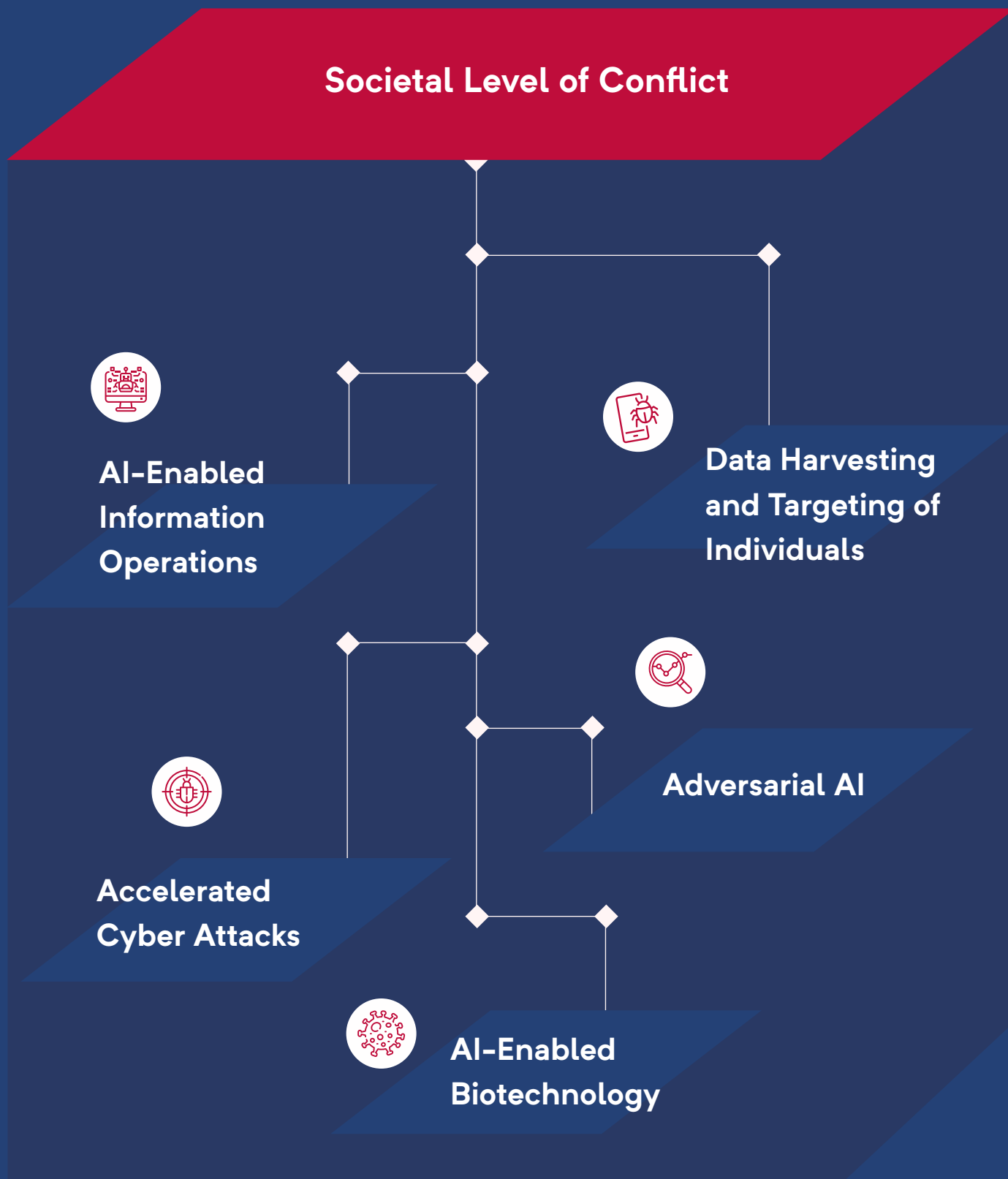


Chapter 1: Emerging Threats in the AI Era





The U.S. government is not prepared to defend the United States in the coming artificial intelligence (AI) era. AI applications are transforming existing threats, creating new classes of threats, and further emboldening state and non-state adversaries to exploit vulnerabilities in our open society.¹ AI systems will extend the range and reach of adversaries into the United States just as the missile age and terrorism brought threats closer to home. Because of AI, adversaries will be able to act with micro-precision, but at macro-scale and with greater speed. They will use AI to enhance cyber attacks and digital disinformation campaigns and to target individuals in new ways. AI will also help create precisely engineered biological agents. And adversaries will manipulate the AI systems we will rely upon.

