Executive Summary: Building Trust through Testing

Michèle Flournoy  
Co-Founder, WestExec Advisors, Washington, DC

Avril Haines  
Former Assistant and Advisor to President Obama and Deputy Director of the CIA

Gabrielle Chefitz  
Former Sr. Associate at WestExec Advisors

The United States is at an inflection point in an age of mounting transnational threats, unprecedented global interdependence, and resurgent great power competition. This moment is taking place in the context of a technological revolution that exacerbates the challenges we face while simultaneously offering potential solutions, providing breakthroughs in climate, medicine, communications, transportation, intelligence, and many other fields. Many of these breakthroughs will come through the exploitation of artificial intelligence (AI) and its related technologies – chief among them machine learning (ML). These advances will likely shape the economic and military balance of power among nations and the future of work, wealth, and inequality within them.

Innovations in ML have the potential to fundamentally transform how the US military fights and the Department of Defense (DoD) operates. Machine learning applications can increase the speed and quality of human decision-making on the battlefield, enable human-machine teaming to maximize performance and minimize the risk to soldiers, and greatly improve the accuracy and speed of analysis that relies on very large data sets. ML can also strengthen the United States’ ability to defend its networks against cyber attacks at machine speeds and has the power to automate critical components of labor-intensive enterprise functions, such as predictive maintenance and personnel management.

Advances in AI and machine learning are not the sole province of the United States, however. Indeed, US global leadership in AI remains in doubt in the face of an aggressive Chinese challenge in the field. Numerous DoD and academic reports reflect on the need to invest more in AI research and development, train and recruit a skilled workforce, and promote an international environment supportive of American AI innovation – all while promoting safety, security, privacy, and ethical development and use. However, far too little attention is placed on the issue of trust, and especially in testing, evaluation, verification, and validation (TEVV) of these systems. Building a robust testing and evaluation ecosystem is a critical component of harnessing this
technology responsibly, reliably, and urgently. Failure to do so will mean falling behind.

This report will first highlight the technological and organizational barriers to adapting DoD’s existing TEVV ecosystem for AI-enabled systems, with a particular emphasis on ML and its associated techniques of deep learning (DL), which we predict will be critical to future deterrence and warfighting while presenting unique challenges in terms of explainability, governability, traceability, and trust. Second, this report will offer concrete, actionable recommendations to DoD leadership, working with the intelligence community, the State Department, Congress, industry, and academia on how to advance the TEVV system for ML/DL by reforming processes, policy, and organizational structures while investing in research, infrastructure, and personnel. These recommendations are based on the authors’ decades of experience working in the US government national security and dozens of interviews with experts from government, industry, and academia working on ML/DL and test and evaluation.

New Technologies Require New Testing Approaches

The Defense Department needs to reform its existing testing and verification system—its methods, processes, infrastructure, and workforce—in order to help decision-makers and operators understand and manage the risks of developing, producing, operating, and sustaining AI-enabled systems. Several DoD reports and policy documents identify TEVV as a barrier to AI adoption and call for increased research into new methodologies, including the Pentagon’s AI Ethics Principles1 and AI Strategy,2 which states, “we will invest in the research and development of AI systems that are resilient, robust, reliable, and secure; we will continue to fund research into techniques that produce more explainable AI; and we will pioneer approaches for AI test, evaluation, verification, and validation.”

However, DoD has yet to translate this stated goal into a real plan of action. Advancing the Defense Department’s TEVV enterprise for ML/DL systems is critical for several reasons.

First, developing an effective TEVV approach that is sufficiently predictive of performance is critical to building the trust in these systems necessary to deploy and leverage these capabilities at scale. The United States has already seen this dynamic with nuclear power, for example, where lost trust in the technology has prevented policymakers from harnessing nuclear power for clean energy.

The Pentagon cannot let TEVV become a barrier to fielding AI-enabled systems in an operationally relevant time frame but must do so in a manner that engenders trust in such systems and is consistent with US values and principles. The ultimate goal of any TEVV system should be to build trust—with a commander who is responsible for deploying a system and an operator who will decide whether to delegate a task to such system—by providing relevant, easily understandable data to inform decision-making.

Fielding AI systems before our competitors may not matter if DoD systems are brittle and break in an operational environment, are easily manipulated, or operators consequently lose faith in them. Military operations present a challenging environment. The Defense Department needs ML/DL systems that are robust and secure. They need to be able to function in a range of environmental conditions, against adversaries who are adaptive and clever, and in a manner that engenders trust by the warfighter.

Second, the context in which DoD operates means these technologies are prone to adversary attack and system failure, with very real consequences. Machine learning systems have an increased potential for failure modes relative to other systems, such as bias due to a distribution shift in data, as well as novel vulnerabilities to attacks ranging from data poisoning to adversarial attacks. One could easily imagine an image classifier that accidentally classifies a civilian school bus as a tank or an adversary exfiltrating a model processing sensitive intelligence, surveillance, and reconnaissance or communications data. Image classification algorithms developed for one environment (e.g., the desert) could turn out to work incorrectly in another environment (e.g., cities).

