



Scale AI Framework for Successful Government AI Adoption



The race for global Artificial Intelligence (AI) leadership is well underway, and the country that most readily embraces AI will reap the national security and economic benefits that accompany this leadership. For this reason, the U.S. government, led by the Department of Defense (DoD), has long invested in AI-enabled platforms like computer vision and machine learning algorithms. It is critical that the United States incorporate the lessons that the DoD has learned related to the acquisition and use of AI as the rest of the federal government embraces AI.

Scale AI (Scale) was founded in 2016 to accelerate the development of AI by providing the data infrastructure that powers AI. From our earliest days of data labeling for the autonomous vehicle industry to our work today to fine tune the data that powers the world's leading large language models, Scale has been on the forefront of AI innovation. Since 2020, we have worked with the federal government to better enable our government's AI capabilities and to position the United States to lead the world in the adoption of AI.

The following lays out Scale's recommendations to ensure that the federal government embraces AI as efficiently as possible.

AI-ready data is the key to unlocking the power of AI

Both the commercial industry and the federal government have learned the importance of AI-ready data that is labeled and annotated to truly unlock the power of AI. For this reason, the leading commercial entities spend billions per year to ensure they have the right data strategy, management procedures, and retention policies in place. The DoD also learned this lesson in its early work on AI, most notably with Project Maven—which was launched in 2017 to speed up AI adoption in the military—and now the DoD fully recognizes that AI-ready data is critical for the Department's own data strategy.

Today, the Chief Digital and AI Office (CDAO) has been established with the responsibility of creating a Centralized Data Repository to build the DoD's "AI Scaffolding" or data infrastructure to power AI. While this top-down leadership is promising, it is also imperative that efforts take place at every level of the DoD to ensure that harnessing

the power of AI-ready data is prioritized.

Additionally, the Intelligence Community has also released a long-term data strategy that highlights the need for AI-ready data. While these efforts are maturing, they can serve as a blueprint for the rest of the federal government, in conjunction with the DoD's work. Every individual agency has numerous use cases for AI to improve their day-to-day efficiencies. For example, a large language model at the Department of Energy could see complex information, which currently takes teams of researchers hours or days to search through manually, almost immediately accessible to anyone who needs it. This vision only becomes reality if this raw and siloed data becomes AI-ready.

For that reason, it is clear that long-term plans and funding is necessary within each agency to accelerate the development and prioritization of AI-ready data.

Scale recommends that all government agencies develop AI scaffolding and create AI-ready data strategies that will truly unlock the potential of AI across the myriad of government use cases.

Test and Evaluation is critical to protect U.S. investment and to ensure high-quality AI systems

Over the next year, the federal government is likely to spend millions of dollars acquiring AI systems, which are critical to maintain our place as the global leader in AI adoption. Understanding whether AI is safe to deploy is one of the most important questions the federal government must answer as AI is inevitably adopted broadly.

Scale has worked for years across the leading AI developers and this experience has demonstrated that the best way to ensure responsible AI is through a risk-based approach to Test and Evaluation with human oversight. This not only protects taxpayer resources by ensuring

that the government acquires high-quality AI systems, but also is one of the strongest methods to ensure accuracy, limit bias, and uphold the Responsible AI Principles, such as those outlined in the Biden Administration's AI Bill of Rights.

Test and Evaluation has long been a key part of the product development cycle for responsibly bringing consumer-facing technologies to market and military technologies into production. This is essential for AI applications because they are rapidly developing and constantly iterating, and therefore constantly presenting new opportunities and risks to the end user. A risk-based approach to Test and Evaluation will ensure that AI is factual, accurate, and explainable regardless of the underlying model or data being used. If the product—including the data infrastructure and underlying model—does not meet these requirements, we risk sacrificing user trust in the technology. These standards should be set by the specific agency based on the intended use case.

While red teaming is necessary to understand the unanticipated vulnerabilities associated with an AI system, a comprehensive Test and Evaluation framework is still imperative to ensure that those vulnerabilities are addressed. In practice, this means that the safety bar is aligned to the risk of the activity the AI is being asked to support. For example, a large language model used for mission critical intelligence analysis will need to meet a higher bar prior to deployment than a large language model used to compile routine reports.

The DoD has already begun to understand the best approaches to Test and Evaluation for Large Language Models (LLMs) with the idea to require this prior to acquisition. Industry is also working towards the development of commercial Test and Evaluation platforms, and standards, which should be a critical part of the development process, prior to bringing a product to market.

All Federal agencies should learn from the DoD's years of experience working on this topic and update their acquisition guidelines accordingly.

Scale recommends that the federal government mandate a risk-based approach to Test and Evaluation with the deployment of any LLM on a government network.

Innovative and expedited paths to acquisition will allow the United States to maintain global competitiveness

The federal acquisition process was designed for legacy platforms, such as aircraft carriers and weapons systems, and can take years to complete. Many tweaks have been considered and adopted to help accelerate the process. However, the acquisition process has not yet adapted to reflect the rapid development of leading technologies today. In addition to the legacy platforms, the government is now acquiring software and other technologies that are iterating rapidly and may be outdated by the time a single acquisition process takes place.

For this reason, industry has long highlighted the “Valley of Death” concept that has hampered the federal government’s ability to acquire innovative technology. To address this issue, the DoD developed Tradewinds Solutions Marketplace, a rapid acquisition vehicle that seeks to accelerate the procurement and adoption of AI and Machine Learning technology. Scale recently was awarded a contract through Tradewinds that occurred in near record time for a federal government acquisition.

Beyond the DoD, the rest of the federal government should be encouraged to expedite the acquisition process by leveraging all available contracting authorities that could make a significant difference for any agency seeking commercial emerging technologies. A slow acquisition process can hamper the government’s ability to utilize innovative technologies, but with AI, it will directly impact our nation’s global leadership because the technology is proving vital to national and economic security. Our adversaries, like China, are not letting bureaucratic processes stand in their way of working towards global leadership.

When acquiring emerging technologies like AI, Scale strongly recommends that all federal agencies embrace, to the maximum extent possible, the full range of acquisition authorities to use frameworks such as Tradewinds or similar tools which allow for the timely acquisition of the best-in-class commercial technologies. Doing so will directly help the U.S. lead the world in the adoption of AI.