SYM-AM-26-069

Excerpt from the Proceedings

of the

Twenty-Third Annual Acquisition Research Symposium and Innovation Summit

Wednesday, May 6, 2026 Sessions

Volume I

"Accelerating Warfighting Capabilities"

The Pentagon's Revolution in Software-Defined Warfare and Its Testing Dilemma

Published: April 30, 2026

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program, Graduate School of Defense Management at the Naval Postgraduate School.

To request defense acquisition research, please contact:

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
E: arp@nps.edu
www.acquisitionresearch.net

Copies of Symposium Proceedings and Presentations; and Acquisition Sponsored Faculty and Student Research Reports and Posters may be printed from the **NPS Defense Acquisition & Innovation Repository** at **https://dair.nps.edu/**.

ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF ACQUISITION, FINANCE, AND MANPOWER
NAVAL POSTGRADUATE SCHOOL

The Pentagon's Revolution in Software-Defined Warfare and Its Testing Dilemma

Douglas C. Schmidt—is the Dean of William & Mary's School of Computing, Data Sciences & Physics, after spending more than 20 years as a CS professor at Vanderbilt University. He recently served as the Pentagon's Director, Operational Test and Evaluation and was a program manager at DARPA and the Chief Technology Officer at Carnegie Mellon University's Software Engineering Institute. He has published many papers and books on software-related topics. He received Bachelor and Master of Arts degrees in Sociology from William & Mary and Master of Science and Doctorate degrees in Computer Science from the University of California, Irvine. [dcschmidt@wm.edu]

Nickolas H. Guertin—is a defense acquisition executive with more than 30 years of experience delivering complex naval, aerospace, and software-intensive systems. He served as Assistant Secretary of the Navy for Research, Development, Acquisition, and Sustainment (2023–2025), overseeing a $130 billion portfolio and leading 130,000 professionals. Previously, as Director of Operational Test and Evaluation, he provided independent assessments to Congress and the Secretary of Defense. A former enlisted submariner and Navy Reserve Engineering Duty Officer, he is now a Virginia Tech Senior Research Fellow, advisor, and published thought leader in acquisition reform, digital engineering, and open systems. [nickolashg@vt.edu]

John E. Robert—is a Principal Engineer at Carnegie Mellon University's Software Engineering Institute (SEI) and Deputy Director of the Software Solutions Division, leading software engineering R&D and Navy program support teams. He co-authored Architecting the Future of Software Engineering: A National Agenda for Software Engineering Research & Development, a multi-year roadmap for next-generation software-reliant systems. Previously, he served as a Department of the Navy civilian in the distributed simulation team at the Naval Air Warfare Center Aircraft Division (NAWCAD). [jer@sei.cmu.edu]

Abstract

Warfare is inherently messy and adaptive—Sun Tzu's observation that "all warfare is based on deception" remains relevant—but today's tempo of capability delivery is outpacing hardware-centric acquisition and legacy warfighting patterns. This paper argues that military preeminence increasingly depends on software-defined warfare, where code—not platforms—becomes the decisive differentiator. We characterize this shift through six tenets: rapid adaptability, AI-driven decision support, digital twins and simulation, reprogrammable weapons, autonomous systems, and cyber operations. Together, these tenets demand unprecedented operational agility, enabling forces to reconfigure tactics, platforms, and effects during conflict.

The same features that enable overmatch also introduce fragility: tightly coupled "kill webs," vulnerabilities in AI reasoning, and the risk of cascading failure from a single software update. This creates a central Pentagon dilemma: software-enabled capabilities can be fielded faster than they can be objectively assessed. Traditional test and evaluation (T&E), optimized for static hardware designs, is straining under continuous updates and complex interdependencies. We propose a reinvention of T&E—supported by digital twins, AI-augmented testing, DevSecOps pipelines, and independent oversight—and offer recommendations to balance rapid innovation with assurance so software-defined arsenals remain agile and dependable in the fog of war.

Introduction

At a desert airstrip, dozens of autonomous drones lift off in unison—a buzzing swarm that embodies the Pentagon's emerging operational concept and supporting acquisition reforms of software-defined warfare. Unlike past eras defined by hardware-centric superiority, today's battlespace is increasingly shaped by the code that runs in these systems. In this emerging paradigm, software upgrades, artificial intelligence (AI) models, and digital reconfigurations—in addition to new airframes, hulls, ground vehicles, and the systems they carry—dictate the pace

of adaptation. U.S. Navy warships now received combat software patches mid-deployment, uncrewed aerial systems (UAS) can rapidly be retasked with new behaviors, and AI tools can generate operational plans or prototype code in hours rather than months. These advances collectively promise unmatched agility: the ability to reprogram the rapidly evolving military capability as quickly as adversaries shift their tactics.

This revolution is needed to win future battles but also includes new challenges. In a software-defined approach, new defense system capabilities are identified based on analysis of operational data, and new capabilities are created and integrated into systems (McNamara, 2025; Mulchandani, 2022). Each tenet of software-defined warfare shown in Figure 1 also introduces enormous, and currently unavoidable, complexity into the defense ecosystem. Systems once evaluated in isolation are now interdependent nodes in sprawling kill webs. A single line of faulty code, a misaligned interface, or an unanticipated AI decision can ripple through the joint force with cascading consequences. The very features that make software-defined warfare so powerful—speed, malleability, scale—also require technology insertion at the speed of relevance (DoW, 2026) and rethinking the testing and modernization activities (DoD, 2025).

The revolution in software-defined warfare is also why the Pentagon now faces a profound testing dilemma (Schmidt & Guertin, 2025). While development and deployment cycles accelerate, the processes to verify and validate these capabilities have not kept pace. Traditional test ranges, certification checklists, and siloed evaluations were not designed for swarming drones, cyber-resilient kill webs, or continuously updated digital twins. Without rigorous and independent test and evaluation (T&E), the United States risks fielding systems that look formidable on paper but falter in the fog of war. Realizing the promise of software-defined warfare requires not only ingenuity in development but also innovation—and renewed discipline—in testing. For example, the Pentagon could increase the synergies between developmental test and evaluation (DT&E) and operational test and evaluation (OT&E) so that validated developmental test (DT) evidence can count toward operational test (OT) findings earlier in system life cycles.

This paper explores how the Pentagon is transforming its acquisition and operational models, and why ensuring trust in these software-driven systems may prove the hardest battle of all. The remainder of this paper is organized as follows: Software-Define Warfare: Key Tenets elaborates on the tenets of software-defined warfare; A Revolution in How the Pentagon Builds Technology examines AI-augmented development and accelerated acquisition practices; Fielding Faster Than We Can Test analyzes the widening gap between rapid fielding and the pace of testing; Trust, Technology, and the Testing Tightrope highlights both the enduring importance of independent comprehensive test and evaluation, along with both the promises and pitfalls of automation; Risks of Undertested Warfare describes key risks of under-tested warfare; Recommendations: Building a Test Enterprise for Software-Defined Warfare offers recommendations on how to build a resilient test enterprise for the software-defined warfare era; and Concluding Remarks presents concluding remarks.