How AI is
Transforming the
Threat Landscape

**Current Threats
Advanced BY AI Systems**

AI transforms existing range and reach of threats

- Self-replicating AI-generated malware
- Improved and autonomous disinformation campaigns
- AI-engineered and targeted pathogens

**New Threats
FROM AI Systems**

AI creates new threat phenomena

- Deepfakes and computational propaganda
- Micro-targeting: AI-fused data for targeting or blackmail
- AI swarms and nano-swarms

**Threats TO AI Stacks
Themselves**

AI itself is also a new attack surface

- AI attack involves the whole “AI stack”. Examples include:
 - Model inversion
 - Training data manipulation
 - “Data lake” poisoning

**Future Threats
VIA AI Systems**

Examples of potential threats to keep in view


- Rapid machine-to-machine escalation via automated C2
- AI-enabled human augmentation by peer competitors
- Proliferation of simple lethal autonomous weapons to terrorists

AI technologies exacerbate two existing national security challenges:

- First, digital dependence in all walks of life increases vulnerabilities to cyber intrusion across every segment of our society: corporations, universities, government, private organizations, and the homes of individual citizens. In parallel, new sensors have flooded the modern world. The internet of things (IoT), cars, phones, homes, and social media platforms collect streams of data, which can then be fed into AI systems that can identify, target, and manipulate or coerce our citizens.²
- Second, state and non-state adversaries are challenging the United States below the threshold of direct military confrontation by using cyber attacks, espionage, psychological and political warfare, and financial instruments. Adversaries do not need AI to conduct widespread cyber attacks, exfiltrate troves of sensitive data about American citizens, interfere in our elections, or bombard us with malign information on

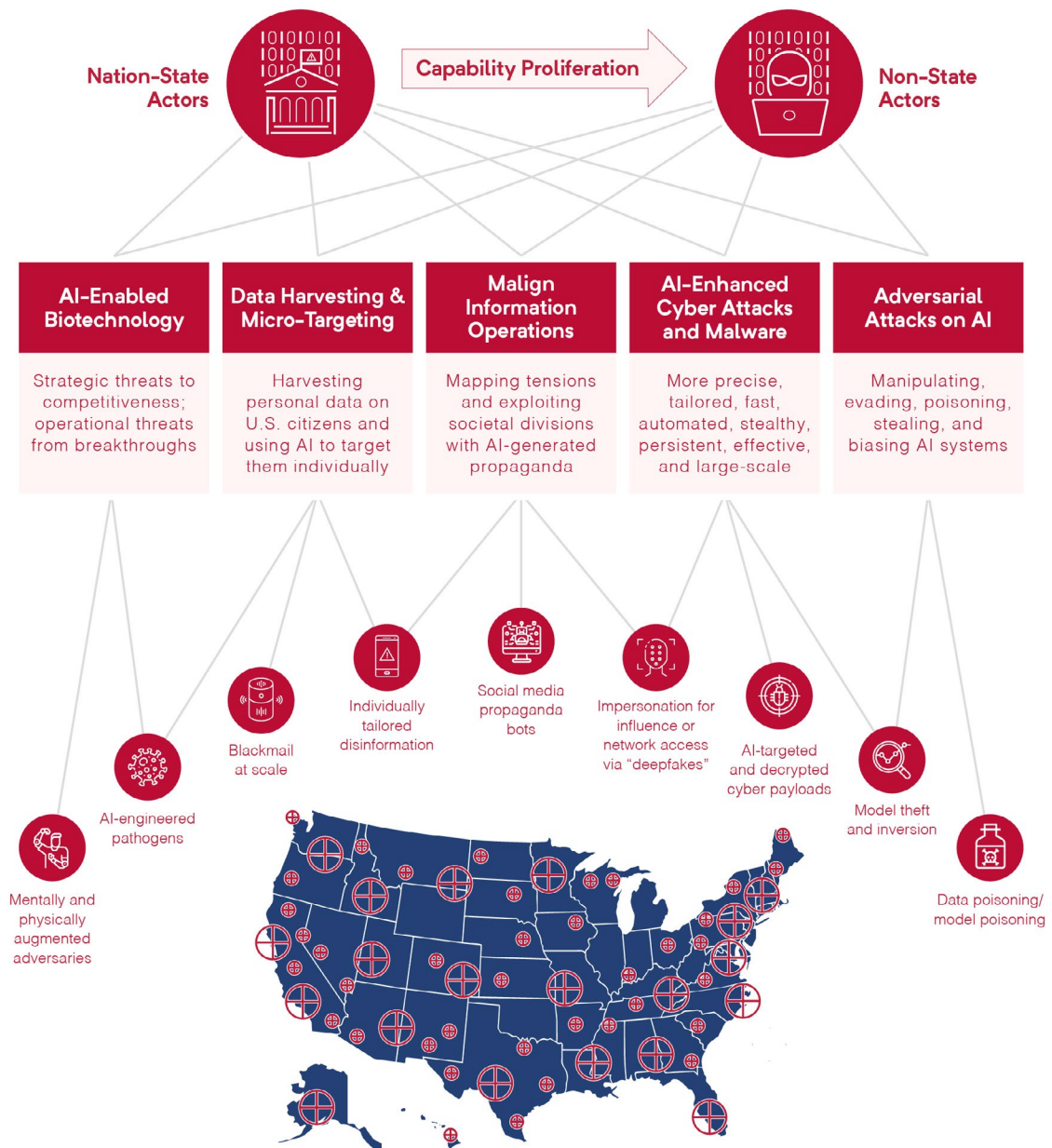
digital platforms. However, AI is starting to change these attacks in kind and in degree, creating new threats to the U.S. economy, critical infrastructure, and societal cohesion.³ Moreover, these AI-enabled capabilities will be used across the spectrum of conflict. They will be used as tools of first resort in non-military conflicts, as a prelude to military actions, or in concert with military actions in war.

