

Much of today's AI infrastructure was built for research labs and hyperscaler data centers, not air-gapped forward sites, shipboard operations or the bandwidth- and power-constrained environments where Special Operations Forces (SOF) work. Many AI platforms were optimized for data scientists, not operators, leaving lifecycle management, resiliency, security patching and observability as afterthoughts. At the same time, hardware accelerators change on 3–6-month cycles and model providers continuously release new architectures, making the cost of constantly re-architecting unsustainable for defense organizations.

Private AI at the Edge

VMware by Broadcom has spent four years solving how to deliver sovereign, edge-first AI infrastructure that works wherever the mission goes, preserves data chain of custody and achieves a total cost of ownership hyperscalers cannot match.

VMware Cloud Foundation 9 (VCF 9) delivers integrated compute, networking, storage and AI runtime support as a production-grade IaaS platform that extends from enterprise data centers to single-node tactical edge sites. By abstracting the underlying infrastructure, VCF 9 helps defense teams adopt new accelerators and models without rebuilding the stack.

VCF supports frontier and specialized open source models at the edge, scaling from single-node deployments to full clusters without infrastructure redesign. It also supports fully disconnected operations across the data plane, control plane, deployment, monitoring and key management, giving defense teams complete control in air-gapped environments.

Governance, Security and the Agentic Threat Surface

As AI agents expand across defense software development and operations, the attack surface grows with them. Agents can be granted access to sensitive data sources, creating risk if tool usage, identity controls and prompt flows are not centrally governed.

VCF addresses this through an AI gateway that governs agent access, enforces identity-aware controls and limits exposure from unauthorized tools, prompt injection and unmanaged token usage. For agentic workloads, VCF can also apply access controls and content security at the agent level rather than just the perimeter.

Shared GPU infrastructure allows multiple AI PaaS layers to coexist securely, improving density and reducing cost while preserving operational control. Confidential computing capabilities further protect model weights in transit and at rest, supporting the assurances required to bring frontier models to edge deployments.

Observe, Orient, Decide, Act

For SOF applications, VMware's dark vessel alerting demonstration shows how AI can fuse sensor data, prioritize signals and route insights to human decision-makers while keeping operators in control.

Pull-based software updates help edge nodes stay current without reliable connectivity. Zero-touch provisioning allows nodes to be deployed with minimal technical skill on-site. Lifecycle management across distributed sites is treated as a core design requirement, not a later integration challenge.

For more, visit: vmware.com/solutions/industry/federal-government-it-solutions