Figure 1. Key Tenets of Software-Defined Warfare

Software-Defined Warfare: Key Tenets

Software-defined warfare is an emerging paradigm where software rather than hardware dictates the pace, form, and outcome of modern conflict (Clark et al., 2023). This paradigm will require generalizing and expanding our thinking of military programs. We must now reorient our actions to consider the composite capability as a larger system (or more likely a system-of-systems) and how it will be built and tested over many iterations of new functionality. These products must interact across kill-webs and embrace a different design practice of fluid integration to deliver new performance. In this way, militaries can now evolve warfighting capabilities rapidly via code into targeted sub-systems (and systems). Scaling and expanding these concepts across deployment nodes for associated mission capabilities requires fundamental changes or key tenets of software-defined warfare, as shown in Figure 1 and described below:

- **Rapid adaptability—**Weapon systems and critical supporting information, like user training and updates to tactics, can be reprogrammed or updated in hours, days, or weeks rather than over decades-long development cycles. Software updates, new AI models, or modular code allow forces to swiftly adjust to emerging threats or missions. For example, a warship's critical systems can *"get smarter overnight with a download"* as seen when the Navy updated its Aegis combat software at sea in 2024 (Eckstein, 2024). This type of agility, performed as a normal course of events, means battlefield advantages can be gained (or lost) through fast code deployment, and gain rapid mission impact through multi-platform capabilities, placing a premium on the ability to test updates quickly and confidently.
- **AI-driven decision support**—AI helps process the flood of sensor and intelligence data to accelerate targeting, logistics, and command decisions. Machine-learning algorithms serve as copilots, offering recommendations or automating analysis for human commanders. In practice, AI-driven battle management might flag threats or suggest courses of action within seconds. The challenge for testers is validating that these AI systems are reliable and bias-free under the pressures of real operations. Ensuring that AI makes sound recommendations on a jammed, contested, chaotic environment—be it in grey-zone conflict or full-scale war—requires extensive scenario testing (Goode et al., 2025) and new T&E methods (e.g., red-teaming the AI's decision logic; Freeman et al., 2025).
- **Digital twins and simulation**—Militaries increasingly rely on high-fidelity software models—digital twins—to evaluate new concepts of operation, predict outcomes, and rehearse

missions through live-virtual-constructive environments. When continuously updated, these simulations can provide commanders with predictive insight by exploring “what-if” scenarios without real-world risk. The Pentagon envisions using such joint battlespace simulations to assess tactics and system integration in silico. However, their value depends entirely on fidelity. Test agencies must verify that digital twins accurately reflect operational reality, or they risk producing overly optimistic results that mislead decision-makers (AIAA ,202). Simulation is a powerful complement to live testing, but it requires rigorous validation and, given current technological and budget constraints, cannot fully replace real-world trials.

- **Reprogrammable weapons systems**—Missiles, drones, ships, and tanks are becoming software-reliant systems whose capabilities can shift through code rather than hardware changes. For example, a single UAS can move from reconnaissance to electronic attack by loading different software. Replicator swarms illustrate this flexibility: their behaviors and targeting priorities evolve through updates, not redesigns. This agility offers significant operational advantages but demands that each new release or configurable module truly adds value in the field. AI-augmented development accelerates reconfiguration and customization, increasing the need for tailored testing to ensure reliability. Testers must validate these rapidly changing systems under combat conditions, confirming that enhancements do not introduce instability or unintended interactions.
- **Autonomous and semi-autonomous systems**—UAS swarms, uncrewed surface/undersea and robotic ground vehicles, provide an expanding range of systems now operating with minimal human intervention, guided by software-based rules of engagement. These autonomous and semi-autonomous weapons adapt in real time to sensor inputs and shifting conditions, serving as powerful force multipliers but dramatically increasing testing demands. For example, how can testers ensure a drone distinguishes targets correctly in the fog of war, or that human–AI teams perform reliably under fire? Answering these questions requires new T&E methods and scenarios—live swarm exercises, virtual war-gaming, and stress tests across contested environments, such as GPS jamming, electronic warfare, and adversary deception. No test fully anticipates a wily opponent, so system architectures must also flex rapidly to incorporate responses when new behaviors emerge in the field.
- **Cyber operations as warfare**—Cyberspace is now a battlefield in its own right, where attacks on networks, satellites, and critical infrastructure can be as decisive as kinetic strikes. In software-defined warfare, adversaries may seek to defeat systems through hacking, spoofing, or software supply-chain compromise rather than physical destruction, driving the need for rigorous cybersecurity and resilience testing. Networked weapons and command systems must be hardened and evaluated against realistic cyber-attack scenarios. Modern operational testing increasingly incorporates red-team cyber assaults on systems, networks, and development pipelines alongside physical stress tests. To address these threats, the Pentagon is investing in advanced cyber ranges and simulations (Podnar et al., 2021) to assess system performance under sustained cyber pressure, recognizing that even the most advanced digital capabilities are ineffective if their software can be silently subverted.

Speed, accuracy and scale of new solutions to the fleet and force are necessary for victory (McNamara et al., 2025). Pursuing a strategy of software-defined warfare offers unprecedented adaptability and capability by leveraging software’s malleability. With this promise comes heightened complexity since success depends on whether these software-driven capabilities work as intended in the chaos of war. Each of the above tenets introduces new dimensions for T&E, from assuring AI decision aids, to validating sprawling simulations, to stress-testing autonomous behaviors and cyber defenses. The following sections examine how the Pentagon’s development and acquisition practices are evolving in this software-centric era,

and why rigorous, innovative, and independent testing is critical to realizing the potential of software-defined warfare.

A Revolution in How the Pentagon Builds Technology

The U.S. military must advance beyond upgrading individual program-of-record systems—it now must reinvent how it conceives, develops, coordinates, and fields its arsenal. At the core of this concept is software-defined warfare: a reoriented focus on adaptability through code, powered through AI-driven decision-making, supported by modular upgrades, and continuous improvement in performance of forces. What once required decades of hardware cycles can now happen in months or even weeks through software-driven iteration. This section highlights two of the most dramatic shifts—AI-augmented development and accelerated acquisition—showing how they operationalize the tenets of software-defined warfare while creating urgent new challenges for T&E.

AI-Augmented Development: Speeding the Software Cycle

A key driver of software-defined warfare is the integration of AI directly into the development process—an embodiment of software-defined warfare's principle of *rapid adaptability*. Generative AI and machine-learning tools now help design systems, modernize legacy code, and even draft acquisition documents. The U.S. Army's experimental "AI Flow" program, for instance, can generate an acquisition document 80% complete in minutes—a task that once took weeks (Fox, 2024). In 2023, the Pentagon underscored this shift with an $800 million contract for advanced large language models (LLMs; CDAO, 2025). From automating code and documentation to shaping design decisions, AI copilots are accelerating delivery and embedding speed and flexibility into every stage of the cycle.

AI's role is expanding beyond writing code to actively supporting testing and evaluation, as shown in Figure 2. Here, AI-driven tools extend the digital twin concept by using validated data to create simulations of complex systems and highlight vulnerabilities in real time. Early applications include automatically generating test cases from requirements, tracing system dependencies, and surfacing compliance gaps that might otherwise go unnoticed (Bain, 2025). By producing dozens of scenarios or code fixes rapidly, AI has the potential to accelerate delivery while catching subtle flaws that human testers could miss. This capability points toward a future where T&E itself becomes iterative and adaptive, evolving at the same speed as the software-defined warfare systems it safeguards.

Figure 2. Applying Digital Twins to Accelerate Software Development and Testing

Human experts (including testers) must therefore learn new workflows to collaborate with AI assistants, double-check AI outputs, and guard against failure modes introduced by AI-created

code or data. For instance, if an LLM recommends a code change, who ensures that the change is secure and reliable? The burden falls on T&E professionals to develop new verification frameworks suited to AI-generated software—frameworks that may involve algorithmic transparency, adversarial stress tests, and continuous monitoring (Schmidt, 2025). In short, AI has turbocharged the Pentagon's software cycle, aligning with software-defined warfare's tenet of rapid adaptability, but it also amplifies the imperative for agile testing at machine speed.

Accelerated Acquisition and Agile Deployment

Another pillar of the Pentagon's software revolution is reforming how systems are acquired and deployed, treating individual platforms as components of larger, *reprogrammable capabilities* that span software-defined weapons and cyber warfare. Traditional acquisition cycles take years, yet modern capabilities—assembled from distributed subsystems across the battlespace—must evolve in hours, days, or weeks, creating a fundamental mismatch. To address this gap, the Pentagon has begun using flexible contracting authorities and agile methods to dramatically shorten acquisition timelines (AFLCMC, 2025).

