness. Definitions or assumptions about fairness and metrics for identifying fair inferences and allocations should be explicitly documented. This should be accompanied by a discussion of alternate definitions and rationales for the current choice. These elements should be documented internally as machine-learning components and larger systems are developed. This is especially important as establishing alignment on the metrics to use for assessing fairness encounters

deliberations made on the appropriate fairness metrics to use.[358] Agencies should also conduct outcome and impact analysis to detect when subtle assumptions in the system concept of operations and requirements are showing up as unexpected and undesired outcomes in the operational environment.[359]

   d. Articulation of performance standards and metrics. This includes ways to communicate to the end user the meaning/significance of performance metrics, e.g., through a probability assessment, based on sensitivity and specificity. It also requires clear documentation of system performance (across diverse environments or contexts), including information content of model output.

2. **Representativeness of the data and model for the specific context at hand.** For machine learning models, challenges exist when transferring a model to a context/setting that differs from the one for which it was trained and tested. When using classification and prediction technologies, challenges with representativeness of data used in analyses, and fairness/accuracy of inferences and recommendations made with systems leveraging that data when applied in different populations/contexts, should be considered explicitly and documented. As appropriate, robust and reliable methods can be used to enable model generalization and transfer beyond the training context.
3. **Evaluating an AI system's performance relative to current benchmarks** where possible. Benchmarks should assist in determining if an AI system's performance meets or exceeds current best performance.
4. **Evaluating aggregate performance of human-machine teams.** Consider that the current benchmark might be the current best performance of a human operator or the composed performance of the human-machine team. Where humans and machines interact, it is important to measure the aggregate performance of the team rather than the AI system alone. [360]
5. **Reliability and Robustness:** Various kinds of AI systems often demonstrate impressive performance on average, but can fail in ways that are unexpected in any specific instance. The performance potential of an AI system is often roughly determined by experiment and test, rather than by

---

an added challenge when different cultural and policy norms are involved when collaborating on development and use with allies.

[358] Examples of tools available to assist in assessing and mitigating bias in systems relying on machine learning include Aequitas by the University of Chicago, Fairlearn by Microsoft, AI Fairness 360 by IBM, and PAIR and ML-fairness-gym by Google.

[359] See Microsoft's AI Fairness checklist as an example of an industry tool to support fairness assessments, Michael A. Madaio et al., *Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI*, CHI 2020 (Apr. 25-30, 2020), http://www.jennwv.com/papers/checklists.pdf [hereinafter Madaio, Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI].

[360] Kamar, Combining Human and Machine Intelligence in Large-scale Crowdsourcing.

any predictive analytics. AI can have blinds spots and unknown fragilities.[361] Focus on tools and techniques to carefully bound assumptions of robustness of the AI component in the larger system architecture, and provide sustained attention to characterizing the actual performance envelope for nominal and off-nominal conditions throughout development and deployment. [362]

6. **For systems of systems, testing machine-machine/multi-agent interaction**. Individual AI systems will be combined in various ways in an enterprise to accomplish broader missions beyond the scope of any single system. For example, pipelines of AI systems will exist where the output of one system serves as the input for another AI system. (The output of a track management and classifier system might be input to a target prioritization system which might in turn provide input to a weapon/target pairing tool.) Multiple relatively independent AI systems can be viewed as distinct agents interacting in the environment of the system of systems, and some of these agents will be humans in and on the loop. Industry has encountered and documented problems in building 'systems of systems' out of multiple AI systems[363] A related problem is poor backward compatibility when the performance of one model in a pipeline is enhanced and may result in degrading the overall system of system behavior.[364] These problems in composition illustrate emergent performance, as described in the conceptual overview portion of this section.

   A frequent cause of failures in composed systems is the violation of assumptions that were not previously challenged; therefore, a priority during testing should be to challenge ("stress test") interfaces and usage patterns with boundary conditions and challenges to assumptions about the operational environment and use. This is focused on both unintended violations of assumptions from system composition and also deliberate challenges to the system by adversarial attacks.

B. **Maintenance and deployment**

   Given the dynamic nature of AI systems, recommended practices for maintenance are also critically important. These include:

   1. **Specifying maintenance requirements** for datasets as well as for systems, given that their performance can degrade over time.[365]

---

[361] John Launchbury, *A DARPA Perspective on Artificial Intelligence*, DARPA, (last accessed June 18, 2020), https://www.darpa.mil/about-us/darpa-perspective-on-ai (noting that machine learning is "statistically impressive, but individually unreliable").

[362] Shneiderman, Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy.

[363] One example is "Hidden Feedback Loops", where systems that learn from external world behavior may also shape the behavior they are monitoring. See Sculley, Machine Learning: The High Interest Credit Card of Technical Debt. See also Cynthia Dwork, et al., *Individual Fairness in Pipelines*, (apr. 12, 2020), https://arxiv.org/abs/2004.05167; Megha Srivastava, et al., *An Empirical Analysis of Backward Compatibility in Machine Learning Systems*, KDD '20 (forthcoming, August 2020) [hereinafter Srivastava, An Empirical Analysis of Backward Compatibility in Machine Learning Systems].

[364] Srivastava, An Empirical Analysis of Backward Compatibility in Machine Learning Systems.

[365] *Artificial Intelligence (AI) Playbook for the U.S. Federal Government*, Artificial Intelligence Working Group, ACT-IAC Emerging Technology Community of Interest, (January 22, 2020),

2. **Continuously monitoring and evaluating AI system performance**, including the use of high-fidelity traces to determine continuously if a system is going outside of acceptable parameters (including operational performance measures and established constraints for fairness and core values), both during pre-deployment and operation.[366] This includes measuring system performance per acceptable parameters in terms of both reliability and values.[367] It also includes assessing statistical results for performance over time, for example, to detect emergent bias or anomalies.[368]
3. **Iterative and sustained testing and validation**. Be wary that training and testing that provide characteristics on capabilities might not transfer or generalize to specific settings of usage (for example lighting conditions in some applications may be very different for scene interpretation); thus, testing and validation may need to be done recurrently, and at strategic intervention points, but especially for new deployments and classes of task.[369]
4. **Monitoring and mitigating emergent behavior**. There will be instances where systems are composed in ways not anticipated by the developers (e.g., opportunistic integration with an ally's system). These use cases clearly can't be adequately addressed at development time; some aspects of confidence in the composition must be shifted to monitoring the actual performance of the composed system and its components. For

---

https://www.actiac.org/act-iac-white-paper-artificial-intelligence-playbook.

[366] Beyond accuracy, high-fidelity traces capture other parameters of interest/musts, including fairness, fragility (e.g. whether a system degrades gracefully versus unexpectedly fails), security/attack resilience, and privacy leakage. Often instrumentation results from execution are treated as time-series data and can be analyzed by a variety of anomaly detection techniques to identify unexpected or changing characteristics of system performance. See Meir Toledano et al., *Real-Time Anomaly Detection System for Time Series at Scale*, KDD 2017: Workshop on Anomaly Detection in Finance (2017), http://proceedings.mlr.press/v71/toledano18a/toledano18a.pdf. DOD recently updated its acquisition processes to improve "the ability to deliver warfighting capability at the speed of relevance" See *DoD 5000 Series Acquisition Policy Transformation Handbook*, Department of Defense (Jan. 15, 2020), https://www.acq.osd.mil/ae/assets/docs/DoD%205000%20Series%20Handbook%20(15Jan2020).pdf. These include revised policies for acquiring software-intensive systems and components. Relevant here, program managers are now required to "ensure that software teams use iterative and incremental software development methodologies," and use modern technologies "to achieve automated testing, continuous integration and continuous delivery of user capabilities, frequent user feedback/engagement (at every iteration if possible), security and authorization processes, and continuous runtime monitoring of operational software" Ellen Lord, *Software Acquisition Pathway Interim Policy and Procedures*, Memorandum from the Undersecretary of Defense, to Joint Chiefs of Staff and Department of Defense Staff (Jan. 3, 2020), https://www.acq.osd.mil/ae/assets/docs/USA002825-19%20Signed%20Memo%20(Software).pdf. See also Ori Cohen, *Monitor! Stop Being A Blind Data-Scientist* (Oct. 8, 2019), https://towardsdatascience.com/monitor-stop-being-a-blind-data-scientist-ac915286075f; Mace, Pivot Tracing.

[367] Values parameters could include pre-determined thresholds for acceptable false positive or false negative rates for fairness, or parameters set regarding data or model leakage in the context of privacy.

[368] Lehman, Evolutionary Computation and AI Safety.

[369] Eric Breck, et al., *The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction*, 2017 IEEE International Conference on Big Data, (Dec. 11-14, 2017), https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8258038&tag=1.

emergent performance concerns when AI systems are composed, there are advances in runtime assurance/verification[370] and feature interaction management[371] that can be adapted.

