Time for a Counter-AI Strategy

The United States and China have each vowed to become the global leader in artificial intelligence (AI). In 2016, the United States published its National Artificial Intelligence Research and Development Strategic Plan. In 2017, China released its "New Generation Artificial Intelligence Development Plan," announcing its intention to leapfrog the United States to become the global leader in AI by 2030 by combining government and private sector efforts.[1] The United States countered with the publication of the 2018 Department of Defense Artificial Intelligence Strategy, focused on maintaining AI leadership through faster innovation and adoption, and in 2019 updated its original plan.[2]

The competition has been characterized as an "AI arms race," measured by expenditure, number of patents filed, or speed of adoption. On the battlefield, the perceived benefits of AI are increased speed and precision as AI systems rapidly handle tasks such as target identification, freeing humans for higher-level cognitive tasks. AI will, in theory, help the military to act faster, eclipsing its adversary's ability to observe, orient, decide, and act.

The singular strategic focus on gaining and maintaining leadership and the metaphor of an "arms race" are unhelpful, however. Races are unidimensional, and the winner takes all. Previous arms races in long-range naval artillery or nuclear weapons were predicated on the idea that advanced tech would create standoff, nullifying the effects of the adversary's weapons and deterring attack. But AI is not unidimensional; it is a diverse collection of applications, from AI-supported logistics and personnel systems to AI-enabled drones and autonomous vehicles. Nor does broadly better tech necessarily create standoff, as the US military learned from improvised explosive devices in Afghanistan. This means that in addition to improving its own capabilities, the United States must be able to respond effectively to the capabilities of others. In addition to its artificial intelligence strategy, the United States needs a counter-AI strategy.

The AI Challenge

US competitors are already making military use of AI. In the military parade that marked the 70th anniversary of the Chinese Communist Party, the People's Liberation Army displayed autonomous vehicles and drones.[3] At the same time, Russia is forging ahead with the Status-6, a nuclear autonomous torpedo.[4] Less capable countries will acquire AI-enabled weapons and systems through purchases or security cooperation.

The popular focus on military AI has been on tactical applications such as weapons targeting, and AI will be most successful when applied to static, simple problems. However, AI-enabled competitors and adversaries will develop new decision-making processes, modes of operation and coordination, battlefield capabilities, and weapons. Enterprise systems in human resources, logistics, procurement, equipment management and maintenance, accounting, intelligence collection and analysis, and reporting may also be AI-enabled. Operational and strategic leaders may turn to AI systems to suggest or test courses of action.

AI will likely create vulnerabilities as well as advantages. It may be error prone or biased, unpredictable, unreliable, opaque, and less capable of fine discrimination. Paul Scharre of the Center for a New American Security warns of the possibility of "a million mistakes a second" and rapid AI-enabled escalation of the kind illustrated by the 2010 Wall Street "flash crash" driven by automated trading programs.[5] Although he calls for a greater investment in testing to ensure the reliability of AI systems, AI may be intrinsically unreliable. For example, the problems to which AI is applied may be dynamic, or the AI itself may be constantly updated with new data.[6] Further, the interaction of multiple, different AI systems may produce unanticipated emergent behaviors.

Humans may hesitate to trust their own AIs—there is active research in developing "explainable AI" to foster human trust—but it is more likely that they will trust them too much.[7] Just as there is a generation of "digital natives" who grew up with computers, there will be a new generation of "AI natives" who are sophisticated users but take the technology for granted, do not know how it operates, do not understand its limitations, and lack the skills to operate without it. To the extent that they habitually use AI to tee up choices, it may be more difficult for them to generate creative options.

Strategic Counter-AI Initiatives

A counter-AI strategy would seek to harden the United States as a target for AI-enabled attacks, reduce the advantages of AI to an adversary, and predict and adapt to changes in behavior that are consequences of reliance on AI. Among other measures, the United States could take more aggressive steps to protect US data that could be used for training AI models, invest in counter-AI tactics, and change how it comprehends AI behavior. Finally, the United States should cultivate self-awareness of the vulnerabilities created by its own increasing reliance on AI systems.