For instance, program offices are using the authorities for using Other Transaction Authority agreements and Commercial Solutions Openings to bypass the usual red tape and reach non-traditional vendors (Gallium Solutions, 2025). Under Project Replicator, the Pentagon awarded contracts for new drone prototypes in as little as 110 days from concept to award, compared to the 12–18 months typical of older processes (Tucker, 2025). This pace matches the logic of software-defined warfare: field more adaptable, updatable systems that can be kept up to date faster than adversaries can counter them. It can be done today with significant effort for high priority mission capabilities—but the whole environment must be reoriented to delivering rapid capability at scale instead of systems.

Defense acquisition teams are also experimenting with LLM-based assistants and automation throughout the procurement and integration pipeline (Magnuson, 2025). By generating interface code, drafting requirements, and even allocating software tasks, these tools echo the *AI-driven decision support* described in the Software-Defined Warfare: Key Tenets section. Combined with iterative DevSecOps methods, they allow the Pentagon to push out software updates and modular hardware prototypes at a pace unthinkable a decade ago. This adaptability makes the force smarter and faster, but it also places immense pressure on T&E processes that must validate these updates in near real time.

Figure 3 depicts this transformation, where the mission systems hosted in combat platforms, from tanks to ships to UAS swarms, now receive over-the-air updates that alter their capabilities rapidly (Sherbinin & Gray, 2025). These AI-enabled systems not only generate tactics and recommend strategies but also deliver software patches that can reconfigure weapon systems at the speed of need, giving U.S. forces agility through code rather than costly and time-consuming hardware redesign. Figure 3 depicts software-defined warfare in action—an adaptable, reprogrammable force that can outpace adversaries by shifting faster than they can respond. Yet the very speed and scale that make this revolution possible also heighten the stakes: without rigorous T&E, rapid reprogramming risks introducing unseen vulnerabilities into the fight.

Figure 3. An Agile Force is a Reprogrammable Force

The surge in speed and quantity of new software-defined warfare systems underscores the Pentagon's testing dilemma. The central question looms: how do we know these rapidly fielded, software-defined tools will actually work in the chaos of war? Contracting and delivering a system quickly is one thing; making sure it will work, with real operators, fully supported, in a tactical environment is another. As the next sections explore, the imbalance between rapid fielding and thorough testing is growing—and with it, the danger that America's software-defined arsenal could prove brittle when it matters most.

Fielding Faster Than We Can Test

This section examines the growing tension between the Pentagon's accelerating software cycle and the slower pace of assurance. Software-defined warfare promises rapid adaptability, autonomy at scale, and reprogrammable weapons updated overnight—but each breakthrough multiplies the necessary action of verification and validation. Fielding cycles now measured in weeks outstrip test capacity, forcing greater reliance on models and new methods even as interdependent, software-updatable systems flood the force. The hardest challenge lies in proving interoperability and resilience across the joint "kill web"—autonomy, swarms, electronic warfare, cyber, and over-the-air updates—where failures often surface at the seams.

The Growing Imbalance Between Fielding and Testing

Even as military arsenals evolve at accelerated speed, the T&E mechanisms needed to validate those new tools are struggling to keep up. According to a recent Defense Science Board (DSB) study, the gap between fielding and testing is not closing but widening, with potentially serious consequences (Evans et al., 2024). Officials and experts are increasingly worried about a mismatch where the Pentagon can deploy new software-defined systems in months but cannot test them thoroughly in time. The very tenet of *rapid adaptability* means capabilities can be reprogrammed at will, yet each change requires verification and an appreciation that other things didn't break when a fix is introduced, which is increasingly hard to assure at the same pace. Accidental introduction of new bugs with software updates is a known issue and requires continuing focus on resilient architectures, automated processes, and regression testing to reduce the risks (DOT&E, 2024).

The DSB study also found that many emerging technologies "often require development of new analytical tools and testing capabilities for verifying and validating system performance," and in some cases physical testing might not even be possible, forcing greater reliance on modeling and simulation study (Evans et al., 2024]) This reliance connects directly to the *digital*

twin concept of software-defined warfare. Models and simulators may allow updates to be stress-tested virtually, but unless they are themselves validated, they risk creating false confidence. This DSB study emphasized that to test leap-ahead technologies, such as hypersonic missiles, directed-energy weapons, or AI/ML systems, the Pentagon must invest in novel verification methods—and validate those methods themselves—to fill gaps where traditional live testing falls short, as discussed in the case study of the Joint Simulation Environment (JSF) shown on the previous page.

Case Study of the Joint Simulation Environment.

The F-35 Joint Simulation Environment (JSE) stands as both an essential achievement and a cautionary example. It ultimately delivered sophisticated, physics-based testing capability needed to evaluate combat scenarios that cannot be affordably or safely executed on an open range, as shown below.

However, its development demonstrated how hard it is to build a simulation environment that decision-makers can trust—one that accurately reflects real-world data, supports dynamic interaction, and produces results credible enough to inform combat-readiness decisions. Achieving this level of fidelity requires rigorous upfront planning and a clear understanding of the underlying physics, data dependencies, and system integration challenges.

The initial JSE effort underestimated this complexity and began with an incomplete vision of what a credible environment required (DOT&E, 2023), leading to abandonment and restart of the original design. The reconstituted JSE ultimately succeeded, enabling the F-35 program to reach full-rate production and establishing a gold standard for validated warfighting simulation (Naegele, 2024). However, it became operational only after more than 1,000 aircraft had already been produced under incremental acquisition authorities. Consequently, issues discovered during IOT&E—though correctable—were more expensive to resolve and introduced greater early operational risk than if identified earlier in the development life cycle (Guertin, 2022).

Without significant changes in technology, workforce, and oversight, imbalance between fielding and testing could undermine both safety and effectiveness. The reality today is that dozens of AI-driven drones, autonomous vehicles, and software applications are pouring into the joint force each year—an embodiment of the autonomy and reprogrammable weapons systems tenets. Each may be deployable, but can the Pentagon ensure that they are safe, effective, and interoperable? As another DSB study on applying AI in DoD missions starkly put it: if that trend continues, how can the Pentagon ensure that each system, much less the whole force, will work reliably under combat stress (DSB, 2023).

Unlike earlier eras when a single platform could be tested in isolation, today's systems-of-systems are woven into sprawling, interdependent networks. For example, a drone's effectiveness may rely on satellite feeds, AI-enabled command centers, and/or coordination with swarms of other autonomous assets. These dependencies are the essence of the Pentagon's Joint All-Domain Command and Control (JADC2) concept, which involves an intricate "kill web" linking sensors, shooters, and decision nodes across domains. As shown in Figure 4, unvalidated or untested interactions within this web can introduce hidden vulnerabilities or faulty

assumptions of behavior, creating risks that ripple through the entire network. Testing one component in isolation is no longer sufficient—testers will now need to understand the behavior of the entire web of connections is order to assess effectiveness, a challenge that grows exponentially with every new link.

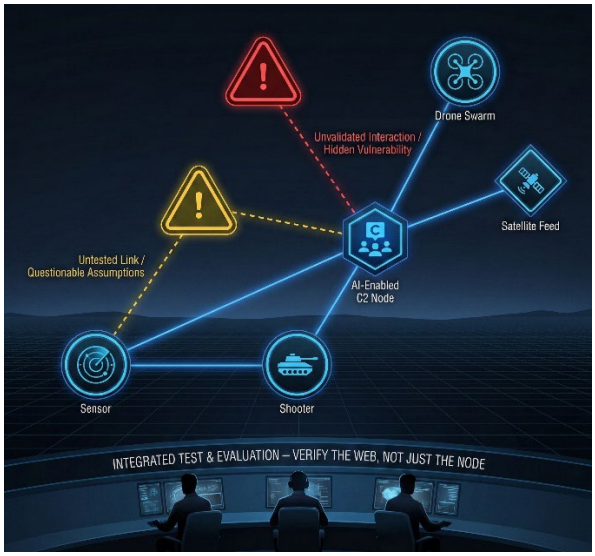

Figure 4. Unvalidated Connections Create Unseen Vulnerabilities in JADC2 Kill Webs

The bottom line is that the U.S. military's ability to rapidly field new software-defined capabilities now outpaces its capacity to test them under realistic conditions. This imbalance creates a risk of deploying systems that are not fully vetted for the rigors of combat. Recognizing this, defense analysts are emphasizing the need to cover the gap between the rate of fielding new systems and the capacity to test them thoroughly (Schmidt & Guertin, 2025).