(4) Recommendations for Future Action

- Future R&D is needed to advance capabilities for:
    - Testing, Evaluation, Verification, and Validation (TEVV) of AI systems - to develop a better understanding of how to conduct TEVV and build checks and balances into an AI system. Includes complex system testing - to increase our understanding of and ability to have confidence in emergent performance of composed AI systems. Improved methods are needed to understand, predict, and control systems-of-systems so that when AI systems interact with each other, their interaction does not lead to unexpected negative outcomes.
    - Multi-agent scenario understanding - to advance the understanding of interacting AI systems, including the application of game theory to varied and complex scenarios, and interactions between cohorts composed of a mixture of humans and AI technologies.
- Basic definitional work has been ongoing for years on how to characterize key properties such as fairness and explainability. Progress on a common understanding of the concepts and requirements is critical for progress in widely used metrics for performance.
- Significant work is needed to establish what appropriate metrics should be to assess system performance across attributes for responsible AI and across profiles for particular applications/contexts. (Such attributes, for example, include fairness, interpretability, reliability and robustness.)
- International collaboration and cooperation is needed to:
    - Align on how to test and verify AI system reliability and performance along shared values (such as fairness and privacy). Establishing how to test systems will include measures of performance based on common standards, and may have implications for the types of traceability that will need to be incorporated into system design and development.

---

[370] Shuvendu Lahiri, et al., *Runtime Verification*, 17th International Conference on Runtime Verification (Sept. 13-16, 2017), https://link.springer.com/book/10.1007/978-3-319-67531-2; Christian Colombo ,et al., *Runtime Verification*, 18th International Conference on Runtime Verification (Nov. 10-13, 2018), https://link.springer.com/book/10.1007/978-3-030-03769-7; Sanjit A. Seshia, *Compositional Verification without Compositional Specification for Learning-Based Systems*, UC Berkeley (Nov. 26, 2017), https://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-164.pdf.

[371] Larissa Rocha Soares, et al., *Feature Interaction in Software Product Line Engineering: A Systematic Mapping Study*, Information and Software Technology at 44-58 (June 2018), https://www.sciencedirect.com/science/article/abs/pii/S0950584917302690; Seregy Kolesnikov, *Feature Interactions in Configurable Software Systems*, Universität Passau (Aug. 2019), https://www.researchgate.net/publication/334926566_Feature_Interactions_in_Configurable_Software_Systems; Bryan Muscedere, et al., *Detecting Feature-Interaction Symptoms in Automotive Software using Lightweight Analysis*, 2019 IEEE 26th International Conference on Software Analysis, Evolution and Reengineering at 175-185 (2019), https://ieeexplore.ieee.org/document/8668042.

> Such collaboration on common testing for reliability and adherence to values will be critical among allies and partners to enable interoperability and trust. Additionally, these efforts could potentially include dialogues between the United States and strategic competitors regarding establishing common standards of AI safety and reliability testing in order to reduce the chances of inadvertent escalation.[372]

IV. Human-AI Interaction

(1) Overview

Responsible AI development and fielding requires striking the right balance of leveraging human and AI reasoning, recommendation, and decision-making processes. Ultimately, all AI systems will have some degree of human-AI interaction as they will all be developed to support humans. In some settings, the best outcomes will be achieved when AI is designed to augment human intellect, or to support human-AI collaboration more generally. In other settings, however, time-criticality and the nature of tasks may make some aspects of human-AI interaction difficult or suboptimal.[373] Where the human role is critical in real-time decisions because it is more appropriate, valuable, or designated as such by our values, AI should be intentionally designed to effectively augment and support human understanding, decision making, and intellect. Sustained attention must be focused on optimizing the desired human-machine interaction throughout the AI system lifecycle. It is important to think through the use criteria that are most relevant depending on the model. Models are different for human-assisted AI decision-making, AI-assisted human decision-making, pure AI decision-making, and AI-assisted machine decision-making.

(2) Examples of Current Challenges

There is an opportunity to develop AI systems to complement and augment human understanding, decision making, and capabilities. Decisions about developing and fielding AI systems aimed at specific domains or scenarios should consider the relative strengths of AI capabilities and human intellect across expected distributions of tasks, considering AI system maturity or capability and how people and machines might coordinate.

---

[372] For research regarding common interests in ensuring safety-critical systems work as intended (e.g. in a reliable manner) to avoid destabilization/escalatory dynamics, see Andrew Imbrie & Elsa Kania, *AI Safety, Security, and Stability Among Great Powers Options, Challenges, and Lessons Learned for Pragmatic Engagement,* CSET, (Dec. 2019), https://cset.georgetown.edu/wp-content/uploads/AI-Safety-Security-and-Stability-Among-the-Great-Powers.pdf.

[373] The need for striking the right balance of human involvement in situations of time criticality is not unique to AI. For instance, DoD systems dating back to the 80s have been designed to react to airborne threats at speeds faster than a human would be capable of. See *MK 15 - Phalanx Close-In Weapons System (CIWS)*, U.S. Navy (last accessed June 18, 2020), https://www.public.navy.mil/surfor/Pages/Phalanx-CIWS.aspx.

Designs and methods for human-AI interaction can be employed to enhance human-AI teaming.[374] Methods in support of effective human-AI interaction can help AI systems to understand when and how to engage humans for assistance, when AI systems should take initiative to assist human operators, and, more generally, how to support the creation of effective human-AI teams. In engaging with end users, it may be important for AI systems to infer and share with end users well-calibrated levels of confidence about their inferences, so as to provide human operators with an ability to weigh the importance of machine output or pause to consider details behind a recommendation more carefully. Methods, representations, and machinery can be employed to provide insight about AI inferences, including the use of interpretable machine learning.[375] Research directions include developing and fielding machinery aimed at reasoning about human strengths and weaknesses, such as recognizing and responding to the potential for costly human biases of judgment and decision making in specific settings.[376] Other work centers on mechanisms that consider the ideal mix of initiatives, including when and how to rely on human expertise versus on AI inferences.[377] As part of effective teaming, AI systems can be endowed with the ability to detect the focus of attention, workload, and interruptability of human operators and consider these inferences in decisions about when and how to engage with the operators.[378] Directions of effort include developing mechanisms for identifying the most relevant information or inferences to provide end users of different skills in different settings.[379] Consideration must be given to the prospect introducing bias, including potential biases that may arise because of the configuration and sequencing of rendered data. For example, IC research[380] shows

---

[374] Saleema Amershi, et al., *Guidelines for Human-AI Interaction*, Proceedings of the CHI Conference on Human Factors in Computing Systems (2019), https://dl.acm.org/doi/10.1145/3290605.3300233

[375] Rich Caruana, et al., Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission, Semantic Scholar (Aug. 2015), https://www.semanticscholar.org/paper/Intelligible-Models-for-HealthCare%3A-Predicting-Risk-Caruana-Lou/cb030975a3dbcdf52a01cbd1c140711332313e13.

[376] Eric Horvitz, *Reflections on Challenges and Promises of Mixed-Initiative Interaction*, AAAI Magazine 28 Special Issue on Mixed-Initiative Assistants (2007), http://erichorvitz.com/mixed_initiative_reflections.pdf.

[377] Eric Horvitz, *Principles of Mixed-Initiative User Interfaces*, Proceedings of CHI '99 ACM SIGCHI Conference on Human Factors in Computing Systems (May 1999), https://dl.acm.org/doi/10.1145/302979.303030; Kamar, Combining Human and Machine Intelligence in Large-scale Crowdsourcing.

[378] Eric Horvitz, et al., *Models of Attention in Computing and Communications: From Principles to Applications*, Communications of the ACM 46(3) at 52-59 (Mar. 2003), https://cacm.acm.org/magazines/2003/3/6879-models-of-attention-in-computing-and-communication/fulltext.

[379] Eric Horvitz & Matthew Barry, *Display of Information for Time-Critical Decision Making*, Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (Aug. 1995), https://arxiv.org/pdf/1302.4959.pdf.

[380] There has been considerable research in the IC on the challenges of confirmation bias for analysts. Some experiments demonstrated a strong effect that the sequence in which information is presented alone can shape analyst interpretations and hypotheses. Brant Cheikes, et al., *Confirmation Bias in Complex Analyses*, MITRE (Oct. 2004), https://www.mitre.org/sites/default/files/pdf/04_0985.pdf. This highlights the care that is required when designing the human machine teaming when complex, critical, and potentially ambiguous information is presented to analysts and decision makers.

that confirmation bias can be triggered by the order in which information is displayed, and this order can consequently impact or sway intel analyst decisions. Careful design and study can help to identify and mitigate such bias.

(3) Recommendations for Adoption

Critical practices to ensure optimal human-AI interaction are described in the non-exhaustive list below. These recommended practices span the entire AI lifecycle.

Human-AI Interaction Recommended Practices

A. **Identification of functions of human in design, engineering, and fielding of AI**
   1. **Define functions and responsibilities of human operators and assign them to specific individuals**. Functions will vary for each domain and each project within a domain; they should be periodically revisited as model maturity and human expertise evolve over time.
   2. **Given the nature of the mission and current competencies of AI, policies should define the tasks of humans across the AI lifecycle,** noting needs for feedback loops, including opportunities for oversight.
   3. **Enable feedback and oversight to ensure that systems operate as they should -** algorithmic accountability means that there is a governance structure in place to correct grievances if systems fail.