Third, with an effective TEVV system, the United States can reduce barriers to innovation and facilitate US leadership in ML/DL technologies. As most of the innovation in ML/DL will come from the private sector, unless the US government is able to effectively draw on private sector work in this arena, it will not be able to leverage the best cutting-edge technology. Research on new TEVV methods and organizational reforms to adapt the current system is simply not keeping pace with private sector development. Without urgent reforms and prioritized investment in new research and infrastructure, the Defense Department will lose its chance to shape industry’s approach to ML/DL development in a manner consistent with DoD standards for safety, reliability, and accountability. It will lose the opportunity to take advantage of new private sector
developments, while allowing other nations without such standards to adopt the latest innovations. It is critical that the US government not only shape its own US industry standards but also promote compatible global standards and norms.

Fourth, adversary advancements will likely increase pressure to field AI-enabled systems faster, even if testing and assurance are lacking. China has elevated AI to be a major national priority across sectors and is already exporting armed drones with varying degrees of autonomy. Russia is also pursuing research and development (R&D) on AI for military purposes and fields AI-enabled robotic systems in Syria with little regard for ethical considerations. However, it shouldn’t be a race against our competitors to field AI systems at any cost. It’s a race to field robust, lawful, and secure AI systems that can be trusted to perform as intended.

Finally, high standards for robustness, assurance, interpretability, and governability can ultimately be a tremendous source of strategic advantage, incentivizing industry to harden systems to adversary attack.

Taken together, these risks and opportunities suggest that devising an effective, efficient, and ethical TEVV process is critical for maintaining the US military and economic competitive edge, as well as deploying reliable and trustworthy ML/DL systems.

Recommendations for Adapting DoD’s TEVV Enterprise for AI/ML

1. Create an Office of the Secretary of Defense (OSD) coordinating body to lead on AI/ML TEVV and incentivize strong cooperation with the services.
2. Invest in priority areas of research in partnership with industry and academia.
3. Develop a tailored, risk-based framework for ML/DL testing and safety.
4. Translate the testing framework into testable, verifiable requirements to be used by the private sector and build an integrated team to leverage this approach.
5. Bridge the gap between development and testing.
6. Increase and integrate spending for T&E research and infrastructure.
7. Develop industry/US government TEVV standards and promote them internationally.
8. Test, train, and certify human-machine teams through wargaming, simulation, and experimentation.
10. Increase resources for and attention on adversarial testing and red-teaming.
11. Promote greater cooperation on ML/DL between DoD and the Intelligence Community (IC).

Conclusion

The future of US leadership on ML/DL and DoD’s ability to harness these critical technologies depends on DoD investing in the science of ML/DL TEVV to develop new approaches and metrics, as well as standing up the coordination and governance mechanisms to accelerate progress and scale solutions. It will require developing the testing frameworks, requirements, and standards to bridge the gap between industry and government and shape a more iterative development and testing approach; shifting culture and practice toward the testing and certification of human-machine teams; and securing the talent, infrastructure, and resources to implement this new approach. Finally, DoD will need to deepen partnerships with the private sector, academia, non-governmental organizations, international organizations, and international partners to realize a multi-stakeholder approach to ML/DL development, testing, and deployment.

Adapting the TEVV enterprise for ML/DL is critical to increasing trust in and, consequently, accelerating the deployment of these systems on a timeline consistent with the rate of innovation, operational need, and US ethics and principles. The steps DoD and the broader US government take now to adapt the ML/DL testing ecosystem will determine the long-term safety, reliability, and relevance of these systems in the coming decades.

MICHÈLE FLOURNOY is Co-Founder and Managing Partner of WestExec Advisors, and former Co-Founder and Chief Executive Officer of the Center for a New American Security (CNAS), where she currently serves on the Board. Michèle served as the Under Secretary of Defense for Policy from February 2009 to February 2012. She was the principal advisor to the Secretary of Defense in the formulation of national security and defense policy, oversight of military plans and operations, and in National Security Council deliberations. She led the development of the Department of Defense’s 2012 Strategic Guidance and represented the Department in dozens of foreign engagements, in the media and before Congress. Michèle earned a bachelor’s degree in social studies from Harvard University and a master’s degree in international relations from Balliol College, Oxford University, where she was a Newton-Tatum scholar.
AVRIL HAINES served as Assistant to the President and Principal Deputy National Security Advisor to President Obama and as Deputy Director of the Central Intelligence Agency. Avril also held a number of senior legal positions in the government, including Legal Adviser to the National Security Council and Assistant Legal Adviser for Treaty Affairs. Avril received her bachelor’s degree in Physics from the University of Chicago and a law degree from Georgetown University Law Center.

GABRIELLE CHEFITZ previously served as Senior Associate at WestExec Advisors. Prior to her work at WestExec, she was a Research Assistant to the Director of the Project on the Middle East Peace Process at the Washington Institute for Near East Policy and worked at the Senate Foreign Relations Committee. Gabrielle earned a master’s degree in Public Policy from the Harvard Kennedy School of Government and a bachelor’s degree from Northwestern University’s Medill School of Journalism.

Endnotes