Testing Challenges in the Era of Kill Webs and Autonomy

Ensuring that a single new weapon works as designed is hard enough; ensuring that hundreds of new, software-defined systems all work together, each system also possibly being updated, in a dynamic battlespace is a monumental challenge that doesn't scale with traditional testing methods. The Pentagon's push for integrated, all-domain operations means that victory rests on cohesion and interoperability, e.g., Air Force systems must seamlessly coordinate with Army, Naval, Space Force, and allied systems in real time. This ambition reflects the *autonomy* and *cyber-as-warfare* tenets of software-defined warfare but also magnifies the difficulty of verification.

Recent experience underscores the difficulties. In one experiment involving Project Overmatch, a capability was fielded rapidly via over-the-air updates (a textbook example of *reprogrammable weapons systems*) but validating its interoperability at scale proved complex and time-consuming (DOT&E, 2024). Large-scale joint exercises and integrated tests—where multiple services stress their systems together—are logistically hard, expensive, and often cut for schedule or budget reasons. Under tight timelines, many rapid acquisition programs simply lack operationally realistic test environments, undermining the *rapid adaptability* that over-the-air updates were meant to provide.

Program offices or contractors, eager to deliver quickly, sometimes choose lower programmatic cost or schedule risk at the expense of higher risk-to-operational failure. This tendency is especially true for integration efforts and how systems interact with each other as a harbinger of broken schedules and unfulfilled promises. These shortcuts are particularly risky in

the context of software-defined warfare, where each system's function may depend on another's data or algorithms. The result is that systems reach deployable status with only limited understanding of their true performance in a system-of-systems context. Warfighters may be handed new drones, sensors, or AI tools that individually passed lab tests, but no one knows for sure if they will mesh when plugged into the joint kill web.

This challenge is not merely hypothetical—the gap between lab success and battlefield reality has already been exposed. Troops have fielded new technologies only to discover too late that a drone could not link with a Navy ship or that an AI guidance system collapsed under electronic jamming (Albon, 2025). As shown in Figure 5, swarms that appear flawless in controlled test environments can unravel amid the chaos of a contested battlefield, where cyber-attacks and spectrum interference often degrade communications and coordination. Failures like incompatible radios and software crashes underscore the danger of testing *autonomy* and *AI-driven decision systems* only under ideal conditions. Building resilience for dozens of interconnected agents demands joint, all-domain trials at scale—exercises that remain rare despite their necessity.

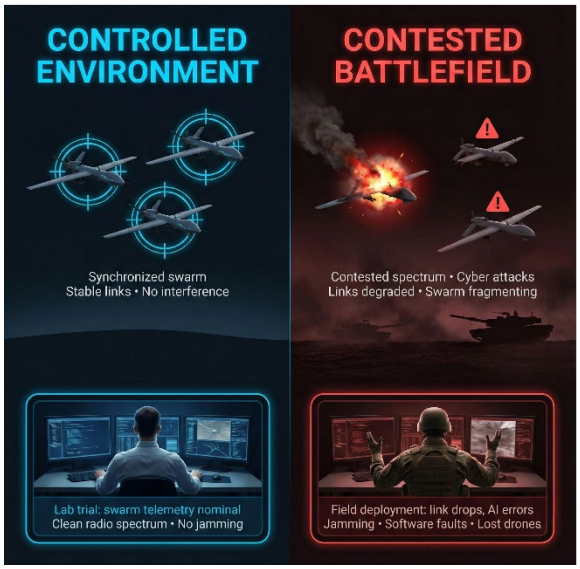

Figure 5. Swarm Success in the Lab, Chaos in the Fight: Why Testing Must Match Reality

To their credit, Pentagon stakeholders are not blind to these challenges. There have been efforts to modernize testing itself—to "test the way we fight," meaning more joint, all-domain evaluations in environments that mirror the drone-saturated, data-driven battlefields of tomorrow (DoD, 2024). The updated operational T&E guidance in DoDI 5000.98 aligns with the *digital twin and simulation* tenet, which envisions more advanced models and automated test tools for AI-enabled capabilities. National defense research teams are exploring how to close the "software understanding gap" (CISA, 2025) and the Armed Services are also investing in exercises like Project Convergence to identify interoperability issues early (South, 2022).

However, these efforts face steep institutional and resource hurdles. Testing in an all-domain, software-defined context demands not only new technology—simulators, distributed ranges, AI-driven test orchestration—but also sufficient skilled people and resources. Defense analysts are observing a troubling inconsistency: at the very moment such broad, innovative testing is most needed (Schmidt & Guertin, 2025). In addition, the Pentagon has dramatically downsized the very office responsible for independent operational test and evaluation across the services (Hegseth, 2025). This irony underscores the theme of this paper: the Pentagon is

charging into the era of software-defined warfare, but unless it invests equally in testing and oversight, speed could become fragility.

Trust, Technology, and the Testing Tightrope

This section explores the paradox at the heart of modern T&E: the Pentagon is best served by preserving independent oversight while also embracing new simulation and automation tools. The stakes high in this era of software-defined warfare, where systems are rapidly adaptable, digitally twinned, reprogrammable, autonomous, and cyber-contested. Unlike traditional weapons platforms, software-reliant systems can change rapidly through updates that introduce new features *and* new risks, making independent T&E critical to operational safety and trust. Without adequate scrutiny, there is a real danger of deploying unproven code or brittle autonomy into contested battlespaces, where the margin for error is razor-thin. This section also examines the promise and peril of digital twins, AI-driven test agents, and large-scale simulation, emphasizing how these tools can expand coverage but are not yet replacements for seasoned testers.

The Urgency to Modernize Independent Test and Evaluation

For decades, verifying the adequacy of test plans and the validity of their results has consistently demonstrated its value. The Director of Operational Test & Evaluation (DOT&E)'s annual reports are replete with examples of programs falling short of delivering needed capability. An independent evaluator of T&E performance and data analysis has proven value in supporting software-defined warfare. An organization like DOT&E and the affiliated Service's operational test agencies ensure that teams responsible for product development are held to high standards for how tests are written and how the results are analyzed.

New approaches are also needed to subject teams developing new weapons systems—and major upgrades to legacy systems—to rigorous external validation to ensure their products are tested under realistic combat conditions and evaluated objectively. This role is essential in the context of software-defined warfare because code can be changed overnight, and the challenges of not having a test plan and the resources to execute it risk deploying unverified updates into combat. In essence, DOT&E's job is to ensure weapons systems are "battle-tested before battle," identifying critical flaws that must be fixed before warfighters' lives are at stake.

Realistic testing is indispensable in an era of *AI-driven decision support* and *autonomous systems*, where critical failure modes often emerge only in contested, unpredictable environments. Meeting this challenge requires enterprise-level resources to design and orchestrate rigorous field events and multi-service exercises. These integrated, cross-domain testing are needed to ensure true interoperability and realize the kill-web vision of software-defined conflict.

Modern testing must validate not just a single platform in isolation, but the complex web of interactions across satellites, ships, aircraft, drones, and ground systems that now define how wars are fought. As shown in Figure 6, illuminating and closing gaps in kill webs is needed to highlight untested links, validate cross-domain connections, and ensure interoperability under real-world stresses. Without this type of holistic system validation, many critical vulnerabilities may remain hidden, e.g., a contractor may certify a drone as mission-ready yet fail to test it under electronic jamming or in coordination with allied assets—precisely the kinds of weaknesses adversaries exploit in cyber-as-warfare environments. Oversight ensures that speed and innovation do not outpace resilience and trustworthiness.