B. **Explicit support of human-AI interaction and collaboration**
   1. **Human-AI design guidelines**. AI systems designs should take into consideration the defined tasks of humans in human-AI collaborations in different scenarios; ensure the mix of human-machine actions in the aggregate is consistent with the intended behavior, and accounting for the ways that human and machine behavior can co-evolve;[381] and also avoid automation bias and unjustified reliance on humans in the loop as failsafe mechanisms. Allow for auditing of the human-AI pair, not only the AI in isolation, which could be a secondary expert examining a subset of cases. Designs should be transparent (e.g., about why and how a system did what it did, system updates, or new capabilities) so that there is an understanding the AI is working day-to-day and to allow for an audit trail if things go wrong .[382] Based on context and mission need, designs should ensure usability of AI systems by AI experts, domain experts, and novices, as appropriate.[383] Both transparency and usability will depend on the audience.

---

[381] Patricia L. McDermott et al., *Human-machine Teaming Systems Engineering Guide*, MITRE (Dec. 2018), https://www.mitre.org/publications/technical-papers/human-machine-teaming-systems-engineering-guide; Shneiderman, Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy.

[382] For additional examples, see *Guidelines for Human AI Interaction*, Microsoft (June 4, 2019), https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/.

[383] Systems are sometimes designed with the assumption of a human in the loop as the failsafe or interlock, but humans often defer to computer generated results and get in the habit of confirming machine results without scrutiny.

2. **Algorithms and functions in support of interpretability and explanation.** Algorithms and functions that provide individuals with task-relevant knowledge and understanding need to take into consideration that key factors in an AI system's inferences and actions can be understood differently by various audiences. These audiences span real-time operators who need to understand inferences and recommendations for decision support, engineers and data scientists involved in developing and debugging systems, and other stakeholders including those involved in oversight. Interpretability and explainability exists in degrees; what's needed in terms of explainability will depend on who is receiving the explanation, what the context is, and the amount of time available to deliver and process this explanation. In this regard, interpretability intersects with traceability, audit, and documentation practices.
3. **Designs that provide cues to the human operator(s) about the level of confidence the system has in the results or behaviors of the system**.[384] AI system designs should appropriately convey uncertainty and error bounding. For instance, a user interface should convey system self-assessment of confidence alerts when the operational environment is significantly different from the environment the system was trained for, and indicate internal inconsistencies that call for caution.
4. **Policies for machine-human initiative and handoff.** Policies, and aspects of human computer interaction, system interface, and operational design, should define when and how information or tasks should be handed off from a machine to a human operator and vice versa. Include checks to continually evaluate whether distribution of tasks is working. Special attention should be given to the fact that humans may freeze during an unexpected handoff due to the processing time the brain needs, potential distractions, or the condition during which the handoff occurs. The same may be true with an AI system which may not fully understand the human's intent during the handoff and may consequently make unexpected actions.
5. **Leveraging traceability to assist with system development and understanding**. Traceability processes must include audit logs or other traceability mechanisms to retroactively understand if something went wrong, and why, in order to improve systems and their use in the future and for redress. Infrastructure and instrumentation[385] can also help assess humans, systems, and environments to gauge the impact of AI at all levels of system maturity; and to measure the effectiveness and performance for hybrid human-AI systems in a mission context.

---

[384] When systems report confidence in probabilities of correctness, these should be well calibrated. At the same time, it is important to acknowledge that there are limits to the confidence that can be assigned to a system estimate of correctness.

[385] Infrastructure includes tools (hardware and software) in the test environment that support monitoring system performance (such as the timing of exchanges among systems, or the ability to generate test data). Instrumentation refers to the presence of monitoring and additional interfaces to provide insight into a specific system under test.

6. **Training**. Train and educate individuals responsible for AI development and fielding, including human operators, decision makers, and procurement officers. Training should include experiences with use of systems in realistic situations. Beyond training in the specifics of the system and application, operators of systems with AI components, especially systems that perform classification or pattern recognition, should receive education that includes fundamentals of AI and data science, including coverage of key descriptors of performance, including rates of false negatives and false positives, precision and recall, and sensitivity and specificity.

   Periodic certification and refresh. In addition to initial programs of training, operators should receive ongoing refresher trainings. Beyond being scheduled periodically, refresher trainings are appropriate when systems are deployed in new settings and unfamiliar scenarios. Refresh on training is also needed when predictive models are revised with new or additional data as the performance of systems may shift with such updates introducing behaviors that are unfamiliar to human operators.[386]

(4) Recommendations for Future Action

- Future R&D is needed to advance capabilities for:
  - Enhanced human-AI interaction -
    - To progress the ability of AI technologies to perceive and understand the meaning of human communication, including spoken speech, written text, and gestures. This research should account for varying languages and cultures, with special attention to diversity given that AI typically performs worse in cases with gender and racial minorities.
    - To improve human-machine teaming. This should include disciplines and technologies centered on decision sciences, control theory, psychology, economics (human aspects and incentives), and human factors engineering, such as human-AI interfaces**,** to enhance situational awareness and make it easier for users to do their work. Human-AI interaction and the mechanisms and interfaces that support such interactions, including richer human-AI *collaborations*, will depend upon mission needs and appropriate degrees of autonomy versus human oversight and control. R&D for human-machine teaming should also focus on helping systems understand human blind spots and biases, and optimizing factors such as human attention, human workload, ideal mixing of human and machine initiatives, and passing control between the human and machine. For effective passing of control, and to

[386] Gagan Bansal et al., *Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff*, AAAI (Jul. 2019), https://www.aaai.org/ojs/index.php/AAAI/article/view/4087.

have effective and trusted teaming, R&D should further enable humans and machines to better understand intent and context of handoff.

- Ongoing work is needed to train the workforce that will interact with, collaborate with, and be supported by AI systems. In its First Quarter Recommendations, the Commission provided recommendations for such training.[387]
    - **Workforce training**. A complementary best practice for Human-AI Interaction is training the workforce to understand tools they're using; as AI gets democratized, it will also get misused. For probabilistic systems, concepts and ideas that are important in system operation should be understood; for operators this includes understanding concepts such as precision, recall, sensitivity and specificity, and ensuring operators know how to interpret the confidence in inferences that well-calibrated systems convey.

V. Accountability and Governance

(1) Overview

National security departments and agencies must specify who will be held accountable for both specific system outcomes and general system maintenance and auditing, in what way, and for what purpose. Government must address the difficulties in preserving human accountability, including for end users, developers, testers, and the organizations employing AI systems. End users and those ultimately affected by the actions of an AI system should be offered the opportunity to appeal an AI system's determinations. And, finally, accountability and appellate processes must exist not only for AI decisions, but also for AI system inferences, recommendations, and actions.

(2) Examples of Current Challenges

Overseeing entities must have the technological capacity to understand what in the AI system caused the contentious outcome. For example, if a soldier uses an AI-enabled weapon and the result violates international law of war standards, an investigating body or military tribunal should be able to re-create what happened through auditing trails and other documentation. Without policies requiring such technology and the enforcement of those policies, proper accountability would be elusive if not impossible. Moreover, auditing trails and documentation will prove critical as courts begin to grapple with whether AI system's determinations reach the requisite standards to be admitted as evidence.[388] Building the traceability infrastructure to permit auditing (as described in the Engineering Practices section)

---

[387] See *First Quarter Recommendations*, NSCAI (Mar. 2020), https://www.nscai.gov/reports.

[388] For more on the difficulties of admitting ML evidence, see Patrick Nutter, *Machine Learning Evidence: Admissibility and Weight*, University of Pennsylvania Law (Feb. 2019), https://scholarship.law.upenn.edu/jcl/vol21/iss3/8/.

will increase the costs of building AI systems and take significant work -- a necessary investment given our commitment to accountability, discoverability, and legal compliance.

(3) Recommendations for Adoption

Critical accountability and governance practices are identified in the non-exhaustive list below.

Accountability and Governance Recommended Practices

1. **Identify responsible actors.** Determine and document the human beings accountable for a specific AI system or any given part of an AI system and the processes involved with it. This includes identifying persons responsible for the operation of an AI system including the system's inferences, recommendations, and actions during usage, as well as the enforcement of policies for using a system. Determine and document the mechanism/structure for holding such actors accountable and to whom should that mechanism/structure be disclosed to ensure proper oversight.
2. **Adopt technology to strengthen accountability processes and goals**. Document the chains of custody and command involved in developing and fielding AI systems. This will allow the government to know who was responsible at which point in time. Improving traceability and auditability capabilities will allow agencies to better track a system's performance and outcomes. [389]
3. **Adopt policies to strengthen accountability.** Identify or, if lacking, establish policies that allow individuals to raise concerns about irresponsible AI, e.g. via an ombudsman. Agencies should institute specific oversight and enforcement practices, including: auditing and reporting requirements, a mechanism that would allow thorough review of the most sensitive/high-risk AI systems to ensure auditability and compliance with other responsible use and fielding requirements, an appealable process for those who have been found at fault of developing or using AI irresponsibly, and grievance processes for those affected by the actions of AI systems. Agencies should leverage best practices from academia and industry for conducting internal audits and assessments,[390] while also acknowledging the benefits offered by external audits.[391]

---

[389] See Raji, Closing the AI Accountability Gap.

