# Accelerating Air Force outcomes with AI

> "AI will supercharge Intelligence, Surveillance, and Reconnaissance (ISR) by providing networked sensors capable of identifying hidden 'needles in a haystack' without prior threat knowledge. AI will fuel robotics advancements and enable intelligent swarms of autonomous agents to perform tasks once thought to exist solely under human control."
>
> **Air Force Doctrine Note 25-1: Artificial Intelligence[4]**

## Outpacing peer adversaries requires swift AI adoption

The White House reaffirmed the importance of artificial intelligence (AI) to U.S. global AI dominance in the 2025 Executive Order 14179, "Removing Barriers to American Leadership in Artificial Intelligence."[1] Department of Defense (DoD) components that do not accelerate AI adoption risk being outpaced by a peer adversary already investing heavily in AI.

The U.S. Air Force has used predictive AI for more than a decade—for example, for predictive maintenance of aircraft and carrier elevators. Today the contest for information dominance has shifted to generative AI (gen AI), which produces, translates, or transforms original content by learning from large quantities of data. The Air Force has proceeded cautiously with gen AI, for 2 main reasons. One is the time and work required to scope the use case, data requirements, constraints, and performance metrics. The other is the massive cost and effort involved in building large language models (LLMs) from scratch. Building a specialized LLM can take years, and the resulting model typically lacks the capabilities of commercial models. Accelerated AI adoption is urgent for maintaining information dominance.

## Commercial tools cost less, help speed up implementation, and align with White House mandates

Use of commercial platforms and tools to build models and AI-powered applications helps the Air Force comply with multiple 2025 Executive Orders (EOs) intended to save money and shorten lengthy technology acquisition cycles. Guidance includes:

▸ "Modernizing Defense Acquisition and Spurring Innovation in the Defense Industrial Base," signed in April 2025, states a preference for commercial solutions.[2] The U.S. Space Force had previously published a roadmap for commercial partnerships.[3]

▸ Another EO signed in April 2025 directs the DoD to "prioritize the procurement of commercially available products and services...rather than non-commercial, custom products or services," adding that contracting officers must justify the necessity of non-commercial products.[4]

▸ The Defense Secretary has directed the DoD to remove barriers to the procurement of digital tools for warfighters by using the software acquisition pathway (SWP). The goal is, "...cutting red tape to deliver software faster and tap into commercial innovation—ensuring warfighters stay ahead."[5]

**1** The White House. "White House Releases New Policies on Federal Agency AI Use and Procurement." 25 April 2025.

**2** The White House. "Modernizing Defense Acquisitions and Spurring Innovation in the Defense Industrial Base." 9 April 2025.

**3** U.S. Space Force. "Space Force official outlines roadmap for commercial partnerships." 24 April 2024.

**4** The White House. "Fact Sheet: President Donald J. Trump Enforces Requirement of Cost-Effective Commercial Solutions in Federal Contracts." 15 April 2025.

**5** *Department of Defense. "Directing Modern Software Acquisition to Maximize Lethality." 10 March 2025.*

To specialize a commercial LLM for the mission, Air Force data scientists and developers need platforms and tools to:

- ▶ Control the data set used for training.

- ▶ Scope model parameters. An LLM might have billions of parameters, of which only a small fraction are needed for a particular Air Force use case.

- ▶ Apply weights that determine the influence of each model input.

- ▶ Integrate with new or existing rules engines and fine-tune rules based on domain data.

- ▶ Deploy the model in the optimum location for the mission: a public cloud, datacenter, or the tactical edge.

- ▶ Continually monitor and fine-tune the model.

## Accelerate AI adoption using Red Hat AI technologies

Red Hat provides platforms and tools the Air Force can use to build, test, and rapidly scale AI models that are precise and accurate, more quickly than if the model were built from scratch.

Start with a proof of concept by adapting a commercial LLM for the use case. Red Hat® Enterprise Linux® AI is a gen AI foundation model platform that includes the IBM Granite family of LLMs and InstructLab tooling to train and fine-tune models. During the proof of concept, compare the new model's performance to benchmarks for accuracy, response time, and hardware resource usage.

After vetting the model's mission value, deploy it at scale on Red Hat OpenShift® AI. Air Force data scientists and developers can use OpenShift AI to develop and serve predictive and gen AI models in any location across the Air Force's hybrid cloud environment—public cloud, datacenter, or small, power-constrained servers at the tactical edge. Use OpenShift AI for data acquisition and preparation, and for model training, fine-tuning, serving, and monitoring.

## High-value AI proofs of concept for the Air Force

### Space domain awareness: predicting solar events

Current process. Solar flares can disrupt radio communications, global positioning system (GPS), satellites, and operational planning. Currently, historical and newly captured data are fed into a predictive model running on a high-performance computing (HPC) system. The model typically generates a 7-10 day prediction window in 5 or more hours.

Red Hat AI proof of concept. Speed up decision support for responding to and mitigating the consequences of solar events by building and serving models in a federated, distributed architecture–on-premise, in a public cloud environment, or at the edge. As a proof of concept, Air Force data scientists and developers can demonstrate the value of AIOps with continuous integration/continuous delivery (CI/CD) for predictive analytics with inferencing. The mission value is to improve data analysis and fidelity, provide more timely warning about solar events, and continually improve accuracy and precision.

