

# TASK FORCE LIMA

## Executive Summary



# Task Force Lima Executive Summary

In August 2023, the Deputy Secretary of Defense (DSD) established Task Force Lima within the Office of the Chief Digital and Artificial Intelligence Officer (CDAO) to develop, evaluate, recommend, and monitor Generative AI (GenAI) capabilities – particularly large language models (LLMs) across the Department of Defense. This tasking and the report presented here is in keeping with the National Security Memorandum on Advancing the United States’ Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence (hereafter 2024 NSM on AI). The 2024 NSM on AI directs the national security community to “utilize advanced AI technologies, with proper safeguards in place, to meet national security goals” and highlights the critical role of LLMs in maintaining the United States’ advantage during a “transformational shift in the AI landscape.” The 2024 NSM on AI also provides guidance to the national security community to “harness powerful AI, with appropriate safeguards, to achieve national security objectives,” and specifically identified LLMs as important to maintaining the U.S. advantage amid a “paradigm shift within the AI field.”

This report conveys the accomplishments and findings of Task Force Lima along the five objectives articulated in the DSD memorandum establishing the task force:

- accelerate promising GenAI initiatives and joint solutions;
- federate disparate developmental and research efforts into a Department of Defense (DoD) community of practice;
- evaluate solutions across doctrine, organization, training, materiel, leadership and education, personnel, facilities, and policies (DOTMLPF-P);
- drive education and build a culture of responsible implementation; and
- ensure coordinated DoD engagement with interagency, international, educational, civil society, and industry partners.

## Key Findings:

Task Force Lima’s key findings include:

**There are widespread opportunities for operational and administrative applications that require experimentation and iteration to realize.** As units are beginning to pilot applications of GenAI, they are discovering technical and policy challenges that need to be addressed as well as new opportunities that can be seized. These pilots are a critical first step to scaling GenAI solutions for the DoD but are already encountering substantial challenges, especially when it comes to scaling solutions.

**GenAI’s capabilities have important limitations that contribute to the risks of applying GenAI in specific use cases, though the capabilities and limitations are changing rapidly.** Current technical issues with GenAI, like hallucinations, the lack of explainability, security vulnerabilities, and limited test and evaluation techniques, condition its deployment. While the technology will never be perfect, it is imperative that leaders only deploy GenAI in situations for which it is well suited and that users understand the

capabilities and limitations before deploying GenAI in safety or security critical applications. At the same time, DoD must stay abreast of improvements and developmental progress to accelerate use responsibly.

**The future of GenAI in the DoD is predicated on having sufficient technical talent and computational resources.** GenAI is available today in the private sector and commercial partners are open to collaborating, but to adopt these tools, DoD needs leaders and users who understand how to employ these tools and enough compute to support scaled operations with GenAI. Many pilots underway today are enabled by visionary leaders with deep technical knowledge and an ability to work with commercial developers, but even the most advanced pilots are already challenged by a lack of access to compute resources. Furthermore, the long timeline for training and high cost of computing resources, particularly for advanced models, pose a significant barrier to widespread GenAI adoption that the DoD must urgently address.

**DoD policies will need to evolve to meet the unique aspects of GenAI.** GenAI's unique development pipeline, continuous test, evaluation, verification and validation (TEVV) approaches, and maintenance requirements are significantly hindered by traditional, hardware-centric procurement processes. Similarly, GenAI brings with it unique cybersecurity concerns that evolve at a pace that outstrips legacy network authorization processes and policies. Acquiring and adopting GenAI at the speed and scale needed to drive U.S. advantage will take a level of flexibility that allows for new opportunities to be quickly seized, and emerging issues rapidly addressed.

### **Summary of Recommendations:**

**Continue to rapidly pilot GenAI to discover and evolve its most impactful applications.** Enable those pilots by improving accessibility to compute infrastructure and commercial technical expertise, establishing provisional authorizations for LLM services across the main cloud service providers, embedding test and evaluation teams in pilot efforts, and ensuring frontier models can be licensed in the DoD domain. Equip users for these pilots by investing in education and plain-language guidance on capabilities and limitations of GenAI systems.