[390] See Raji, Closing the AI Accountability Gap ("In this paper, we present internal algorithmic audits as a mechanism to check that the engineering processes involved in AI system creation and deployment meet declared ethical expectations and standards, such as organizational AI principles"); see also Madaio, Co-Designing Checklists to Understand Organizational Challenges and Opportunities Around Fairness in AI.

[391] For more on the benefits of external audits, see Brundage, Toward Trustworthy AI Development. For an agency example, see Aaron Boyd, *CBP Is Upgrading to a New Facial Recognition Algorithm in March*, Nextgov.com (Feb. 7, 2020), https://www.nextgov.com/emerging-tech/2020/02/cbp-upgrading-

4. **External oversight support**. Remain responsive and facilitate Congressional oversight through documentation processes and other policy decisions.[392] For instance, supporting traceability and specifically documentation to audit trails, will allow for external oversight.[393] Internal self-assessment alone might prove to be inadequate in all scenarios.[394] Congress can provide a key oversight function throughout the AI lifecycle, asking critical questions of agency leadership and those responsible for AI systems.

(4) Recommendations for Future Action

- Currently no external oversight mechanism exists specific to AI in national security. Notwithstanding the important work of Inspectors General in conducting internal oversight, open questions remain as to how to complement current practices and structures.

---

new-facial-recognition-algorithm-march/162959/ (highlighting a NIST algorithmic assessment on behalf of U.S. Customs and Border Protection).

[392] Maranke Wieringa, *What to Account for When Accounting for Algorithms*, Proceedings of the 2020 ACM FAT Conference, (Jan. 2020), https://dl.acm.org/doi/10.1145/3351095.3372833.

[393] Raji, Closing the AI Accountability Gap.

[394] Brundage, Toward Trustworthy AI Development.

Appendix A-3 — DoD AI Principles Alignment Table

NSCAI staff developed the below table to illustrate how U.S. government AI ethics principles, like those recently issued by the DoD, can be operationalized through NSCAI's Key Considerations for Responsible Development and Fielding of AI (See Appendix A-1 and A-2). Other Federal agencies and departments can use this table to visualize how NSCAI's recommended practices align with their own AI principles, or as guidance in the absence of internal AI ethics principles. In the table below, an "X" indicates that the NSCAI recommended practice on the left operationalizes the DoD principle at the top. As the table shows, every NSCAI key consideration recommended practice implements one or more DOD AI ethics principles. And every DoD AI ethics principle has at least one Key Considerations Recommended Practice that implements the principle.

NSCAI Recommended Practices:		DOD PRINCIPLES OF AI ETHICS				
		Responsible	Equitable	Traceable	Reliable	Governable
Core Values	A1 - Employ technologies and operational policies for privacy, fairness, inclusion, human rights, and law of armed conflict	X	X			X
	B1 - Consider and document value considerations based on how tradeoffs with accuracy are handled	X	X	X	X	
	B2 - Consider and document value considerations in systems that rely on representations of objective or utility functions	X	X	X	X	
	B3 - Conduct documentation, reviews, and set limits on disallowed outcomes	X	X	X	X	X
Engineering	1 - Concept of operations development, and design and requirements definition and analysis	X	X	X	X	X
	2 - Documentation of the AI lifecycle			X		
	3 - Infrastructure to support traceability, including auditability and forensics		X	X		
	4 - Security and robustness: addressing intentional and unintentional failures				X	X
	5 - Conduct red-teaming				X	
System Performance	A1 - Standards for metrics & reporting	X	X	X	X	
	A2 - Representativeness of data and model for the specific context at hand	X	X	X	X	
	A3 - Evaluating an AI system's performance relative to current benchmarks			X		X
	A4 - Evaluating aggregate performance of human-machine teams	X				
	A5 - Reliability and robustness	X		X	X	
	A6 - For systems of systems, testing machine-machine/multi-agent interaction	X			X	
	B1 - Specifying maintenance requirements	X	X	X	X	
	B2 - Continuously monitoring and evaluating AI system performance	X	X	X	X	X
	B3 - Iterative and sustained testing and validation	X			X	X
	B4 - Monitoring and mitigating emergent behavior	X			X	X
Human-AI Interaction	A1 - Define functions and responsibilities of human operators and assign them to specific individuals	X		X		
	A2 - Policies should define the tasks of humans across the AI lifecycle	X				
	A3 - Enable feedback and oversight to ensure that systems operate as they should	X			X	
	B1 - Human-AI design guidelines	X	X	X		X
	B2 - Algorithms and functions in support of interpretability and explanation	X		X		X
	B3 - Designs that provide cues to human operator(s) about the confidence a system has in its results or behaviors	X		X		X
	B4 - Policies for machine-human handoff	X		X		X
	B5 - Leveraging traceability to assist with system development and understanding	X		X	X	X
	B6 - Training	X	X	X	X	X
Accountability/ Governance	1 - Identify responsible actors	X		X		X
	2 - Adopt technology to strengthen accountability processes and goals	X		X		X
	3 - Adopt policies to strengthen accountability	X		X		
	4 - External oversight support			X		

Appendix B — Draft Proposed Executive Order on Applying Export Control and Investment Screening Mechanisms to Artificial Intelligence and Related Technologies

By the authority vested in me as President by the Constitution and the laws of the United States of America, and in order to promote U.S. innovation and leadership in emerging and foundational technologies while protecting U.S. national security, it is hereby ordered as follows:

Section 1. *Policy.* It is the policy of the United States that export controls and investment screening mechanisms must be used in targeted, clearly defined, and strategic ways to protect U.S. national security, in pursuit of the broader policy of promoting U.S. innovation and leadership in emerging and foundational technologies, to include dual-use technologies such as artificial intelligence (AI).

The United States must be tailored and discrete in implementing export controls on general purpose and dual-use technologies, such as AI. To ensure maximum effectiveness and minimize the adverse impact on U.S. industry, the United States Government should be guided by the following principles:

(1) Principle One: Export Controls Cannot Supplant Investment and Innovation. Technology protection policies are intended to slow U.S. competitors' pursuit and development of key strategic technologies for national security purposes, not stop them in their tracks. The United States must cultivate investment in these technologies through direct federal funding or changes to the regulatory environment in order to preserve existing U.S. advantages.

(2) Principle Two: U.S. Promote and Protect Strategies Must Be Integrated. The U.S. strategy to protect emerging technologies, including but not limited to AI, must be integrated with targeted efforts to promote U.S. leadership in such technologies. When choosing to implement controls, the United States should simultaneously consider policies to spur domestic research and development (R&D) in key industries to partially offset the resulting costs to U.S. firms, create alternative global markets, or encourage new investment to strengthen the U.S. industrial position.

(3) Principle Three: Export Controls Must Be Targeted, Strategic, and Coordinated with Allies. In devising new export controls on widespread and dual-use technologies such as AI, the United States must be careful and

selective in the implementation of export controls. To ensure maximum effectiveness and minimize the adverse impact on U.S. industry, the United States Government should be guided by the following three-part test:

a. Export controls must be targeted, clearly defined, and focused on choke points where they will have a strategic impact on the national security capabilities of competitors, but smaller repercussions on U.S. industry.

b. Export controls must have a clear strategic objective, seeking to deter competitors from pursuing paths that endanger U.S. national security interests, and account for the projected cost and timeframe for competitors to create a domestic alternative.

c. Export controls must be coordinated with key U.S. allies which are also capable of producing the given technology, in order to effectively restrict the supply to adversaries and also prevent circumstances where unilateral controls cut off U.S. market access but competitors are able to purchase the same technology from other countries.

(4) Principle Four: Use Discrete Export Controls, But Broader Investment Screening. While broad and sweeping export controls on AI and other dual-use emerging technologies could result in significant blowback on U.S. industry, which would harm overall U.S. strategic competitiveness, investment screening presents opportunities to take a more proactive regulatory approach while minimizing risk to U.S. industry. Provided the United States can continue approving benign transactions expeditiously, enhancing investment screening presents significant potential to blunt concerning transfers of technology.

Section 2. *Objective.* In 2018, the Congress enacted the Export Control Reform Act of 2018 (ECRA) and the Foreign Investment Risk Reduction Modernization Act of 2018 (FIRRMA) to provide the United States Government with additional mechanisms to control exports and screen investments. The United States Government must take steps to provide the private sector and foreign governments with clarity about the application of these laws to emerging and foundational technologies and enhance U.S. national security in the process.

Section 3. *Establishment of Interagency Task Force on Emerging and Foundational Technologies.*

(a) Pursuant to Section 1758 of the Export Control Reform Act of 2018 (ECRA), there is hereby established an Interagency Task Force on Emerging and Foundational Technologies (Task Force) to identify emerging and foundational technologies that are essential to the national security of the United States and are not critical technologies described in clauses (i) through (v) of 50 U.S.C. 4565(a)(6)(A).