Americans are waking to some of the privacy implications of their digital dependence and the potential threats from AI-powered malign information, like deep fakes. However, debate in the United States has not yet accounted for the full scope and danger of the AI-enabled threats and the overall security risks to the AI systems all around us. The prospect of adversaries using machine learning (ML), planning, and optimization to create systems to manipulate citizens' beliefs and behavior in undetectable ways is a gathering storm.⁴ Most concerning is the prospect that adversaries will use AI to create weapons of mass influence to use as leverage during future wars, in which every citizen and organization becomes a potential target.



“The prospect of adversaries using machine learning, planning, and optimization to create systems to manipulate citizens’ beliefs and behavior in undetectable ways is a gathering storm. Most concerning is the prospect that adversaries will use AI to create weapons of mass influence to use as leverage during future wars, in which every citizen and organization becomes a potential target.”

Societal
Level Impact.



The rest of this chapter discusses five AI-related threats that already have been, or soon will be, developed and used against the United States.

1. AI-Enabled Information Operations.

AI and associated technologies will increase the magnitude, precision, and persistence of adversarial information operations. AI exacerbates the problem of malign information in three ways:

- **Message.** AI can produce original text-based content and manipulate images, audio, and video, including through generative adversarial network (GAN)-enabled and reinforcement learning (RL) deep fakes that will be very difficult to distinguish from authentic messages.

- **Audience.** AI can construct profiles of individuals' preferences, behaviors, and beliefs to target specific audiences with specific messages.
- **Medium.** AI can be embedded within platforms, such as through ranking algorithms, to proliferate malign information.

AI-enabled malign information campaigns will not just send one powerful message to 1 million people, like 20th century propaganda. They also will send a million individualized messages—configured on the basis of a detailed understanding of the targets' digital lives, emotional states, and social networks.⁵ Rival states are already using AI-powered malign information. For example, according to Taiwan authorities, China's government tested its AI-powered malign information capacities during the 2020 Taiwan elections.⁶ A National Basketball Association general manager was harassed on social media for supporting protesters in Hong Kong, in an effort that may have involved autonomous bots.⁷ Other techniques rely on AI-generated fake personas.⁸ The control and manipulation of digital information has become central to the Kremlin's strategy, including in efforts to undermine the integrity of the democratic process in the United States and elsewhere.⁹

In the United States, the private sector has taken the leading role in combating foreign malign information. Social media companies in particular have extensive operations to track and manage information on their platforms. But coordination between the government and the social media firms remains ad hoc. We need a more integrated public-private response to the problem of foreign-generated disinformation. Moreover, the government needs to devote greater attention and resources to the technical challenges of detection, attribution, and media authentication. The government should:

Create a Joint Interagency Task Force and Operations Center. Congress has authorized a Foreign Malign Influence Response Center to be established within the Office of the Director of National Intelligence (ODNI).¹⁰ The government should use this authority to create a technologically advanced, 24-hour task force and operations center to lead and integrate government efforts to counter foreign-sourced malign information. It would survey the landscape of relevant public and private actors, coordinate among them, and act in real time to counter foreign information campaigns. To expose, attribute, and respond effectively, the center must be equipped with modern AI-enabled digital tools and staff with specialized expertise.

Recommendation

Fund the Defense Advanced Research Projects Agency (DARPA) to coordinate multiple research programs to detect, attribute, and disrupt AI-enabled malign information campaigns and to authenticate the provenance of digital media. Additional funding would amplify ongoing DARPA research programs to detect synthetic media and expand its efforts into attributing and disrupting malign information campaigns.¹¹ However promising some of these detection technologies may prove to be individually, funding to develop alternative technologies to authenticate the provenance of the digital media will provide a more technologically robust means to prevent the impersonation of trusted sources of information.¹² DARPA should pursue these programs and help transition all of these

Recommendation

technologies and applications to government departments and agencies, in order to assist with detecting, attributing, and disrupting malign information campaigns in real time.

Recommendation

Create a task force to study the use of AI and complementary technologies, including the development and deployment of standards and technologies, for certifying content authenticity and provenance. The White House Office of Science and Technology Policy should take the lead in creating this task force. In response to the challenges of misinformation, efforts are underway to develop standards and pipelines aimed at certifying the authenticity and provenance of audiovisual content.¹³ These efforts make use of technologies, including encryption and fragile watermarking, to secure and track the expected transformations of content via production and transmission pipelines. These efforts offer the opportunity to mitigate malign information campaigns that seek to corrupt or spoof highly trusted sources of information across our digital ecosystem. This technology area is ripe for public-private partnership. Several private organizations are already forming to fight disinformation efforts in this realm.¹⁴

2. Data Harvesting and Targeting of Individuals.

“Potential adversaries will recognize what every advertiser and social media company knows: AI is a powerful targeting tool.”