**Work with industry and academia to improve and appropriately harness GenAI.** GenAI techniques are still rapidly being developed in the commercial sector and with the help of academia. This development is already beginning to address some of the known shortfalls of GenAI, which impede its broad application, such as hallucinations. DoD's strategy should be to leverage commercial solutions and talent in this rapidly changing field and focus any research and development efforts on instances where commercial solutions are insufficient for military tasks.

**Clear the pathway for the best pilots to scale** by streamlining GenAI policies across the DoD, elevating the baseline of AI literacy across the force, and acquiring the necessary compute (cloud or on-premises). Establishing the compute resources necessary for scaling will be especially important, and DoD will need to signal its projected requirements for classified cloud computing as well as the scale of need for on-premises computing.

### **Immediate Next Steps:**

The way forward: CDAO recommends the sunset of Task Force Lima as an independent unit and federation of the following GenAI adoption efforts across the offices of primary responsibility (OPR) listed in parenthesis:

- **Support the development of experimental environments and exercises** for proof of concepts and pilot projects, enabling a deeper understanding of GenAI capabilities and limitations before enterprise-wide deployment. (OPR: CDAO; OCR: DIU, CCMDs)
- **Monitor and adopt new GenAI and GenAI TEVV** techniques to address technology shortfalls in tandem with organizational or training efforts to reduce the probability and impact of technical shortfalls in deployments. Establish proactive outreach and monitoring of GenAI advancements in industry and academia to inform DoD activities and ensure the Department remains at the forefront of technology. (OPR: CDAO; OCR: DOT&E)
- **Sustain DoD involvement in interagency efforts to develop security standards for GenAI** and work with academia and commercial industry to develop affordable solutions for start-ups and smaller vendors. (OPR: CDAO; OCR: CIO, A&S, Policy, I&S)
- **Sustain research and development efforts to address DoD-unique classification and security concerns with GenAI**, such as classification by aggregation. Additionally, to avoid inadvertent spills through employee queries to commercial and unsecured GenAI services, accelerate the scaling of DoD-secured GenAI experimental platforms like NIPRGPT and CamoGPT to give interested GenAI users an alternative to using commercial and less secure models. (OPR: CDAO; OCR: CIO, I&S)
- **Develop and resource a comprehensive acquisition strategy for GenAI**, encompassing development, acquisition, evaluation, and sustainment of both pilot projects and scaled capabilities across research and development and operational networks. Address hosting sustainment of GenAI tools and compute capabilities to maximize effectiveness for operations as well as research and development. (OPR: CDAO; OCR: DIU, CIO)
- **Continue to build and harmonize the guidelines, guardrails, frameworks, and policies** for innovating with or deploying GenAI solutions. (OPR: CDAO; OCR: Policy).
- **Explore and accelerate options to make the ATO process more efficient**, including through the application of GenAI. Regularly update cyber risk policies that may be specific to GenAI and, wherever possible, harmonize policies to allow for ATOs acceptable across the DoD. Involve Authorizing Officials (AOs) in the GenAI community of practice or discussions. Maintain the LLM Authority to Operate (ATO) dashboard so that commands have the relevant information about LLM platforms with Interim Authorization to Test (IATT) or ATOs on DoD enclaves. (OPR: CDAO; OCR: CIO)
- **Continue to solicit, develop, exchange, and promulgate information on GenAI to support better understanding** of GenAI capabilities and limitations and unify responsible AI approaches. Develop guidance on conducting low-risk experimentation and recognize and reward leadership for initiating and contributing to responsible pilots. Sustain a community of practice to facilitate knowledge exchange, problem-solving, and accelerated GenAI capability development. (OPR: CDAO, OCR: MILDEPs)