(b) The Task Force shall be chaired by the Secretary of Commerce (Chair) and consist of senior-level officials from the following executive departments and agencies (agencies) designated by the heads of those agencies:

(i) Department of State;

(ii) Department of the Treasury;

(iii) Department of Defense;

(iv) Department of Energy; and

(vi) such other agencies as the President, or the Chair, may designate.

(c) The Chair shall designate a senior-level official of the Department of Commerce as the Executive Director of the Task Force, who shall be responsible for regularly convening and presiding over the meetings of the Task Force, determining its agenda, and guiding its work in fulfilling its functions under this Order, in coordination with the Bureau of Industry and Security (BIS) at the Department of Commerce.

Section 4. *Functions of the Task Force.*

(a) The Task Force shall meet regularly to identify emerging and foundational technologies that are essential to the national security of the United States for purposes of establishing export controls and investment screening mechanisms, as appropriate, related to those technologies.

(b) Within 120 days, the Task Force shall finalize lists of emerging and foundational technologies pursuant to section 1758 of ECRA. The Secretary of Commerce shall thereafter issue proposed rules on emerging and foundational technologies and proceed expeditiously to issue final rules at the conclusion of the notice and comment period.

(c) The Task Force shall review the lists of emerging and foundational technologies and issue amendments as needed on no less than an annual basis.

Section 5. *Process for Identifying Emerging and Foundational Technologies.*

(a) In identifying emerging and foundational technologies pursuant to this Order, the Task Force shall consider information from multiple sources, including:

(i) publicly available information;

(ii) classified information, including relevant information provided by the Director of National Intelligence;

(iii) information relating to reviews and investigations of transactions by the Committee on Foreign Investment in the United States under 50 U.S.C. 4565; and

(iv) information provided by the advisory committees established by the Secretary to advise the Under Secretary of Commerce for Industry and Security on controls under the Export Administration Regulations, including the Emerging Technology Technical Advisory Committee.

(b) In identifying emerging and foundational technologies pursuant to this Order, the Task Force shall take into account:

(i) the development of emerging and foundational technologies in foreign countries;

(ii) the effect export controls imposed pursuant to this section may have on the development of such technologies in the United States;

(iii) the effectiveness of export controls imposed pursuant to this section on limiting the proliferation of emerging and foundational technologies to foreign countries; and

(iv) the policy and principles reflected in section 1 of this Order.

Section 6. *Improving Coordination with Expert Advisory Groups.*

(a) The Secretary of Commerce shall review existing technical advisory committees (TACs) at the Department of Commerce, including the Emerging Technology Technical Advisory Committee (ETTAC), to ensure that each TAC is comprised of members from industry and academia with deep subject matter expertise to assess the need for export controls for emerging and foundational technologies.

(b) The Secretary of Commerce, as Chair of the Task Force, shall ensure that the Task Force has solicited and received feedback from the ETTAC and other relevant TACs at the Department of Commerce on the text of any proposed or final rule on emerging or foundational technologies, prior to issuance of such rule.

(c) The Secretary of Commerce shall ensure that senior officials at the Departments of State and the Treasury are granted non-voting observer access at all ETTAC meetings.

Section 7. *Improving International Coordination on Export Controls on Semiconductor Manufacturing Equipment.* Within 180 days, the Secretary of State, in consultation with the Secretary of Commerce and the Secretary of Defense, shall host a multilateral engagement with senior-level representatives of Japan, the Netherlands, and if deemed appropriate, other U.S. allies and partners that produce semiconductor

manufacturing equipment, including EUV lithography equipment and ArF immersion lithography equipment, listed by the Wassenaar Arrangement or identified by the Task Force. The purpose of this meeting will be to align export licensing policies toward a presumptive denial of export licenses for exports of semiconductor manufacturing equipment to China. The Secretary of State shall provide a report to the President within 60 days of the meeting assessing:

(i) whether U.S. allies and partners are currently exporting such equipment to China;

(ii) what steps each country which manufactures such equipment must take to ensure its regulatory regime is aligned with that of the United States, and its willingness to take those steps; and

(iii) whether additional opportunities exist to strengthen international cooperation on export controls on semiconductor manufacturing equipment which are consistent with the policy and principles reflected in section 1 of this Order.

Section 8. *Engaging Technical Experts for Export Control Review.*

(a) The Secretary of Commerce, in consultation with the Secretaries of the Treasury and Defense, shall establish a network within existing federally funded research and development centers (FFRDCs) and university affiliated research centers (UARCs) to provide technical expertise to all departments and agencies for issues relating to export controls and investment screening related to emerging and foundational technologies. The network shall encompass a regional distribution of FFRDCs and UARCs located in areas of the United States with a concentration of technology expertise in emerging and foundational technologies.

(b) Individuals selected to participate in the network shall provide real-time technical input to all policy discussions on export controls and review of export control license applications, including those of the Task Force, those conducted pursuant to EO 12981 or a successor order, and any other interagency policy discussions pertaining to export controls, as well as the investment screening processes of the Committee on Foreign Investment in the United States (CFIUS).

Section 9. *Automating Export Control and Investment Screening Reviews.* The Secretaries of Commerce and the Treasury shall task the aforementioned network with exploring using AI-based systems to assist in the evaluation of applications for export control licenses and CFIUS filings and shall provide a report to the President on the use of AI-based systems for such purposes within 180 days. This report shall include an evaluation of—

(i) how AI-based systems could assist existing review processes;

(ii) whether incorporating such systems could enhance the accuracy and speed of the review processes;

(iii) whether relevant Departments and Agencies have sufficient quantity and quality of data to train AI-based review systems, and how existing data can be improved;

(iv) what information technology infrastructure inside relevant Departments and Agencies needs to be improved to fully utilize such systems; and

(iv) an approximate timeline and cost for deploying a system or systems, and the projected savings per year in labor-hours once deployed.

Section 10. *General Provisions.*

(a) Nothing in this order shall be construed to impair or otherwise affect:

(i) the authority granted by law, regulation, Executive Order, or Presidential Directive to an executive department, agency, or the head thereof; or

(ii) the functions of the Director of the Office of Management and Budget relating to budgetary, administrative, or legislative proposals.

(b) This order shall be implemented consistent with applicable law and subject to the availability of appropriations.

(c) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

Appendix C — Legislative Language

The below legislative text represents the Commission staff's best effort to capture the Commission's second quarter recommendations. The Commission defers to the House and Senate members, staff, and legislative counsels as to appropriate drafting and policy.

TAB 1 – Legislative Language

Recommendation 4: Expand Section 219 Laboratory Initiated Research Authority funding to support AI infrastructure and software investments at DoD laboratories.

SEC. ___.—MECHANISMS TO PROVIDE FUNDS FOR DEFENSE LABORATORIES FOR EXPANDED INVESTMENTS IN INFRASTRUCTURE AND SOFTWARE ASSETS TO SUPPORT ARTIFICIAL INTELLIGENCE.—

(a) AMENDMENTS TO TITLE 10, UNITED STATES CODE.—

(1) Section 2363 of title 10, United States Code is amended—

(A) In paragraph (a)(1)(D), by striking "infrastructure and equipment" and inserting "infrastructure and equipment, including but not limited to infrastructure and software assets to support AI research, prototyping, and testing,"; and

(B) In paragraph (a)(2), by adding at the end the following:

"Such mechanisms may include the use of a working capital fund in accordance with the requirements of section 2208 of this title."

(2) Section 2805 of title 10, United States Code is amended—

(A) In paragraph (d)(1), by adding a new subparagraph (C), as follows:

"(C) in the case of an investment in infrastructure and software assets to support AI research, prototyping, and testing, up to two times the amounts otherwise applicable under paragraphs (A) and (B)."; and

(B) In paragraph (d)(2), by striking the period and inserting the following:

"(or, in the case of an investment in infrastructure and software assets to support AI research, prototyping, and testing, two times that amount)."

(b) SENSE OF CONGRESS.—It is the Sense of Congress that the Directors of the Defense laboratories should use an amount of funds as close as possible to four percent of all funds available to the defense laboratory for the purposes specified in section 2363 of title 10, United States Code, to enable higher-level dollar investments in infrastructure and software assets to support AI research, prototyping, and testing.

TAB 3 – Legislative Language

Recommendation 1: Create a National Reserve Digital Corps.

SEC. 1. SHORT TITLE.—This Act may be cited as the "National Reserve Digital Corps Act".

SEC. 2. ESTABLISHMENT OF NATIONAL RESERVE DIGITAL CORPS.—

(a) IN GENERAL.—Subpart I of Part III of title 5, United States Code, is amended by inserting after chapter 102 the following new chapter:

CHAPTER 103—NATIONAL RESERVE DIGITAL CORPS
Sec. 10301. Establishment.
Sec. 10302. Definitions.
Sec. 10303. Organization.
Sec. 10304. Work on Behalf of Federal Agencies.
Sec. 10305. Digital Corps Scholarship Program.
Sec. 10306. Duration of Pilot Program.
Sec. 10307. Authorization of Appropriation.