Given the increasing complexity of software-defined, networked systems, independent oversight of testing is more crucial than ever. Yet recently resources for advancing the practice of T&E have been downsized significantly (Hegseth, 2025). However, reducing oversight before

those methods and tools mature runs counter to the very principles of *responsible autonomy* and *digital twin validation* (Fredenburg, 2025).

Figure 6. DOT&E Helps Validate the Webs of War

In the context of software-defined warfare, independent comprehensive testing provides something intangible yet invaluable: *trust*. When DOT&E signs off on a system after rigorous evaluation, military leaders and troops can have confidence that T&E professionals tried hard to break the system before it reached them. That trust is hard to quantify, but easy to appreciate when lives are on the line. If independent oversight is weakened, the Pentagon risks succumbing to optimistic bias or deploying unvetted updates. In a software-defined era, DOT&E is not just an overseer—it is the steward of confidence.

Embracing Simulation and Automation: Opportunities and Risks

Facing an explosion of complex, software-defined systems to test (and with fewer human testers on hand), the Pentagon is understandably looking to technology for help. Advanced simulation environments, digital twins, and automated test tools embody the tenets of digital twins and simulation and rapid adaptability. If implemented properly, they could help bridge the gap between rapid fielding and rigorous evaluation, allowing the Pentagon to scale its testing to match the speed of its software-defined arsenal (Software Engineering Institute [SEI], 2020).

For example, instead of conducting dozens of live-fire exercises to ensure a new UAS works in multiple scenarios, testers could run thousands of virtual trials in a high-fidelity simulator. Such tools could evaluate interoperability by linking virtual instances of an Air Force command system and an Army missile battery. Automated cyber agents could probe vulnerabilities, reflecting the cyber warfare domain, while digital twins could predict system performance after software updates. In theory, these methods could vastly expand what testers can cover, especially with limited time and resources.

However, current reality has not caught up to this vision. The advanced test automation needed for realistic testing remains at relatively low technology readiness levels—essentially prototypes themselves (Guertin, 2022). Many tools remain too immature for large-scale use, a serious liability for reprogrammable weapons that require continuous validation. Without robust research and development (R&D) investment, the promise of trustworthy *simulation* and *automation* remains out of reach.

In practice, building a credible digital twin of a complex weapon system is hard. High-fidelity models demand massive datasets, specialized expertise, and significant computing power to achieve accuracy (Hahn et al., 2023). Even small modeling errors can yield misleading "rosy outcomes" that collapse under combat stress—a risk that is especially acute for *AI-driven decision support*, where flawed assumptions in simulation can translate into catastrophic real-world failures. As a result, the Pentagon cannot simply simulate its way out of live testing until confidence in these models improves substantially.

The over-reliance on automation without human judgment is another risk. While a simulation may appear to provide near-perfect stability and success, the actual reality under combat stress may be quite different, e.g., systems failing, swarms unraveling, and missions collapsing, as shown in Figure 7. In practice, testing is not a simple box-checking exercise. Instead, intuition, creativity, and skepticism are needed that only experienced testers bring. If seasoned professionals are replaced too quickly with immature AI tools, the Pentagon risks gaining neither trusted human insight nor validated digital capability. This danger is especially perilous with *autonomous systems*, whose unpredictable behaviors cannot be fully captured by models. Test tools themselves must therefore be tested and accredited as rigorously as the systems they claim to evaluate.

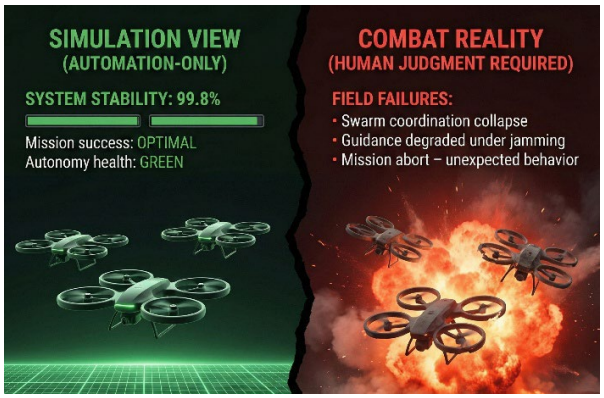

Figure 7. Simulation vs. Reality—the Cost of Over-Reliance

The Pentagon leadership recognizes both the promise and pitfalls here. They argue that a leaner workforce forces innovation—indeed Pentagon leadership and the Armed Services are experimenting with analytics and automated testing to cover more ground with fewer people. Going forward, the most likely outcome is a hybrid approach: human-led design and analysis paired with machine-speed execution of simulations. This hybrid aligns with the human-machine teaming ethos already central to software-defined warfare.

For instance, an AI-driven orchestration system could coordinate distributed simulations across labs and crunch terabytes of test data, spotting anomalies faster than humans alone. This approach could allow small test teams to supervise continuous, large-scale war-gaming of our systems versus simulated adversaries. However, achieving this vision requires both patience and investment. The technologies needed to automate operational testing at scale are still nascent and cutting human testers before it matures risks creating oversight gaps. Thus, replacing seasoned testers faster than reliable automated tools can be fielded is not a recipe for confidence.

In summary, advanced simulations and test automation will be indispensable for tackling the complexity of software-defined warfare. They embody its tenets by offering scalability, adaptability, and speed. Yet in 2025 they remain supplements to—not substitutes for—experienced testers and realistic trials. The Pentagon must walk a fine line: pursuing new

technologies aggressively while acknowledging their current limits. To succeed, Pentagon leaders must remember a core principle: in software-defined warfare, agility without trust is fragility. The next section explains why getting this balance right is so critical, by examining the risks of under-tested systems in the field.

Risks of Under-Tested Warfare

Modern U.S. military strategy assumes that our high-tech weapons will work as intended when called upon. If that assumption fails due to inadequate testing, the consequences could be dire. This section examines several distinct risks that can arise if the military embraces software-defined warfare without sufficiently comprehensive T&E.

Integration Failures in the Kill Web

One risk is that the fragility of integration—the chance that our mosaic of integrated systems needed to support the joint force will not mesh into the coherent kill web envisioned by strategists and war planners. Software-defined warfare thrives on *AI-driven decision support* and *autonomous systems* coordinating across domains. Without robust joint T&E, however, there is no assurance that this myriad of disparate platforms will interact seamlessly.

Figure 8 depicts a central danger in software-defined warfare: peer adversaries won't strike at our strongest systems, but at the fragile seams between them. The Navy ship on one side and the Army missile battery on the other symbolize powerful platforms, yet a hacker gaining entry to the network underscores how integration fragility can undermine them both. AI-driven decision support and cross-domain coordination promise a coherent kill web. If systems are validated only in isolation, however, hidden incompatibilities persist. Without robust joint testing these inter-connections will remain brittle and unseen, exposing vulnerabilities that adversaries are poised to exploit, creating vulnerabilities at the seams.

Figure 8. Adversaries Don't Attack the Strongest Links, They Exploit the Weakest Seams

For example, an adversary exploiting a single untested interface between two Services' systems could turn a seemingly hardened network into a battlefield liability. In an era of software-defined warfare—where capabilities are stitched together dynamically and updated at machine speed—only rigorous, end-to-end joint testing can reveal hidden interdependencies that determine whether kill webs hold or unravel under pressure. Cross-service interoperability trials and large-scale exercises are thus essential, despite their cost and complexity, are indispensable. Without them, the elegant mosaic of interconnected systems we envision and rely

upon risks becoming a brittle patchwork—one that could fracture in the fog and fury of actual conflict.

False Confidence in Untested Systems

Another risk is false confidence—believing a new capability is a significant improvement but later uncover Achilles' heels that only rigorous testing would have exposed. The ethos of rapid adaptability can seduce decision-makers into fielding systems based on demos or contractor reports that fail to reveal the whole story. History is full of examples of weapons that looked promising in controlled trials but stumbled in combat (Wilkinson, 2014).