Data security is a national security problem. “Ad-tech” has become “natsec-tech.” Potential adversaries will recognize what every advertiser and social media company knows: AI is a powerful targeting tool. Just as AI-powered analytics transformed the relationship between companies and consumers, now it is transforming the relationship between governments and individuals. The broad circulation of personal data drives commercial innovation but also creates vulnerabilities.¹⁵ We fear that adversaries’ systematic efforts to harvest data on U.S. companies, individuals, and the government is about more than traditional espionage.¹⁶ Adversaries will combine widely available commercial data with data acquired illicitly—as in the 2015 Office of Personnel Management hack—to track, manipulate, and coerce individuals.¹⁷ The reach of tools that China, for instance, uses

to monitor, control, and coerce its own citizens—big data analytics, surveillance, and propaganda—can be extended beyond its borders and directed at foreigners.¹⁸ Without adequate data protection, AI makes it harder for anyone to hide his or her financial situation, patterns of daily life, relationships, health, and even emotions. Personal and commercial vulnerabilities become national security weaknesses as adversaries map individuals, networks, and social fissures in society; predict responses to different stimuli; and model how best to manipulate behavior or cause harm. The rise and spread of these techniques represent a major counterintelligence challenge.¹⁹

For the government to treat the data of its citizens and businesses as a national security asset, substantial changes are required in the way we think about data security and in our policies and laws to strengthen it. We need to identify categories and combinations of personal and commercial data that are most sensitive. Early efforts to limit foreign adversaries' data harvesting—such as the government's decision to force a Chinese company to relinquish ownership of a popular dating application for fear of what a hostile adversary could do with sensitive private data²⁰—represent important initial steps. However, the government lacks a broad approach with clear policies, criteria, or authorities to confront this multifaceted problem. The government should:

Develop policies that treat data security as national security, including in these areas:

Recommendation

- **First, from a technical standpoint, the government must ensure that a security development lifecycle approach is in place for its own AI systems (including commercial systems it acquires),** which should include a focus on potential privacy attacks.²¹ Red teaming must include privacy expertise. Government databases should be federated and anonymized whenever possible, and personal data retained no longer than is necessary, in order to make it more difficult for adversaries to utilize information for malicious purposes.
- **Second, the government should ensure that data privacy and security are priority considerations** as part of larger efforts to strengthen foreign investment screening and supply chain intelligence and risk management.²²
- **Third, national efforts to legislate and regulate data protection and privacy must integrate national security considerations,** such as limiting the ability of hostile foreign actors to acquire sensitive data on Americans on the commercial market.²³

3. Accelerated Cyber Attacks.

Malware in the AI era will be able to mutate into thousands of different forms once it is lodged on a computer system. Such mutating polymorphic malware already accounts for more than 90% of malicious executable files.²⁴ Deep RL tools can already find vulnerabilities, conceal malware, and attack selectively.²⁵ While it is uncertain which methods will dominate, there is a clear path for U.S. adversaries to transform the effectiveness of cyber attack and espionage campaigns with an ensemble of new and old algorithmic means to automate, optimize, and inform attacks.²⁶ This goes beyond AI-enhanced malware. Machine learning has current and potential applications across all the phases of cyber attack campaigns



“Machine learning has current and potential applications across all the phases of cyber attack campaigns ...”

and will change the nature of cyber warfare and cyber crime.²⁷ The expanding application of existing AI cyber capabilities will make cyber attacks more precise and tailored, further accelerate and automate cyber warfare, enable stealthier and more persistent cyberweapons, and make cyber campaigns more effective on a larger scale.

U.S. defenses have proven incapable of handling even more elementary cyber challenges. Vulnerabilities remain open in outdated infrastructure and medical devices, while new vulnerabilities are proliferating in 5G networks, billions of IoT devices, and in software supply chains.²⁸ The multibillion-dollar global damage caused by Russia’s 2017 NotPetya attack concretely demonstrates the power of even basic automated malware, the risk tolerance of capable state actors, and the consequences of such capabilities proliferating.²⁹ Though defensive applications of AI bring the promise to improve our national cyber defenses, AI can’t defend inherently vulnerable digital infrastructure. To address the present threat, Congress must continue implementing the Cyberspace Solarium Commission’s recommendations.³⁰ With this foundation for cyber defense, the U.S. can prepare for expanding threats via testing and building the instrumented infrastructure required for AI-enabled cyber defenses, establishing better incentives for security, properly organizing to meet the challenge, and keeping attackers off balance. Pervasive cyber-enabled espionage and attacks on U.S. computer networks and critical infrastructure will continue—and will become more damaging with AI—unless urgent federal action is taken. The government should:

Recommendation

Develop and deploy AI-enabled defenses against cyber attacks. National security agencies need to acquire the sensors and instrumentation needed to train AI systems to detect and respond to threats on their networks. AI-enabled cyber defenses will also need large-scale, instrumented, and realistic testing, and they must be robust enough to withstand adversarial attacks. The defenses should be employed to expand machine speed information sharing, behavior-based anomaly detection, and malware mitigation across government networks. To capitalize on these capabilities, the government should accelerate the establishment of a Joint Cyber Planning and Operations Center, modeled after the National Counterterrorism Center.³¹ The Center would serve as a centralized cyber intelligence sharing and collaboration unit with multi-agency jurisdiction and authorities to investigate threats, proactively support defensive mitigations, and coordinate responses.