SEC. 10301. ESTABLISHMENT.—For the purposes of attracting, recruiting, and training a core of world-class digital talent to serve the national interest and enable the Federal Government to become a digitally proficient enterprise, there is established within the Office of Management and Budget a pilot program for a civilian National Reserve Digital Corps, whose members shall serve as special government employees, working not fewer than 30 days per year as short-term advisors, instructors, or developers in the Federal Government.

Sec. 10302. DEFINITIONS.—

(a) DIRECTOR.—The term "Director" means the Director of the Office of Management and Budget.

(b) NODE.—The term "node" means a group of persons or team organized under the direction of a node leader to provide digital service to one or more Federal agencies pursuant to an agreement between the Office of Management Budget and each other Federal agency.

(c) NODE LEADER.—The term "node leader" means a full time government employee, as defined by section 2105 of title 5, United States Code, selected under this Act to lead one or more nodes, who reports to the Director or the Director's designee.

(d) NODE MEMBER.—The term "node member" means a special government employee, as defined by section 202 of title 18, United States Code, selected under this Act to work at least 38 days per fiscal year and report to a node leader in furtherance of the mission of a specified node.

Sec. 10303. ORGANIZATION.—

(a) NODES AND NODE LEADERS. —The National Reserve Digital Corps shall be organized into nodes, each of which shall be under the supervision of a node leader.

(b) ADMINISTRATIVE SUPPORT. —The National Reserve Digital Corps shall receive funding and administrative support from the Office of Management and Budget, which shall be responsible for selecting node leaders, establishing standards, ensuring that nodes meet government client requirements, maintaining security clearances, establishing access to an agile development environment and tools, and facilitating appropriate technical exchange meetings.

(c) HIRING AUTHORITY.—

(1) Direct Hiring Authority of Node Members.—The Director of the Office of Management and Budget, on the recommendation of a node leader, may appoint, without regard to the provisions of subchapter I of chapter 33 (other than sections 3303 and 3328 of such chapter), a qualified candidate to a position in the competitive service in the Office of Management and Budget to serve as a node member. This provision shall not preclude the Director from hiring additional employees, including full time government employees, as defined by section 2105 of title 5, United States Code.

(2) Term and Temporary Appointments of Node Members.—The Director of the Office of Management and Budget, on the recommendation of a node leader, may make a noncompetitive temporary appointment or term appointment for a period of not more than 18 months, of a qualified candidate to serve as a node member in a position in the competitive service for which a critical hiring need exists, as determined under section 3304 of title 5, United States Code, without regard to sections 3327 and 3330 of such title.

Sec. 10304. WORK ON BEHALF OF FEDERAL AGENCIES.—

(a) PURPOSE.—Each node shall undertake projects to assist Federal agencies by providing digital education and training, performing data triage and providing acquisition assistance, helping guide digital projects and frame technical solutions, helping build bridges between public needs and private sector capabilities, and related tasks.

(b) AUTHORITIES.—Projects may be undertaken—

(1) on behalf of a Federal agency—

(A) by direct agreement between the Office of Management and Budget and the Federal agency; or

(B) at the direction of the Office of Management and Budget at the request of the Federal agency; or

(2) to address a digital service need encompassing more than one Federal agency—

(A) at the direction of the Office of Management and Budget; or

(B) on the initiative of a node leader.

Sec. 10305. DIGITAL CORPS SCHOLARSHIP PROGRAM.—

(a) IN GENERAL.—The Director shall establish a National Reserve Digital Corps scholarship program to provide full scholarships to competitively selected students who commit to study specific disciplines related to national security digital technology.

(b) SERVICE OBLIGATION. —Each student, prior to commencing the Digital Corps Scholarship Program, shall sign an agreement with respect to the student's commitment to the United States. The agreement shall provide that the student agree to the following:

(1) a commitment to serve as an intern in a Federal agency for at least six weeks during each of the summers before their junior and senior years; and

(2) a commitment to serve in the National Reserve Digital Corps for six years after graduation.

(c) PROGRAM ELEMENTS.—In establishing the program, the Director shall determine the following—

(1) Eligibility standards for program participation;

(2) Criteria for establishing the dollar amount of a scholarship, including tuition, room and board;

(3) Repayment requirements for students who fail to complete their service obligation;

(4) An approach to ensuring that qualified graduates of the program are promptly hired and assigned to node leaders; and

(5) Resources required for the implementation of the program.

(d) CONTINUING EDUCATION.—The Director shall establish a training and continuing education program to fund educational opportunities for members of the National Digital Reserve Corps, including conferences, seminars, degree and certificate granting programs, and other training opportunities that are expected to increase the digital competencies of the participants.

(e) IMPLEMENTATION.—

(1) Not later than six months after the date of the enactment of this Act, the Director shall establish the administrative support function and issue guidance for the National Reserve Digital Corps, which shall include the identification of points of contact for node leaders at Federal agencies.

(2) Not later than one year after the date of the enactment of this Act, the Director shall appoint not fewer than five node leaders under the National Reserve Digital Corps program and authorize the node leaders to begin recruiting reservists and undertaking projects for Federal agencies.

(3) Beginning two years after the date of the enactment of this Act, the Director shall report annually to Congress on the progress of the National Reserve Digital Corps. The Director's report shall address, at a minimum, the following measures of success:

(A) The number of technologists who participate in the National Reserve Digital Corps annually;

(B) Identification of the Federal agencies that submitted work requests, the nature of the work requests, which work requests were assigned a node, and which work requests were completed or remain in progress;

(C) Evaluations of results of National Reserve Digital Corps projects by Federal agencies; and

(D) Evaluations of results of National Reserve Digital Corps projects by reservists.

Sec. 10306. DURATION OF PILOT PROGRAM.—The pilot program under this Act shall terminate no earlier than six years after its commencement.

Sec. 10307. AUTHORIZATION OF APPROPRIATION.—There is authorized to be appropriated $16,000,000 to remain available until fiscal year 2023 the initial administrative cost, including for the salaries and expenses scholarship and education benefits, for the National Digital Reserve Corps.

Recommendation 3: Create a United States Digital Service Academy.

SEC. 1. SHORT TITLE.—This Act may be cited as the "United States Digital Service Academy Act".

SEC. 2. ESTABLISHMENT OF ACADEMY.—

(a) ESTABLISHMENT.—There is established as an independent entity within the Federal Government a United States Digital Service Academy (hereafter referred to as the "ACADEMY"), at a location to be determined, to serve as a federally-funded, accredited, degree-granting university for the instruction of selected individuals in digital technical fields and the preparation of selected individuals for civil service with the Federal Government.

(b) DIGITAL TECHNICAL FIELDS DEFINED.—The term "digital technical fields" includes artificial intelligence, software engineering, electrical science and engineering, computer science, molecular biology, computational biology, biological engineering, cybersecurity, data science, mathematics, physics, human-computer interaction, robotics, and design and any additional fields specified in regulations by the Board.

SEC. 3. ORGANIZATION.—

(a) BOARD OF REGENTS.– The business of the Academy shall be conducted by a Board of Regents (hereafter referred to as the "Board").

(1) COMPOSITION.– The Board shall consist of nine voting members and ex officio members, as set forth in this subsection.

(2) VOTING MEMBERS.—The President shall appoint, by and with the consent of the Senate, nine persons from civilian life who have demonstrated achievement in one or more digital technical fields, higher education administration, or Federal civilian service, to serve as voting members on the Board. Appointment of the first voting members shall be made not later than 180 days after enactment of this Act.

(3) EX OFFICIO MEMBERS.—Ex officio members shall include—

(A) The Secretary State;

(B) The Secretary of Defense;

(C) The Attorney General;

(D) The Secretary of Commerce;

(E) The Secretary of Energy;

(F) The Secretary of Homeland Security;

(G) The Director of National Intelligence;

(H) The Director of the Office of Personnel Management; and

(I) such other Federal Government officials as determined by the President.

(2) TERM OF VOTING MEMBERS.—The term of office of each voting member of the Board shall be six years, except that initial terms shall be staggered at two year intervals and any member appointed to fill a vacancy occurring before the expiration of a term shall be appointed for the remainder of such term.

(3) PRESIDENT OF THE BOARD.—One of the members (other than an ex officio member) shall be designated by the President as Chairman and shall be the presiding officer of the Board.

(b) KEY POSITIONS.—There shall be at the Academy the following:

(1) A Superintendent;

(2) A Dean of the Academic Board, who is a permanent professor;

(3) A Director of Admissions; and

(4) A Director of Placement.

(c) SUPERINTENDENT.—The Board shall appoint a Superintendent of the Academy, who shall serve for a term of six years. The Superintendent, acting pursuant to the oversight and direction of the Board, shall be responsible for the day-to-day operations of the Academy and the welfare of the students and the staff of the

Academy. The Board shall select the first Superintendent of the Academy no later than 60 days after the Board is established.

(d) ADVISORY BOARD.—The Board of Regents and the Superintendent shall be assisted by an Advisory Board, composed of commercial and academic leaders in digital technical fields and higher education. The Advisory Board shall adhere to the requirements of the Federal Advisory Committee Act, Pub.L. 92–463.