Figure 9 highlights the illusion of being "ready" in the software-defined warfare era. A swarm of drones may appear mission-capable in controlled tests, but the real proving ground comes when adversaries introduce stressors like jamming, GPS spoofing, or hidden software bugs—the very threats depicted beneath the shield. What appears resilient in peacetime can quickly unravel under hostile conditions, exposing cracks in autonomy and code. Without rigorous testing across these edge cases, the confidence inspired by a glowing "ready" status is dangerously misleading. These types of failures quickly expose the fragility of *autonomous and reprogrammable systems* when not fully tested across edge cases.

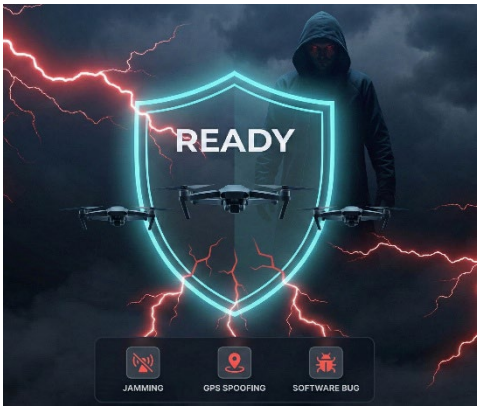

Figure 9. Stress Reveals Hidden Flaws

DOT&E reports are full of such "unknown unknowns," uncovered only during rigorous, realistic testing (DOT&E, 2024). Runtime monitoring and safety guards can reduce these risks, but without a safety net, flaws remain hidden until the battlefield, where they can cause catastrophic consequences. Skipping or truncating operational testing breeds the illusion of readiness—leading commanders and troops to assume systems are reliable when they are not. In software-defined warfare, such misplaced confidence is more dangerous than no capability at all, because it encourages faulty tactical decisions built on brittle foundations.

Eroded Trust and Adoption Delays

Paradoxically, cutting corners on testing can slow down adoption of new technology in the field. Software-defined warfare depends on *human-machine teaming* and trust in *AI-driven decision support*. Yet frontline commanders and troops may hesitate to trust unproven systems with their lives. Trust is built via transparency and rigorous vetting: soldiers gain confidence when they know a system has been tested thoroughly and when they themselves understand its strengths and failure modes.

If a new drone or AI assistant arrives with scant evidence of its performance in contested conditions, operators will naturally be cautious. In well-tested programs, leaders receive detailed reports on what worked, what failed, and what to watch for—establishing a baseline of trust. Without that vetting, doubts linger. Troops might receive powerful new tools but use them

gingerly or not at all, fearing they could backfire, especially for *autonomous decision-making* or *cyber-sensitive systems*.

Figure 10 shows the difference between rushing untested systems into the field versus rigorously validating them before deployment. In the top panel, rushed fielding leaves troops untrained, creates uncertainty, and leads to operational pauses as confidence erodes. Warfighters hesitate to use equipment they cannot trust, undermining the very purpose of *rapid adaptability*. In contrast, the bottom panel shows how rigorous testing earns a DOT&E "approved" stamp, building trust and ensuring systems are combat ready. With confidence in performance instilled by T&E under realistic conditions, warfighter adoption accelerates, and deployment becomes more effective. Speed ultimately flows from trust—and trust in software-defined warfare depends on testing that provides evidence, not assumptions.

Figure 10. Rushed Fielding Yields Delays; Rigorous Testing Builds Trust and Accelerates Deployment

Tactical and Strategic Surprise

It is also essential to consider how the lack of rigorous testing provides adversaries with opportunities to exploit and degrade our capabilities. Our near-peer competitors relentlessly probe for weak links in our digital arsenal. A flaw in a cyber-dependent or software-defined system that slips through testing is a flaw our enemies will eagerly exploit. Figure 11 shows how a single exploited weakness like electronic warfare jamming can ripple throughout an entire kill web. Adversaries constantly search for these fault lines in our networked joint force. A vulnerability in a satellite relay, a radar node, or software in a reprogrammable drone can cascade throughout the system, degrading every connected platform. The lack of rigorous testing doesn't just risk isolated failures—it hands our competitors the opportunity to unravel our whole web of advantage.

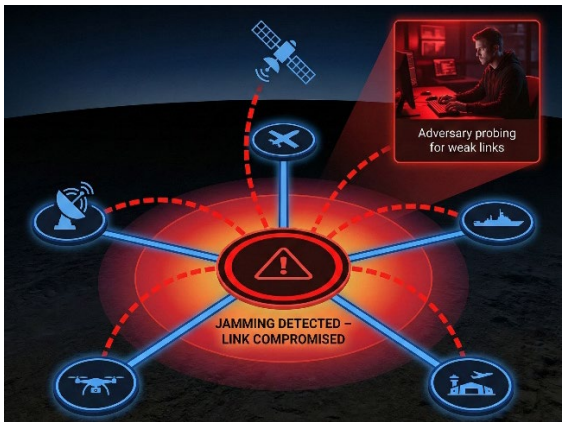

Figure 11. A Single Point of Failure Can Unravel an Entire Kill Web

Thorough testing is how we harden these systems, so our adversaries do not get easy openings. By neglecting testing, we risk strategic surprise. The United States could enter a conflict expecting its software-defined arsenal to dominate, only to see it neutralized—or even turned against us—due to an unvetted defect. The very malleability that makes software-defined warfare powerful also makes it fragile if oversight lags.

In sum, a culture of "field it fast and fix it later" increases operational and safety risk significantly. In contrast, a culture that values rigorous OT&E forces the Pentagon to confront inconvenient truths early, denying adversaries an easy win. The need for rigorous OT&E is especially true in *autonomous and AI-enabled systems*, which adversaries are eager to jam, spoof, or deceive. Only testing across adversary-relevant scenarios can prevent our strengths from becoming liabilities.

Collectively, these risks paint a stark picture: under-tested software yields high-risk warfare. Speed without assurance gambles with lives and strategy alike. The good news is that the Pentagon need not choose between agility and rigor. With the right reforms and investments, it can have both—leveraging the tenets of software-defined warfare for rapid innovation while ensuring those capabilities are battle-ready through robust validation. The next section offers recommendations to achieve this balance and build a resilient test enterprise fit for the digital battlespace.

Recommendations: Building a Test Enterprise for Software-Defined Warfare

To unlock the full potential of software-defined warfare—and avoid the vulnerabilities described in the previous section—the Pentagon must reimagine its T&E enterprise. As shown in Figure 12, this reimagining requires an integrated workflow that combines simulation and digital twins, continuous testing, AI-enabled automation, a collaborative workforce, and a robust and properly resourced DOT&E office that is organizationally independent of systems under test. Together, these elements will form a resilient test ecosystem capable of validating systems at the speed and complexity demanded by modern conflict. This section presents key recommendations for building an enterprise T&E infrastructure and workforce that can support these capabilities effectively in the context of software-defined warfare.

Figure 12. Key Elements in the T&E Enterprise for Software-Defined Warfare

Invest in Advanced Simulation and Digital Twin Infrastructure

The Pentagon should expand and connect its test ranges and labs into a persistent digital proving ground. This investment is central to the *digital twins and simulation* tenet of software-defined warfare, which envisions predictive models that evolve in parallel with production weapons systems. Funding should prioritize high-fidelity simulators and digital twin models for complex systems, networked together to replicate entire kill webs.

For example, a joint simulation might concurrently model an Air Force drone swarm, an Army missile battery, and a Navy command ship to evaluate end-to-end performance. Such environments could run thousands of virtual scenarios, complementing physical tests and stress-testing AI-driven decision support at scale. To ensure credibility, however, validation must be prioritized: digital twins should be compared rigorously to live operational data. A concerted R&D effort—through DARPA, the Test Resource Management Center, and Service labs—is needed to mature these tools from experimental prototypes into trusted T&E assets.