4. Adversarial AI.

AI systems represent a new target for attack. While we are on the front edge of this phenomenon, commercial firms and researchers have documented attacks that involve evasion, data poisoning, model replication, and exploiting traditional software flaws to deceive, manipulate, compromise, and render AI systems ineffective.³² This threat is related to, but distinct from, traditional cyber activities, because AI systems will be vulnerable to adversarial attacks from any domain where AI augments action—civilian or military.³³ Given the reliance of AI systems on large data sets and algorithms, even small manipulations of these data sets or algorithms can lead to consequential changes for how AI systems operate. The threat is not hypothetical: adversarial attacks are happening and already impacting commercial ML systems.³⁴ With rare exceptions, the idea of protecting AI systems has been an afterthought in engineering and fielding AI systems, with inadequate investment in research and development.³⁵ Only three of 28 organizations recently surveyed have “the right tools in place to secure their ML systems.”³⁶ There has not yet been a uniform effort to integrate AI assurance across the entire U.S. national security enterprise. To improve AI “assurance,” the government should:

Create a National AI Assurance Framework. All government agencies will need to develop and apply an adversarial ML threat framework to address how key AI systems could be attacked and should be defended. An analytical framework can help to categorize threats to government AI systems and assist analysts with detecting, responding to, and remediating threats and vulnerabilities.³⁷

Recommendation

Create dedicated red teams for adversarial testing. Red teams should assume an offensive posture, trying to break systems and make them violate rules for appropriate behavior. Because of the scarcity of required expertise and experience for AI red teams, DoD and ODNI should consider establishing government-wide communities of AI red-teaming capabilities that could be applied to multiple AI developments.³⁸

Recommendation

5. AI-Enabled Biotechnology.

Biology is now programmable. New technologies such as the gene editing tool CRISPR ushered in an era where humans are able to edit DNA. Combined with massive computing power and AI, innovations in biotechnology may provide novel solutions for mankind's most vexing challenges, including in health, food production, and environmental sustainability. Like other powerful technologies, however, applications of biotechnology can have a dark side. The COVID-19 pandemic reminded the world of the dangers of a highly contagious pathogen. AI may enable a pathogen to be specifically engineered for lethality or to target a genetic profile—the ultimate range and reach weapon. Also, AI, when applied to biology, could optimize for the physiological enhancement of human beings, including intelligence and physical attributes. To the extent that brain waves can be represented as a machine vision challenge for AI, the mysteries of the brain may be unlocked and programmed.

Individuals, societies, and states will have different moral and ethical views and accept different degrees of risk in the name of progress, and U.S. competitors are comparatively likely to take more risk-tolerant actions and conform less rigidly to bioethical norms and standards. China understands the tremendous upside associated with leading the bio revolution. Massive genomic data sets at places like BGI Group (formerly known as the Beijing Genomics Institute), coupled with China's now-global genetic data collection platform and "all-of-nation" approach to AI, will make them a formidable competitor in the bio realm.³⁹ BGI may be serving, wittingly or unwittingly, as a global collection mechanism for Chinese government genetic databases, providing China with greater raw numbers and diversity of human genome samples as well as access to sensitive personal information about key individuals around the world.⁴⁰ The United States cannot afford to look back in 10 years and be "surprised" by the biotechnology equivalent of Huawei. Additionally, Russia's long-standing disregard for scientific norms and bioethical principles, demonstrated by its development and employment of novel nerve agents such as Novichok for assassination attempts and U.S. government concerns over Russia's compliance with the Biological Weapons Convention, could presage a willingness to utilize advanced biotechnology abilities for nefarious purposes.⁴¹ The government should:

Recommendation

Increase the profile of biosecurity and biotechnology issues within U.S. national security agencies. Given how AI will substantially increase the rate of technical advancement in biotechnology, the government should update the National Biodefense Strategy to include a wider vision of biological threats, such as human enhancement, exploitation of genetic data for malicious ends, and ways U.S. competitors could utilize biotechnology or biodata advantages for novel purposes. Additionally, U.S. officials should warn of the dangers associated with foreign actors obtaining personal genetic information, specifically highlighting concerns about the links between BGI and the Chinese government.⁴²

Chapter 1 - Endnotes

¹ A threat can be understood as an adversary capability paired with a vulnerability that can create a harmful consequence. See Terry L. Deibel, *Foreign Affairs Strategy: Logic for American Statecraft*, Cambridge University Press at 142-150 (2007). Threats can be graded further by their seriousness, likelihood, imminence, and tractability.

² Stuart A. Thompson & Charlie Warzel, *Twelve Million Phones, One Dataset, Zero Privacy*, New York Times (Dec. 19, 2019), <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>. For example, the internet of things (IoT) and AI-powered applications can turn your new robotic vacuum into a listening device. See Sriram Sami, et al., *Spying with Your Robot Vacuum Cleaner: Eavesdropping via Lidar Sensors*, Proceedings of the 18th Conference on Embedded Networked Sensor Systems (Nov. 2020), <https://dl.acm.org/doi/10.1145/3384419.3430781>.

³ This is in some ways analogous to what Cold War strategists called “counter-value targeting.” See Lawrence Freedman, *The Evolution of Nuclear Strategy*, Palgrave Macmillan Vol. 20 at 119-122 (1989). In the realm of nuclear strategy, this was also known as counter-city or counter-economy targeting.