---

**6** U.S. Air Force. "Air Force Doctrine Note 25-1: Artificial Intelligence." 8 April 2025.

### Training maintenance technicians

**Current process.** Today, maintenance technicians refer to lengthy manuals and undergo substantial training to provide precision maintenance on complex aircraft and ancillary equipment. Master aircraft technicians work to make sure that every component of high-performance aircraft is maintained to exacting standards. Sharing knowledge acquired through years of hands-on experience and training is difficult.

**Red Hat AI proof of concept.** Use Red Hat Enterprise Linux AI to build predictive and gen AI applications for maintenance technicians stationed on bases, depots, or downrange. Predictive AI can help identify when flightdeck elevators will need new parts, for example. Gen AI can create training materials for maintenance technicians, answer questions real-time, open tickets for maintenance, obtain permission from the appropriate supervisor, and schedule maintenance windows.

### Improving the performance of autonomous systems

**Current process.** Sensor systems typically transmit raw data to another location for processing. This increases bandwidth requirements and adds latency. A dropped connection can disrupt critical links for command and control (C2), interrupt information sharing, affect operational continuity, and potentially lead to mission failure.

**Red Hat AI proof of concept.** Add intelligence to sensor systems so that the model can continue operating when the connection is dropped. The model executes or recommends the appropriate action based on sensor inputs and the mission payload.

### Enhancing modeling and simulation capabilities and improving outcomes

**Current process.** Air Force training replicates real-world scenarios as closely as possible to give operators necessary skills and decision-making abilities. The goals are operational readiness and confidence that human and nonhuman pilots are safe, reliable, and capable of performing in high-stakes operations.

**Proof of concept.** Enhance modeling and simulation (M&S) capabilities by building sophisticated, data-driven models. Use gen AI to produce a course-of-action analysis with confidence scores to help pilots make decisions more quickly and with greater accuracy. Red Hat AI solutions meet M&S requirements by scaling compute resources to match demand, integrating with existing Air Force tools, and providing robust management capabilities to keep models accurate and efficient over the long term. The goal is better mission outcomes at a lower cost.

### Red Hat AI advantages for the Air Force

**Rapidly train and fine-tune LLMs.** Use Red Hat AI tools to adapt commercial models for domain-specific use in months, not the years needed to build a model from scratch. Capabilities include controlling the data set used for training with retrieval-augmented generation (RAG) and other techniques, scoping model parameters, applying weights to the model, and establishing new rule sets.

**Optimize costs by using hardware resources efficiently.** OpenShift AI increases efficiency by reducing compute load on hardware like graphic processing units (GPUs) and central processing units (CPUs), which lowers hardware costs and energy consumption. As one example, an optimized cache reduces the steps needed to generate an accurate answer. Another way OpenShift AI helps to

---

**Optimized models reduce hardware and power requirements**

"Many advanced AI applications require massive computing power, hence lots of data storage and energy. Even edge computing on the frontlines requires adequate power supplies and storage capability."

Air Force Doctrine Note 25-1: Artificial Intelligence[6]

lower costs is by using the virtual LLM (vLLM) framework to distribute inferencing workloads across multiple resources, using those resources more efficiently. Hardware-optimized inferencing is especially useful for AI applications running in tactical edge locations with size, weight, and power (SWaP) constraints.

**Host the model on a variety of servers.** Serve AI models on any of the following platforms: Red Hat Inference Server, single-model, multimodel, or NVIDIA NIM. The advantage of AI Inference Server is optimizing model inferencing for hybrid cloud environments, which accelerates model deployment and can reduce hardware costs. Inference Server combines 3 components that work together for faster, more accurate, and cost-effective inferencing:

▸ The server runtime, vLLM, maximizes throughput and minimizes latency.

▸ An optimized model repository containing validated models accelerates deployment with performance that meets benchmarks.

▸ An LLM compressor uses advanced quantization techniques to improve inference speeds while maintaining prediction accuracy.

Together, the components of Red Hat AI Inference Server increase the speed and accuracy of inferencing while lowering hardware costs.

**Accelerate inferencing with agentic workflow.** Agentic AI systems reduce the need for human intervention in model training and task execution. AI agents operate by breaking down complex goals into a series of smaller, actionable steps and performing these autonomously. For example, an AI agent controlling a robotic system might use cameras, sensors, and monitors to collect environmental data. The agent then processes the data, possibly feeding it into a model, to determine the next appropriate action or step. This self-directed approach allows the system to adapt and respond to dynamic situations that Airmen might confront. Red Hat AI tools take a distinct approach to agentic AI. Rather than exclusively relying on LLMs for comprehensive responses, our tools use individual AI agents to address specific components or aspects of a query. Agent outputs are then integrated to form a complete response, in less time and with more nuance than the response from traditional agentic AI systems.

**Comply with Executive Orders mandating use of commercial software.** Red Hat AI satisfies requirements for DoD service components to use commercially available products and services rather than noncommercial, custom products or services.

## Learn more

Interested in seeing what our solutions can do? Contact us to discuss an AI proof of concept to support the mission.

Read more about Red Hat AI.

**About Red Hat**

Red Hat helps customers standardize across environments, develop cloud-native applications, and integrate, automate, secure, and manage complex environments with award-winning support, training, and consulting services.