(e) INTERAGENCY WORKING GROUP.—

(1) ESTABLISHMENT.—The Office of Personnel Management shall establish and lead an interagency working group to annually assess and report to the Academy the need for civil servants at agencies in digital technical fields for the purposes of informing Academy student field of study and agency placement.

(2) RESPONSIBILITIES.—The interagency working group shall be responsible for—

(A) establishing a range of Academy graduates needed during the ensuing five-year period, by agency and digital technical field; and

(B) undertaking necessary steps to enable each agency identified to hire Academy graduates into full-time positions in the civil service.

(3) COMPOSITION.—The interagency working group shall consist of the following officials or their designees:

(A) The Secretary State;

(B) The Secretary of Defense;

(C) The Attorney General;

(D) The Secretary of Commerce;

(E) The Secretary of Energy;

(F) The Secretary of Homeland Security;

(G) The Director of National Intelligence;

(H) The Director of the Office of Personnel Management; and

(I) such other Federal Government officials as determined by the Director of the Office of Personnel Management.

SEC. 4. FACULTY.—

(a) NUMBER OF FACULTY.—The Superintendent of the Academy may employ as many professors, instructors, and lecturers at the Academy as the Superintendent considers necessary to achieve academic excellence.

(b) FACULTY COMPENSATION.—The Superintendent may prescribe the compensation of persons employed under this section. Compensation and benefits for faculty members of the Academy shall be sufficiently competitive to achieve academic excellence, as determined by the Superintendent.

(c) FACULTY EXPECTATIONS.—Faculty members shall—

(1) possess academic expertise and teaching prowess;

(2) exemplify high standards of conduct and performance;

(3) be expected to participate in the full spectrum of academy programs, including providing leadership for the curricular and extracurricular activities of students;

(4) comply with the standards of conduct and performance established by the Superintendent; and

(5) participate actively in the development of the students through the enforcement of standards of behavior and conduct, to be established in the Academy's rules and regulations.

(d) DEPARTMENT TITLES.—The Superintendent may prescribe the titles of each of the departments of instruction and the professors of the Academy.

SEC. 5. STUDENT QUALIFICATIONS AND REQUIREMENTS FOR ADMISSION.—

(a) ADMISSIONS REQUIREMENTS.—A student wishing to be admitted to the Academy shall fulfill admission requirements to be determined by the Superintendent and approved by the Board of Regents.

(b) HONOR CODE.—A student wishing to be admitted to the Academy shall sign an Honor Code developed by the Superintendent of the Academy and approved by the Board of Regents. A violation of the honor code may constitute a basis for dismissal from the Academy.

SEC. 6. APPOINTMENT OF STUDENTS.—

(a) NOMINATIONS PROCESS.—Prospective applicants to the Academy for seats described in paragraphs (1) and (2) of subsection (b) shall follow a nomination process established by the Director of Admissions of the Academy that is similar to the process used for admission to the military academies of the United States Armed Forces.

(b) APPOINTMENTS.—

(1) NOMINEES FOR CONGRESSIONAL SEATS.—Each member of the Senate or the House of Representatives may nominate candidates from the State that the member represents for each incoming first-year class of the Academy.

(2) EXECUTIVE BRANCH NOMINEES.—The President may nominate a maximum of 75 candidates to compete for the executive branch seats.

SEC. 7. ACADEMIC FOCUS OF THE UNITED STATES DIGITAL SERVICE ACADEMY—

(a) CURRICULUM.—Each Academy student shall follow a structured curriculum according to the program of study approved by the Board of Regents centered on digital technical fields and incorporating additional core curriculum coursework in history, government, English language arts including composition, and ethics.

(b) DEGREES CONFERRED UPON GRADUATION.—Under such conditions as the Board of Regents may prescribe, once the Academy is accredited, the Superintendent of the Academy may confer a baccalaureate of science or baccalaureate of arts degree upon a graduate of the Academy.

(c) MAJORS AND AREAS OF CONCENTRATION.—Under such conditions as the Board of Regents may prescribe, the Superintendent of the Academy may prescribe requirements for majors and concentrations and requirements for declaring a major or concentration during the course of study.

(d) ADDITIONAL DIGITAL SERVICE OF CIVIL SERVICE PROGRAMMING.—Under such conditions as the Board of Regents may prescribe, the Superintendent of the Academy may prescribe requirements for each Academy student to participate in non-curricular programing during Academy terms and during the summer, which may include internships, summer learning programs, and project-based learning activities.

SEC. 8. CIVIL SERVICE REQUIREMENTS FOLLOWING GRADUATION.—

(a) CIVIL SERVICE AGREEMENT.—Each Academy student, prior to commencing the third year of coursework, shall sign an agreement with respect to

the student's length of civil service to the United States. The agreement shall provide that the student agrees to the following:

(1) The student will complete the course of instruction at the Academy, culminating in graduation from the Academy.

(2) Unless the student pursues graduate education under subsection (f), upon graduation from the Academy, the student agrees to serve in the Federal civil service for not less than five years following graduation from the Academy.

(b) FAILURE TO GRADUATE.—

(1) IN GENERAL.—An Academy student who has completed a minimum of four semesters at the Academy but fails to fulfill the Academy's requirements for graduation shall be—

(A) dismissed from the Academy; and

(B) obligated to repay the Academy for the cost of the delinquent student's education in the amount described in paragraph (2).

(2) AMOUNT OF REPAYMENT.—A student who fails to graduate shall have financial responsibility for certain costs relating to each semester that the student was officially enrolled in the Academy as prescribed by the Superintendent.

(c) FAILURE TO ACCEPT OR COMPLETE ASSIGNED CIVIL SERVICE.—

(1) IN GENERAL.—A student who graduates from the Academy but fails to complete the full term of required civil service shall be obligated to repay the Academy for a portion of the cost of the graduate's education as determined by Academy as set forth in this subsection.

(2) AMOUNT OF REPAYMENT.—In the case of a delinquent graduate who fails to complete all years of public service required under subsection (a)(2) (including any additional years required for graduate education under subsection (f)), the delinquent graduate shall be financially responsible for the cost of the delinquent graduate's education (including the costs of any graduate education), except that the amount of financial responsibility under this paragraph shall be reduced by 20 percent for each year of civil service under subsection (a)(2) that the delinquent graduate did complete.

(d) EXCEPTIONS.—The Superintendent may provide for the partial or total waiver or suspension of any civil service or payment obligation by an individual

under this section whenever compliance by the individual with the obligation is impossible or deemed to involve extreme hardship to the individual, or if enforcement of such obligation with respect to the individual would be unconscionable.

(e) STUDENT SALARIES AND BENEFITS.—The Academy shall not be responsible for the salaries and benefits of graduates of the Academy while the graduates are fulfilling the civilian service assignment under this section. All salaries and benefits shall be paid by the employer with whom the Academy graduate is placed.

(f) GRADUATE EDUCATIONS.—An Academy student and the Superintendent may modify the agreement under subsection (a) to provide that—

(1) the Academy shall—

(A) subsidize an Academy student's graduate education; and

(B) postpone the public service assignment required under subsection (a)(2); and

(2) the student shall—

(A) accept a civil service assignment under subsection (g) upon the student's completion of the graduate program; and

(B) add two additional years to the student's civil service commitment required under the agreement described in subsection (a) for every year of subsidized graduate education.

SEC. 9. IMPLEMENTATION PLAN. –

(a) Not later than 180 days after the enactment of this Act, the Superintendent, in consultation with the Advisory Board, shall develop a detailed plan to implement the Academy that complies with the requirements of this section. Upon approval by the Board of Regents, the Superintendent shall present the implementation plan to Congress.

(b) CONTENTS OF PLAN.—The implementation plan described in section (a) shall provide, a minimum, the following:

(1) Identification and securement of an appropriate site for initial Academy build-out with room for future expansion, to include a construction plan and temporary site plan, if necessary;

(2) Identification of gaps in the government's current and envisioned digital workforce by the interagency working group under the Office of Personnel Management as established by section (3)(e);

(3) Establishment of student qualifications and requirements for admission;

(4) Establishment of the student appointment and nomination process;

(5) Establishment of student honor and conduct code to include a plan for student noncompletion of requirements and obligations;

(6) Establishment of the student curriculum;

(7) Establishment of a mechanism for students to select fields of study and annually select agencies and career fields within the limits prescribed by the interagency working group under the Office of Personnel Management as established by section (3)(e);

(8) Establishment of a mechanism for graduates to transition from the Academy to civil service employment by selected individual agencies;

(9) Determination of the initial Academy departments and faculty needs;

(10) Establishment of faculty and staff requirements and compensation;

(11) Determination of non-academic staff required;

(12) Recruitment and hiring of faculty, including tenure-track faculty, adjunct faculty, part-time faculty and visiting faculty, and other staff as needed;

(13) Identification of nonprofit and private sector partners;

(14) Procurement of outside funds and gifts from individuals and corporations for startup, administrative, maintenance, and infrastructure costs;

(15) Establishment of the process to meet statutory and regulatory requirements for establishing the Academy as an academic institution with degree-granting approval and for applying for degree program specific accreditation and ensuring that the Academy obtains, no later than two years after enactment of this Act, status as an accreditation candidate, as defined by a nationally recognized accrediting agency or association as determined by

the Secretary of Education in accordance with section 1099b in title 10, United States Code, before commencing academic operations;

(16) A plan commencing the Academy with an initial class of 500 students three years after enactment of this Act;

(17) Procedures for incorporating accreditation assessments to facilitate ongoing improvements to the Academy; and,

(18) Procedures for assessing the size of the Academy and potential expansion of student enrollment.