⁴ Some observers have used the concept of “sharp power” to describe such efforts to wield influence in open societies. These uses of power are sharp “in the sense that [authoritarian states aim to] pierce, penetrate, or perforate the information environments in the targeted countries.” *Sharp Power: Rising Authoritarian Influence*, National Endowment for Democracy at 13 (Dec. 5, 2017), <https://www.ned.org/sharp-power-rising-authoritarian-influence-forum-report/>. See also Testimony of Dr. Eric Horvitz, Microsoft, before the U.S. Senate Committee on Commerce, Science, & Transportation, Subcommittee on Space, Science, & Competitiveness, *Hearing on the Dawn of Artificial Intelligence* at 13 (Nov. 30, 2016), http://erichorvitz.com/Senate_Testimony_Eric_Horvitz.pdf.

⁵ Some have characterized AI-driven information operations as “computational propaganda.” See Matt Chessen, *The MADCOM Future: How Artificial Intelligence Will Enhance Computational Propaganda, Reprogram Human Culture, and Threaten Democracy... and What Can Be Done About It*, Atlantic Council (Sept. 2017), https://www.atlanticcouncil.org/wp-content/uploads/2017/09/The_MADCOM_Future_RW_0926.pdf.

⁶ Philip Sherwell, *China Uses Taiwan for AI Target Practice to Influence Elections*, The Australian (Jan. 5, 2020), <https://www.theaustralian.com.au/world/the-times/china-uses-taiwan-for-ai-target-practice-to-influence-elections/news-story/57499d2650d4d359a3857688d416d1e5>.

⁷ Ben Cohen, et al., *How One Tweet Turned Pro-China Trolls Against the NBA*, Wall Street Journal (Oct. 16, 2019), <https://www.wsj.com/articles/how-one-tweet-turned-pro-china-trolls-against-the-nba-11571238943>. On automated bots, see, e.g., Sarah Kreps & Miles McCain, *Not Your Father's Bots*, Foreign Affairs (Aug. 2, 2019), <https://www.foreignaffairs.com/articles/2019-08-02/not-your-fathers-bots>.

⁸ James Vincent, *An Online Propaganda Campaign Used AI-Generated Headshots to Create Fake Journalists*, The Verge (July 7, 2020), <https://www.theverge.com/2020/7/7/21315861/ai-generated-headshots-profile-pictures-fake-journalists-daily-beast-investigation>.

⁹ For recent studies on technical aspects of Russia's interference in the 2016 election, see Alexander Spangher, et al., *Characterizing Search-Engine Traffic to Internet Research Agency Web Properties*, Web Conference (2020), <https://www.microsoft.com/en-us/research/publication/characterizing-search-engine-traffic-to-internet-research-agency-web-properties/>; Ryan Boyd, et al., *Characterizing the Internet Research Agency's Social Media Operations During the 2016 U.S. Presidential Election Using Linguistic Analyses*, PsyArXiv Preprints (2018), <https://psyarxiv.com/ajh2q/>. See also Alina Polyakova, *Weapons of the Weak: Russian and AI-driven Asymmetric Warfare*, Brookings Institution (Nov. 15, 2018), <https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/>.

¹⁰ See Pub. L. 116-92, National Defense Authorization Act for Fiscal Year 2020, 133 Stat. 1198, 2129 (2019).

¹¹ These include the Media Forensics (MediFor) and Semantic Forensics (SemaFor) programs. See Dr. Matt Turek, *Media Forensics*, DARPA (last accessed Jan.10, 2021), <https://www.darpa.mil/program/media-forensics>; Dr. Matt Turek, *Semantic Forensics*, DARPA (last accessed Jan.10, 2021), <https://www.darpa.mil/program/semantic-forensics>.

Chapter 1 - Endnotes

¹² See, e.g., Paul England, et al., *AMP: Authentication of Media via Provenance*, arXiv (June 20, 2020), <https://arxiv.org/abs/2001.07886>.

¹³ See, e.g., Paul England, et al., *AMP: Authentication of Media via Provenance*, arXiv.

¹⁴ See *Creating the Standard for Digital Content Attribution*, Content Authenticity Initiative, <https://contentauthenticity.org/>; and *Project Origin: Protecting Trusted Media*, Project Origin, <https://www.originproject.info/about>.

¹⁵ Robert Williams has described how policy makers face an “innovation-security conundrum,” one aspect of which is “the worry that data privacy and national security are increasingly interconnected. Data (and data networks) can be exploited in ways that threaten security, but they also form the lifeblood of technological innovation on which both economic growth and national security depend.” Robert D. Williams, *Crafting a Multilateral Technology and Cybersecurity Policy*, Brookings at 1 (Nov. 2020), <https://www.brookings.edu/wp-content/uploads/2020/11/Robert-D-Williams.pdf>.

¹⁶ Ellen Nakashima, *With a Series of Major Hacks, China Builds a Database on Americans*, Washington Post (June 5, 2015), https://www.washingtonpost.com/world/national-security/in-a-series-of-hacks-china-appears-to-building-a-database-on-americans/2015/06/05/d2af51fa-0ba3-11e5-95fd-d580f1c5d44e_story.html.