Protect Relevant Data Sets

The United States should seek to better protect sensitive data sets from adversaries that may use them to develop ("train") AI models. A particularly damaging hack in the DOD occurred with the 2015 infiltration of the Office of Personnel Management in which an estimated 21.5 million personnel files were compromised, including the forms submitted by individuals to apply for or maintain the clearances that give them access to classified information.[8] Such data might be used to develop a predictive model for intelligence targeting that estimates the likelihood that a person has a high-level clearance.

At present, US policy on data protection is inconsistent. The executive order Maintaining American Leadership in Artificial Intelligence requires agencies to set as a strategic objective the enhancement of "access to high-quality . . . [f]ederal data [consistent with] safety, security, privacy and confidentiality protections."[9] However, these criteria may not be sufficient because the information can be used to train models even if it is fully anonymized and so does not present privacy concerns.

The handling of private data is also a concern. A number of countries have passed data localization laws that require data collected in country to be stored in country.[10] Localization allows governments to set and enforce standards for the securitization and handling of private data that might otherwise be stored in extraterritorial servers. However, such laws also come at a price of reduced efficiency for global economic exchanges. Authoritarian governments may also use such laws to access their citizens' data and enforce censorship.[11] India is debating data localization while the European Union has explicitly rejected it.[12]

The United States has also rejected localization. The United States Trade Representative has called out China, India, Indonesia, Kenya, Korea, Nigeria, Russia, Saudi Arabia, Turkey, and Vietnam for data restrictions that inhibit digital trade and impair global competitiveness.[13] But at the same time, the Committee on Foreign Investment in the United States has used authority under new legislation to prevent foreign acquisition of private data by, for example, forcing Chinese divestment from Grindr, a dating app that collects personal information.[14] Eric Rosenbach and Katherine Mansted of the Harvard Kennedy School Belfer Center for Science and International Affairs anticipate stepped-up cyberattacks by adversaries on data sets that can be used for training AI and call for a national information policy to protect data.[15]

Invest in Counter-AI Tactics

The United States should invest in research for counter-AI tactics. For example, research on adversarial images focuses on how to defeat AI image recognition systems, which can be thrown off course by subtle changes in the image to be analyzed. Researchers developed an image of a turtle classified by an AI program as a rifle and an image of a baseball classified as espresso.[16] Others have developed an AI program that can subtly tweak facial images to reduce the possibility of detection by AI facial recognition programs.[17] Slight physical defacements can defeat the ability of AI programs to recognize street signs. However, these approaches can be very specific to the implementation of the AI program that they seek to defeat.

More broadly, the United States must invest in developing methods to hack, crack, and outpace an adversary's AI by taking advantage of AI error and biases, the inability of AI to adapt to novelty, and the vulnerability of channels used for developing and pushing software updates. Exploiting such flaws would involve identifying where adversaries rely on AI and for what purposes, reverse engineering AI systems, red teaming the likely decisions of AI programmers (by, for example, identifying the likely source of training data or the algorithms used), and using generative adversarial nets—programs that seek the limits of AI classification abilities. Expertise in counter-AI tactics should be co-located with expertise in offensive cyber capabilities. Tactical counter-AI may need offensive cyber to open the door to AI-enabled systems or to block or spoof pushed software updates, while cyber may need AI expertise to take on AI-enabled cyber adversaries.

Change How We Predict and Understand Adversary Behavior

Analysts charged with assessing and anticipating competitor and adversary behavior will need new approaches. As illustrated by the work on adversarial images, AI programs make mistakes no human would make—which will make those who rely on them less predictable. Sherman Kent, the famed CIA intelligence analysis pioneer, explained why the Central Intelligence Agency estimates during the Cuban missile crisis gave no credence to the idea that Khrushchev had put missiles in Cuba. He wrote, "It is when the other man zigs violently out of the track of 'normal' behavior that you are likely to lose him. If you lack hard evidence of the prospective erratic tack and the zig is so far out of line as to seem to you to be suicidal, you will probably misestimate him every time."[18] It will also become more difficult to ascribe intentionality to adversary actions, a particular concern in situations that may be escalatory. At the same time, the

United States should consider that competitors and adversaries seeking to understand US behavior will have identical challenges.