SEC. 10. ADMINISTRATIVE MATTERS.—

(a) FULLY-SUBSIDIZED EDUCATION.—Each Academy student's tuition and room and board shall be fully subsidized provided that the student completes the requirements of the Academy and fulfills the civil service commitment as determined by the implementation plan in section 9.

(b) GIFT AUTHORITY.—The Board of Regents may accept, hold, administer, and spend any gift, devise, or bequest of real property, personal property, or money made on the condition that the gift, devise, or bequest be used for the benefit, or in connection with, the establishment, operation, or maintenance, of the Academy. The Board of Regents may accept a gift of services, which includes activities that benefit the education, morale, welfare, or recreation of students, faculty or staff, for the Academy.

(1) LIMITATIONS AND PROHIBITIONS.—

(A) IN GENERAL.—The Board of Regents may not accept a gift under this subsection if the acceptance of the gift would reflect unfavorably on the ability of any agency of the Federal Government to carry out any responsibility or duty in a fair and objective manner, or would compromise the integrity or appearance of integrity of any program of the Federal Government or any officer or employee of the Federal Government who is involved in any such program.

(B) FOREIGN GIFTS.—The Board of Regents may not accept a gift of services from a foreign government or international organization under this subsection. A gift of real property, personal property, or money from a foreign government or international organization may be accepted under this subsection only if the gift is not designated for a specific individual.

(C) APPLICABLE LAW.—No gift under this section may be accepted with attached conditions inconsistent with applicable law or regulation.

(D) MISSION.—No gift under this section may be accepted with attached conditions inconsistent with the mission of the Academy.

(E) NAMING RIGHTS.—The Board of Regents may issue regulations governing the circumstances under which gifts conditioned on naming rights may be accepted, appropriate naming conventions, and suitable display standards.

(2) TREATMENT OF GIFTS.—

(A) Gifts and bequests of money, and the proceeds of the sale of property, received under subsection shall be deposited in the Treasury in the account of the Academy as no year money and may be expended in connection with the activities of the Academy as determined by the Board of Regents.

(B) The Board of Regents may pay all necessary expenses in connection with the conveyance or transfer of a gift, devise, or bequest accepted under this section.

(C) For the purposes of Federal income, estate, and gift taxes, any property, money, or services accepted under subsection shall be considered as a gift, devise, or bequest to or for the use of the United States.

(D) The Comptroller General shall make periodic audits of gifts, devises, and bequests accepted under this section at such intervals as the Comptroller General determines to be warranted. The Comptroller General shall submit to Congress a report on the results of each such audit.

SEC.11. INITIAL APPROPRIATION.—There are authorized to be appropriated $40,000,000 to remain available until expended for the Academy's initial administrative cost and salaries and expenses.

TAB 4 – Legislative Language

Recommendation 7: Grant Treasury the authority to mandate CFIUS filings for non-controlling investments in AI from China, Russia, and other competitor nation

SEC. ___. REVIEW OF SENSITIVE TRANSACTIONS INVOLVING COUNTRIES OF SPECIAL CONCERN.

(a) TECHNICAL AMENDMENTS.—Section 721(a) of the Defense Production Act of 1950 (50 USC 4565(a)) is amended by redesignating paragraphs (4), (5), (6), (7), (8), (9), (10), (11), (12), and (13) as paragraphs (5), (6), (7), (9), (10), (11), (12), (13), (15), and (16), respectively.

(b) DEFINITION OF COUNTRY OF SPECIAL CONCERN.—Section 721(a) of the Defense Production Act of 1950 (50 USC 4565(a)) is amended by inserting after paragraph (3) the following:

> "(4) COUNTRY OF SPECIAL CONCERN.—The term "country of special concern" means any country that is—
>
> > "(A) subject to export restrictions pursuant to section 744.21 of title 15, Code of Federal Regulations;
> >
> > "(B) determined by the Secretary of State to be a state sponsor of terrorism; or
> >
> > "(C) determined by the Committee to have a demonstrated or declared strategic goal of acquiring a type of technology or infrastructure that would have an adverse impact on United States leadership in areas related to national security, and is specified in regulations prescribed by the Committee."

(c) DEFINITION OF SENSITIVE TECHNOLOGY.—Section 721(a) of the Defense Production Act of 1950 (50 USC 4565(a)) is amended by inserting after redesignated paragraph (7) the following:

> "(8) SENSITIVE TECHNOLOGY.—The term 'sensitive technology' means any technology that is determined by the Committee to be necessary for maintaining or increasing the technological advantage of the United States over countries of special concern with respect to national defense, intelligence, or other areas of national security, or gaining such an advantage over such countries with respect to national defense, intelligence, or other areas of national security in areas where such an advantage may not exist,

and is not a critical technology as defined in paragraph (7) of this subsection, and is specified in regulations prescribed by the Committee.

(d) DEFINITION OF SENSITIVE TRANSACTION INVOLVING A COUNTRY OF SPECIAL CONCERN.— Section 721(a) of the Defense Production Act of 1950 (50 USC 4565(a)) is amended by inserting after redesignated paragraph (13) the following:

"(14) SENSITIVE TRANSACTION INVOLVING A COUNTRY OF SPECIAL CONCERN.—The term 'sensitive transaction involving a country of special concern' means any investment in an unaffiliated United States business by a foreign person that—

"(A) is—

"(i) a national or a government of, or a foreign entity organized under the laws of, a country of special concern; or

"(ii) a foreign entity—

"(I) over which control is exercised or exercisable by a national or a government of, or by a foreign entity organized under the laws of, a country of special concern; or

"(II) in which the government of a country of special concern has a substantial interest; and

"(B) as a result of the transaction, could achieve——

"(i) influence, other than through voting of shares, on substantive decision making of the United States business regarding the use, development, acquisition, or release of sensitive technologies, as defined in this section; or—

"(ii) access to material nonpublic technical information related to sensitive technologies, as defined in this section, in the possession of the United States business."

(e) DEFINITION OF COVERED TRANSACTIONS.—Section 721(a) of the Defense Production Act of 1950 (50 USC 4565(a)) is amended—

(1) in redesignated paragraph (5)(B)—

(A) in clause (iv)(I), by striking "or";

(B) in clause (iv)(II), by striking the period and inserting "; or"; and

(C) by adding at the end the following:

"(III) a sensitive transaction involving a country of special concern."

(2) by redesignating clause (v) as clause (vi) and inserting after clause (iv) the following:

"(v) Any sensitive transaction involving a country of special concern."

(f) INFORMATION REQUIRED IN ANNUAL REPORT TO CONGRESS.—Section 721(m)(2) of the Defense Production Act of 1950 (50 USC 4565(m)(2)) is amended by adding at the end the following:

"(L) Identification of each country designated as a country of special concern along with an explanation of the rationale for such designation.

"(M) Identification of each technology designated as a sensitive technology along with an explanation of the rationale for such designation."

(g) CONFORMING AMENDMENTS.—Title 50, United States Code, is amended—

(1) in section 4817(a)(1)(B) by striking "section 4565(a)(6)(A)" and inserting "section 4565(a)(7)(A)"; and

(2) in section 4565(b)(4)(B)(ii) (section 721(b)(4)(B)(ii) of the Defense Production Act of 1950) by striking "subsection (a)(4)(B)(ii)" and inserting "subsection (a)(5)(B)(ii)".

Appendix D — Q2 Funding Table

Category	Recommendation & Description	Cabinet Departments and Major Agencies	Amount
Tab 1: Accelerate AI R&D Across the DoD Research Enterprise	8.....Fully fund DoD FY 2021 request for software and digital technologies budget activity pilot.	Department of Defense, RDT&E	$857 million
	1.....Create a National Reserve Digital Corps.	Office of Management and Budget, Salaries & Expenses	$16 million*
Tab 3: Improve the U.S. Government's Digital Workforce	2.....Expand Scholarship for Service Programs. - CyberCorps: Scholarship for Service	National Science Foundation	$6 million
	2.....Expand Scholarship for Service Programs. - SMART: Scholarship for Service	Department of Defense	$7 million
	3.....Create a United States Digital Service Academy.	U.S. Digital Service Academy (new independent, Federal entity)	$40 million*
Tab 4: Improve Export Controls and Foreign Investment Screening	7.....Fully fund Treasury FY 2021 request to upgrade the CFIUS Case Management System IT Infrastructure	Department of the Treasury	$7.3 million
Tab: 5 Reorient the Department of State for Great Power Competition of the Digital Age	2.....Establish the Bureau of Cyberspace Security and Emergining Technology through realignment	Department of State	$17.8 million (FY 2021 realignment)

**Initial funding to be expended over two years*