¹⁷ Another example of an adversary acquiring significant data on U.S. individuals is the hack of the credit reporting agency Equifax. Press Release, Department of Justice, *Chinese Military Personnel Charged with Computer Fraud, Economic Espionage and Wire Fraud for Hacking into Credit Reporting Agency Equifax* (Feb. 10, 2020), <https://www.justice.gov/opa/pr/chinese-military-personnel-charged-computer-fraud-economic-espionage-and-wire-fraud-hacking>; Aruna Viswanatha, et al., *Four Members of China's Military Indicted Over Massive Equifax Breach*, Wall Street Journal (Feb. 11, 2020), <https://www.wsj.com/articles/four-members-of-china-s-military-indicted-for-massive-equifax-breach-11581346824>.

¹⁸ See, e.g., Drew Harwell & Eva Dou, *Huawei Tested AI Software That Could Recognize Uighur Minorities and Alert Police, Report Says*, Washington Post (Dec. 8, 2020), <https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says/>; Hugh Harsono, *China's Surveillance Technology Is Keeping Tabs on Populations Around the World*, The Diplomat (June 18, 2020), <https://thediplomat.com/2020/06/chinas-surveillance-technology-is-keeping-tabs-on-populations-around-the-world/>.

¹⁹ See James Baker, *Counterintelligence Implications of Artificial Intelligence—Part III*, Lawfare (Oct. 10, 2018), <https://www.lawfareblog.com/counterintelligence-implications-artificial-intelligence-part-iii>.

²⁰ Yuan Yang & James Fontanella-Khan, *Grindr Sold by Chinese Owner After US National Security Concerns*, Financial Times (March 7, 2020), <https://www.ft.com/content/a32a740a-5fb3-11ea-8033-fa40a0d65a98>.

²¹ On privacy attacks, see Maria Rigaki & Sebastian Garcia, *A Survey of Privacy Attacks in Machine Learning*, arXiv (July 15, 2020), <https://arxiv.org/abs/2007.07646>.

²² The Committee on Foreign Investment in the United States (CFIUS) has the authority to review transactions that include “sensitive personal data of United States citizens that may be exploited in a manner that threatens national security.” For background, see Laura Jehl, *Spotlight on Sensitive Personal Data As Foreign Investment Rules Take Force*, The National Law Review (Feb. 18, 2020), <https://www.natlawreview.com/article/spotlight-sensitive-personal-data-foreign-investment-rules-take-force>. The National Counterintelligence and Security Center (NCSC) includes “sensitive government data, and personally-identifiable information” in its conception of key supply chain risks. See *Supply Chain Risk Management: Reducing Threats to Key U.S. Supply Chains*, NCSC at 3 (2020), <https://www.dni.gov/files/NCSC/documents/supplychain/20200925-NCSC-Supply-Chain-Risk-Management-tri-fold.pdf>.

²³ See, e.g., Graham Webster, *App Bans Won't Make U.S. Security Risks Disappear*, MIT Technology Review (Sept. 21, 2020), <https://www.technologyreview.com/2020/09/21/1008620/wechat-tiktok-ban-china-us-security-policy-opinion/>.

²⁴ Nicholas Duran, et al., *2018 Webroot Threat Report*, Webroot at 6 (2018), https://www-cdn.webroot.com/9315/2354/6488/2018-Webroot-Threat-Report_US-ONLINE.pdf.

²⁵ Gary J. Saavedra, et al., *A Review of Machine Learning Applications in Fuzzing*, arXiv (Oct. 9, 2019), <https://arxiv.org/pdf/1906.11133.pdf>; Isao Takaesu, *Machine Learning Security: DeepExploit*, GitHub (Aug. 29, 2019), https://github.com/13o-bbr-bbq/machine_learning_security/tree/master/DeepExploit; Marc Ph. Stoecklin, et al., *DeepLocker: How AI Can Power a Stealthy New Breed of Malware*, Security Intelligence (Aug. 8, 2018), <https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/>.

²⁶ *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*, National Academies of Sciences, Engineering, and Medicine (2019), <https://doi.org/10.17226/25488>; Nektaria Kaloudi & Jingyue Li, *The AI-Based Cyber Threat Landscape*, ACM Computing Surveys at 1-34 (Feb. 2020), <https://dl.acm.org/doi/abs/10.1145/3372823>; Ben Buchanan, et al., *Automating Cyber Attacks*, Center for Security and Emerging Technology (Nov. 2020), <https://cset.georgetown.edu/research/automating-cyber-attacks/>; Dakota Cary & Daniel Cebul, *Destructive Cyber Operations and Machine Learning*, Center for Security and Emerging Technology at 5-23 (Nov. 2020), <https://cset.georgetown.edu/research/destructive-cyber-operations-and-machine-learning/>.

²⁷ *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*, National Academies of Sciences, Engineering, and Medicine (2019), <https://doi.org/10.17226/25488>.

²⁸ The recent SolarWinds attack demonstrates deep vulnerabilities in our software supply chains. See *Joint Statement by the Federal Bureau of Investigation (FBI), the Cybersecurity and Infrastructure Security Agency (CISA), and the Office of the Director of National Intelligence (ODNI)*, Office of the Director of National Intelligence (Dec. 16, 2020), <https://www.dni.gov/index.php/newsroom/press-releases/item/2175-joint-statement-by-the-federal-bureau-of-investigation-fbi-the-cybersecurity-and-infrastructure-security-agency-cisa-and-the-office-of-the-director-of-national-intelligence-odni>.