Integrate Continuous Testing into DevSecOps Pipelines

Taking a cue from modern software practice, the Pentagon should adopt a mindset of testing early, often, and automatically across a system's life cycle. An example is using DevSecOps pipelines, an integral part of the preferred Software Acquisition Pathway (Harper, 2025) to automate testing in incremental updates (SEI, 2025) to enable continuous testing as opposed to the testing at the end mindset. This approach directly supports the rapid adaptability and reprogrammable weapons systems tenets, where updates can be deployed quickly and verified immediately. Continuous testing requires embedding automated test suites in DevSecOps pipelines for both software and hardware-in-the-loop. Every new code commit or modular upgrade should trigger regression checks and cybersecurity scans. Whenever feasible, prototypes and updates should be fielded in controlled "beta" capacity that can be rolled back quickly if deviant behavior emerges, capturing operational feedback in near-real time.

An illustrative example would be continuously integrating and testing incremental updates to an autonomous UAS in a sandbox with warfighter participation, rather than deferring all validation to a single T&E event. By the time a capability is fully fielded, it will have already proven itself in multiple iterative trials—ensuring trust in rapid adaptability while reducing risk. This approach would allow operators and warfighters to stress-test the UAS under varied conditions, from electronic warfare interference to extreme weather, surfacing potential weaknesses long before combat deployment.

Leverage AI and Automation as Force Multipliers in Testing

The T&E community should embrace AI tools not as replacements for humans but as amplifiers of their effectiveness. This strategy aligns with the *AI-driven decision support* and *autonomy* tenets of software-defined warfare, where AI agents work alongside humans to manage complexity at scale. AI can rapidly generate test cases, explore edge-condition combinations, and orchestrate distributed test events that would be infeasible to conduct manually (Schmidt, 2025). Generative AI could simulate adversary cyber tactics or electronic jamming against a system under test, exposing vulnerabilities in cyber-as-warfare contexts (Freeman et al., 2025). Machine learning algorithms can process massive test datasets, flag anomalies, and highlight performance trends invisible to humans.

Intelligent orchestration tools could manage networks of test assets—labs, ranges, and virtual environments—running large-scale scenarios overnight while humans review flagged issues in the morning. To realize this vision, the Pentagon should sponsor pilot projects in areas like automated test plan generation, adversarial scenario creation, and anomaly detection. This AI-augmented testing will enable the T&E workforce to scale coverage and focus on higher-order analysis.

Bolster T&E Expertise and a Collaborative Workforce

Technology alone cannot deliver trusted outcomes; people are as essential as machines in the testing enterprise. The human-machine teaming dimension of software-defined warfare requires testers who understand both the software driving modern weapons systems and the operational contexts in which those systems fight. The shortage of experienced testers highlights the need to rebuild this expertise. The Pentagon should prioritize hiring and developing “cyber-warrior testers” with fluency in coding, AI, and mission operations. Recruiting talent from industry, coupled with new military and civilian training pipelines, will accelerate this shift. To capture the wisdom of experienced T&E professionals, the DoD should create cross-agency knowledge exchange programs.

For example, joint T&E task forces could bring together the Service’s test and evaluation agencies in the Army Test and Evaluation Command, the Navy’s Operational Test and Evaluation Force, the Air Force Test and Evaluation Command, and the Joint Interoperability Test Command, the non-Government trusted agents in our research and development teams (e.g. federally funded research and development centers and university applied research centers, as well as other deep experts in academia to tackle high-priority challenges collaboratively. Regular T&E summits and workshops can ensure new programs do not start from scratch when grappling with AI-powered systems. In short, building networks of T&E expertise will ensure the workforce tests not just hardware, but reprogrammable, autonomous, and cyber-contested systems, as well.

Maintain Rigor and Independence in T&E

Finally, the Pentagon must preserve the principle of independent and comprehensive T&E. In the era of software-defined warfare where systems evolve continuously, independent oversight provides the trust anchor for every update. The office of DOT&E—or an equivalent—must retain authority to provide unvarnished assessments, free from political or programmatic pressures. Independent T&E should be reframed not as a bureaucratic brake but as an enabler of speed: by having the discipline of catching problems early, thus preventing catastrophic failures that derail campaigns.

Pentagon leadership must ensure that any workforce reductions are balanced by investments in tools and training. Likewise, Congress should require programs to allocate sufficient time and resources for OT&E events, even if that delays initial fielding. In practice, this process means applying a new rule of thumb for software-defined warfare: “Never trust an

algorithm in battle until it's been stressed in every way feasible." Continuous capability evolution is a strength only if each iteration earns confidence through testing. Independent OT&E ensures that every reprogrammable, AI-enabled system is not just innovative—but dependable.

By pursuing the above initiatives, the Pentagon can build a T&E infrastructure and culture that matches the flexibility and speed of software-defined warfare. The goal is a virtuous cycle where rapid innovation is met with equally agile verification, enabling the United States to confidently field cutting-edge capabilities that not only work as promised but can be trusted in the chaos of war.

Concluding Remarks

Ironically, the Pentagon's digital revolution and its testing crisis are two sides of the same coin—both born from the dizzying pace of technological change. The U.S. military must innovate rapidly to keep its edge, embracing rapid adaptability through software updates, AI-driven decision support to manage complexity, autonomous systems to expand force projection, and cyber-as-warfare tools to contest new domains. Yet technology alone does not win wars; only technology that works when and as needed does. Ensuring that outcome depends on thorough and independent comprehensive T&E to verify that each tenet of software-defined warfare is reliable in the chaos of combat.

In their rush to field the future, defense leaders cannot afford to abandon the proven principles that undergird past successes: realistic training, iterative improvement, and honest evaluation of what works and what fails. There's an old military adage: "Never send a weapon into war until you've seen it fight." In the context of software-defined warfare, that adage must be updated to: "Never deploy an algorithm, a digital twin, or a reprogrammable system to battle until you've stressed it in every way feasible." Achieving that standard will require reinvention and sustained investment in the T&E enterprise—not as an optional safeguard, but as the foundation for realizing the promise of an AI-enabled, software-driven arsenal.

As the Pentagon races ahead with autonomous UAS swarms, AI-powered battle managers, and Internet-of-Things-style weaponry, it must continually ask itself: Who is testing these new technologies, and to what standard? Who ensures that our autonomous systems behave predictably under duress, that our AI-driven decision support resists bias (or at least, incur biases we understand), and that our reprogrammable weapons adapt without introducing new vulnerabilities? The coming years will be telling. We will see whether the United States can balance speed with safety, and innovation with validation, in the realm of defense.

In fact, this piece is all about propelling the kinds of change in outcome the new defense acquisition policy states as an objective. Ultimately, our men and women in uniform deserve to go into the next conflict with systems we know will have their backs—not with question marks. To get there, the Pentagon must rebuild a resilient testing safety net, updated for the digital battlespace. The future of American warfare will depend not just on how fast we can innovate, but also on how well we can test, validate, and trust what we create.