The current strategy of the United States assumes that AI leadership will ensure dominance and deter. The reality of AI is more complicated and ambiguous. The United States needs to consider how it will deal effectively with competitors and adversaries that rely on AI and how it will address the vulnerabilities that arise from its own increasing reliance. SSQ

M. A. Thomas
Professor, US Army School of Advanced Military Studies

Notes

1. Elsa Kania, "China's AI Agenda Advances," *The Diplomat*, 14 February 2018, https://thediplomat.com/.

2. US Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity* (Washington, DC: US Department of Defense, 2019), https://media.defense.gov/.

3. Patrick Tucker, "New Drones, Weapons Get Spotlight in China's Military Parade," *Defense One*, 1 October 2019, https://www.defenseone.com/.

4. Franz-Stefan Gady, "Russia's New Nuclear Torpedo-Carrying Sub to Begin Sea Trials in June 2020," *The Diplomat*, 10 September 2019, https://thediplomat.com/.

5. Paul Scharre, "A Million Mistakes a Second," *Foreign Policy*, 12 September 2018, https://foreignpolicy.com/.

6. Paul Scharre, "Killer Apps: The Real Dangers of an AI Arms Race," *Foreign Affairs*, May/June 2019, https://www.foreignaffairs.com/.

7. See, for example, Matt Turek, "Explainable Artificial Intelligence (XAI)," Defense Advanced Research Projects Agency, accessed 9 October 2019, https://www.darpa.mil/.

8. Brendan I. Koerner, "Inside the Cyberattack That Shocked the US Government," *Wired*, 23 October 2016, https://www.wired.com/.

9. Executive Order 13859 of 11 February 2019, Maintaining American Leadership in Artificial Intelligence, 84 Fed. Reg. 3967–3972 (19 February 2019), https://www.federalregister.gov/.

10. Samm Sacks, "New China Data Privacy Standard Looks More Far-Reaching Than GDPR," Center for Strategic and International Studies, 29 January 2018, https://www.csis.org/; Rogier Creemers, Paul Triolo, and Graham Webster, "Translation: Cybersecurity Law of the People's Republic of China (Effective June 1, 2017)," *DigiChina* (blog), *New America*, 29 June 2018, https://www.newamerica.org/; and Benny Bogaerts and Kara Segers, "The 'Localisation' of Russian Citizens' Personal Data," KPMG, 5 September 2018, https://home.kpmg/.

11. See, for example, Matthew Newton and Julia Summers, "Russian Data Localization Laws: Enriching 'Security' & the Economy," The Henry M. Jackson School of International Studies, University of Washington, 28 February 2018, https://jsis.washington.edu/.

12. Ronak D. Desai, "India's Data Localization Remains a Key Challenge for Foreign Companies," *Forbes*, 6 October 2019, https://www.forbes.com/; "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance)," Pub. L. No. 32016R0679, 119 OJ L (2016), *Official Journal of the European Union*, http://data.europa.eu/; and "Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a Framework for the Free Flow of Non-Personal Data in the European Union (Text with EEA Relevance)," Pub. L. No. 32018R1807, 303 OJ L (2018), *Official Journal of the European Union*, http://data.europa.eu/.

13. Office of the United States Trade Representative, "Fact Sheet on 2019 National Trade Estimate: Key Barriers to Digital Trade," March 2019, https://ustr.gov/.

14. Nevena Simidjiyska, "CFIUS Flexes New Muscles Where Customer Data and Critical Technology Are Involved," *Corporate Compliance Insights*, 24 April 2019, https://www.corporatecomplianceinsights.com/.

15. Eric Rosenbach and Katherine Mansted, "How to Win the Battle over Data," *Foreign Affairs*, 17 September 2019, https://www.foreignaffairs.com/.

16. Anish Athalye et al., "Synthesizing Robust Adversarial Examples," *arXiv:1707.07397v3* [*Cs.CV*], 7 June 2018, http://arxiv.org/.

17. A. J. Bose and P. Aarabi, "Adversarial Attacks on Face Detectors Using Neural Net Based Constrained Optimization," in Institute of Electrical and Electronics Engineers (IEEE), *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, Vancouver, BC, 29–31 August 2018 (Piscataway, NJ: IEEE, 2018), 1–6, https://doi.org/10.1109/MMSP.2018.8547128.

18. Sherman Kent, "A Crucial Estimate Relived," *Studies in Intelligence* 8, no. 2 (Spring 1964): 1–18, posted to CIA Library website 19 March 2007, https://www.cia.gov/.