²⁹ *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*, National Academies of Sciences, Engineering, and Medicine (2019), <https://doi.org/10.17226/25488>; Ben Buchanan, et al., *Automating Cyber Attacks*, Center for Security and Emerging Technology at 3 (Nov. 2020), <https://cset.georgetown.edu/research/automating-cyber-attacks/>.

³⁰ *Cyberspace Solarium Commission Report*, U.S. Cyberspace Solarium Commission (March 2020), <https://www.solarium.gov/report>.

³¹ See recommendation 5.4 in *Cyberspace Solarium Commission Report*, U.S. Cyberspace Solarium Commission at 87 (March 2020), <https://www.solarium.gov/report>.

³² *Adversarial AI Threat Matrix: Case Studies*, GitHub (last accessed Jan. 10, 2021), <https://github.com/mitre/advmthreatmatrix/blob/master/pages/case-studies-page.md>. For more on applications of adversarial AI, see Naveed Akhtar & Ajmal Mian, *Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey*, IEEE (March 28, 2018), <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8294186>.

³³ Adversarial AI is about what can be done to AI systems. The science of protecting and defending AI applications against attacks is called “AI Assurance.” The science of attacking each technological component of AI is called “Counter-AI.”

³⁴ Ram Shankar Siva Kumar & Ann Johnson, *Cyberattacks Against Machine Learning Systems Are More Common Than You Think*, Microsoft Security (Oct. 22, 2020), <https://www.microsoft.com/security/blog/2020/10/22/cyberattacks-against-machine-learning-systems-are-more-common-than-you-think/>.

³⁵ It has been estimated that less than 1% of AI R&D funding is directed toward the security of AI systems. See Nathan Strout, *The Three Major Security Threats to AI*, Center for Security and Emerging Technology (Sept. 10, 2019), <https://cset.georgetown.edu/article/the-three-major-security-threats-to-ai/>.

³⁶ Ram Shankar Siva Kumar, et al., *Adversarial Machine Learning—Industry Perspectives*, arXiv at 2 (May 21, 2020), <https://arxiv.org/pdf/2002.05646.pdf>.

Chapter 1 - Endnotes

³⁷ There are various ongoing public and private efforts including, for instance, the MITRE-Microsoft adversarial ML framework. See Ram Shankar Siva Kumar & Ann Johnson, *Cyberattacks Against Machine Learning Systems Are More Common Than You Think*, Microsoft Security (Oct. 22, 2020), <https://www.microsoft.com/security/blog/2020/10/22/cyberattacks-against-machine-learning-systems-are-more-common-than-you-think/>; *Adversarial AI Threat Matrix: Case Studies*, MITRE (last accessed Jan. 10, 2021), <https://github.com/mitre/advmthreatmatrix/blob/master/pages/case-studies-page.md>.

³⁸ For a similar recommendation, see Michèle Flournoy, et al., *Building Trust Through Testing*, WestExec Advisors at 27 (Oct. 2020), <https://cset.georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf>. (Flournoy, et al., argue for “a national AI and ML red team as a central hub to test against adversarial attacks, pulling together DoD operators and analysts, AI researchers, T&E [Central Intelligence Agency (CIA), Defense Intelligence Agency (DIA), National Security Agency (NSA)], and other IC components, as appropriate. This would be an independent red-teaming organization that would have both the technical and intelligence expertise to mimic realistic adversary attacks in a simulated operational environment.”)

³⁹ BGI built and operates China National GeneBank, the Chinese government's national genetic database. It also is a major global supplier of COVID-19 testing, which potentially provides access to large international genetic data sets; by June 30, 2020, it had supplied more than 35 million test kits to 180 countries, including the United States, and built 58 testing labs in 18 countries. See Kirsty Needham, Special Report: *COVID Opens New Doors for China's Gene Giant*, Reuters (Aug. 5, 2020), <https://www.reuters.com/article/us-health-coronavirus-bgi-specialreport/special-report-covid-opens-new-doors-for-chinas-gene-giant-idUSKCN2511CE>.

⁴⁰ John Wertheim, *China's Push to Control Americans' Health Care Future*, 60 Minutes (Jan. 31, 2021), <https://www.cbsnews.com/news/china-us-biodata-60-minutes-2021-01-28/?ftag=CNM-00-10aab7d&linkId=110169507>; Kirsty Needham, *Special Report: COVID Opens New Doors for China's Gene Giant*, Reuters (Aug. 5, 2020), <https://www.reuters.com/article/us-health-coronavirus-bgi-specialreport/special-report-covid-opens-new-doors-for-chinas-gene-giant-idUSKCN2511CE>.

⁴¹ See Richard Pérez-Peña, *What Is Novichok, the Russian Nerve Agent Tied to Navalny Poisoning?*, New York Times (Sept. 2, 2020), <https://www.nytimes.com/2020/09/02/world/europe/novichok-skripal.html>; *2020 Adherence to and Compliance with Arms Control, Nonproliferation, and Disarmament Agreements and Commitments (Compliance Report)*, U.S. Department of State at Pt. V (2020), https://2017-2021.state.gov/2020-adherence-to-and-compliance-with-arms-control-nonproliferation-and-disarmament-agreements-and-commitments-compliance-report-2/index.html#_Toc43298166.

⁴² See Chapter 16 of this report for additional recommendations pertaining to the nexus of AI and biotechnology.