References

Air Force Life Cycle Management Center. (2025, March 10). *Modern software acquisition to speed delivery, boost warfighter lethality*. https://www.aflcmc.af.mil/NEWS/Article/4115238/modern-software-acquisition-to-speed-delivery-boost-warfighter-lethality/

American Institute of Aeronautics and Astronautics (AIAA), & Aerospace Industries Association (AIA). (2020). *Digital twin: Definition & value—An AIAA and AIA position paper*.

https://www.aiaa.org/docs/default-source/uploadedfiles/advocacy/key-issues/digital-engineering/aiaa-aia-digital-twin-position-paper-(nov-2020).pdf
Albon, C. (2025, July 15). *Jammed and confused: Alaska trial shows pitfalls of fielding US drones.* Defense News. https://www.defensenews.com/pentagon/2025/07/15/jammed-and-confused-alaska-trial-shows-pitfalls-of-fielding-us-drones
Bain, W. (2025, June 11). *Real-time digital twins with AI/ML: A new level of battlefield intelligence.* Military Embedded Systems. https://militaryembedded.com/ai/machine-learning/real-time-digital-twins-with-aiml-a-new-level-of-battlefield-intelligence
Chief Digital & Artificial Intelligence Office. (2025, July 14). *CDAO announces partnerships with frontier AI companies to address national security mission areas*. https://www.ai.mil/Latest/News-Press/PR-View/Article/4242822/cdao-announces-partnerships-with-frontier-ai-companies-to-address-national-secu
CISA. (2025, January 16). *Closing the software understanding gap.* Cybersecurity and Infrastructure Security Agency. https://www.cisa.gov/sites/default/files/2025-01/joint-guidance-closing-the-software-understanding-gap-508c.pdf
Clark, B., Vaddi, D., & Sloman, H. (2023). *From sensor to shooter: Modernizing the kill chain for the era of software-defined warfare.* Hudson Institute. https://www.hudson.org/national-security-defense/sensor-shooter-modernizing-kill-chain-era-software-defined-warfare
Defense Science Board. (2023, July). *Designing and deploying artificial intelligence for DoD missions*. U.S. Department of Defense. https://dsb.cto.mil/reports/2020s/AI_Study_Final.pdf
Director, Operational Test & Evaluation. (2024, January). *FY2023 annual report.* U.S. Department of Defense. https://www.dote.osd.mil/Annual-Report
DoD. (2024, December 9). *DoD instruction 5000.98: Operational test and evaluation and live fire test and evaluation.* https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500098p.PDF
DoD. (2025). *Software modernization implementation plan FY25–26*. https://dodcio.defense.gov/Portals/0/Documents/Library/SW-Mod-I-Plan25-26.pdf
DoW. (2026, January 6).*Transforming the defense innovation ecosystem to accelerate warfighting advantage* [Memorandum for Senior Pentagon Leadership]. https://media.defense.gov/2026/Jan/12/2003855657/-1/-1/0/TRANSFORMING-THE-DEFENSE-INNOVATION-ECOSYSTEM-TO-ACCELERATE-WARFIGHTING-ADVANTAGE.PDF
Eckstein, M. (2024, April 3). *US Navy making Aegis updates, training changes based on Houthi attacks.* Defense News. https://www.defensenews.com/naval/2024/03/21/us-navy-making-aegis-updates-training-changes-based-on-houthi-attacks
Evans, E. D., Van Wie, D., Green, J. (eds.). (2024). *Report of the defense science board task force on test & evaluation.* Defense Science Board. https://dsb.cto.mil/wp-content/uploads/2024/08/DSB_TE-Report_UNCLASS_FINAL_August-2024_Stamped.pdf
Fox, A. (2024, October 29) *A.I. Flow: Pioneering a new approach to artificial intelligence.* U.S. Army. https://www.army.mil/article/280932/a_i_flow_pioneering_a_new_approach_to_artificial_intelligence
Fredenburg, M. (2025, August 4). *Cutting the DoD's testing office won't reverse US military decline.* The National Interest. https://nationalinterest.org/feature/cutting-the-dods-testing-office-wont-reverse-us-military-decline
Freeman, L., Robert, J., & Wojton, H. (2025, June 24). The impact of generative AI on test & evaluation: Challenges and opportunities. *FSE Companion '25: Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering, Association for Computing Machinery (ACM)*. https://doi.org/10.1145/3696630.3728723

Gallium Solutions. (2025, June 19). *What is an OTA contract? A beginner's guide for innovators and startups*. https://www.galliumsolutions.co/post/what-is-an-ota-contract

Goode, W. S., Costello, R., & Shakarian, P. (2025). *Acquiring generative artificial intelligence to improve DoD influence activities* (RAND Report RR-A3157-1). RAND Corporation. https://www.rand.org/pubs/research_reports/RRA3157-1.html

Guertin, N. H. (2022, July). *Digital twin assessment, agile verification processes, and virtualization technology report.* Office of the Director, Operational Test & Evaluation (DOT&E). https://www.dote.osd.mil/Portals/97/pub/reports/(U)%20Digital%20Twin%20Agile%20Verification%20and%20Virtualization%20Technology%20Report.pdf

Hahn, A. S., deCastro, J., Tanaka, M., Lamb, C. (2023, October). *Evaluation of digital twin modeling and simulation* (SAND 2023-09884). Sandia National Laboratories. https://www.sandia.gov/app/uploads/sites/273/2024/11/SAND_Digital_Twins_Final.pdf

Harper, J. (2025, March 7). *Hegseth issues edict on DOD software acquisition.* Defense Scoop. https://defensescoop.com/2025/03/07/hegseth-memo-dod-software-acquisition-pathway-cso-ota

Hegseth, P. (2025, May 27). *Memorandum directing reorganization of the Office of the Director of Operational Test and Evaluation.* U.S. Department of Defense. https://media.defense.gov/2025/May/28/2003725153/-1/-1/1/MEMORANDUM-DIRECTING-REORGANIZATION-OF-THE-OFFICE-OF-THE-DIRECTOR-OF-OPERATIONAL-TEST-AND-EVALUATION.PDF

Magnuson, S. (2025, February 27). Generative AI used to speed up defense acquisitions. *National Defense Magazine*. https://www.nationaldefensemagazine.org/articles/2025/2/27/generative-ai-used-to-speed-up-defense-acquisitions

McNamara, W. M., Modigliani, P., & Nurkin, T. (2025). *Commission on software-defined warfare: Final report.* Atlantic Council, Washington, DC.

Mulchandani, N. & Shanahan, J. N. T. (2022, September). *Software-defined warfare: Architecting the DOD's transition to the digital age.* https://www.csis.org/analysis/software-defined-warfare-architecting-dods-transition-digital-age

Naegele, T. (2024, December 9). JSE: How Air Force aims to get more pilots into world's best F-35 simulator. *Air & Space Forces Magazine*. https://www.airandspaceforces.com/jse-air-force-f-35-simulator-more-pilots/

Podnar, T., Dobson, G., Updyke, D., & Reed, W. (2021, May 19). *Foundation of cyber ranges* (SEI Report CMU/SEI-2021-TR-001). https://doi.org/10.1184/R1/13557566

Schmidt, D. C. (2025). Software testing in the generative artificial intelligence era: A practitioner's playbook. *IEEE Computer, 58*(7), 147–152.

Schmidt, D. C., & Guertin, N. H. (2025, September 3). *The Pentagon's software revolution and its testing dilemma.* War on the Rocks. https://www.warontherocks.com/2025/09/the-pentagons-software-revolution-and-its-testing-dilemma

Software Engineering Institute. (2020). *TwinOps combines digital twins and DevOps for better cyber-physical systems.* https://www.sei.cmu.edu/annual-reviews/2020-year-in-review/twinops-combines-digital-twins-and-devops-for-better-cyber-physical-systems

Software Engineering Institute. (2025, May 12). *SEI study details ongoing DevSecOps adoption in DoD.* Carnegie Mellon University. https://www.sei.cmu.edu/news-events/news/article.cfm?assetId=109121

Sherbinin, A., & Gray, A. (2025, February 10). *Software-defined warships: The Navy's digital future of necessity*. War on the Rocks. https://warontherocks.com/2025/02/software-defined-warships-the-navys-digital-future-of-necessity

South, T. (2022, November 4). *Army's project convergence links sensors, shooters in joint experiments.* Army Times. https://www.armytimes.com/news/your-army/2022/11/04/armys-project-convergence-links-sensors-shooters-in-joint-experiments

Tucker, P. (2025, March 7). *Pentagon aims to accelerate acquisition of new tech through software-contracting change.* Defense One. https://www.defenseone.com/technology/2025/03/pentagon-aims-accelerate-acquisition-new-tech-through-software-contracting-change/403598

Wilkinson, S. (2014, February 26). *10 of history's worst military weapons.* HistoryNet. https://www.historynet.com/10-of-historys-worst-weapons


Acquisition Research Program
Department of Acquisition, Finance, and Manpower
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net