

Operationalizing AI at Scale with AI Infrastructure-as-a-Service for Federal Agencies

As the use of AI permeates throughout federal agencies, the struggle to support the diversity of AI use cases and their workload requirements is forcing infrastructure transformation. To overcome operational burdens, performance challenges, and rising costs from force-fitting the use of legacy hardware to satisfy diverse AI workload requirements, agencies are looking for a next-generation platform that accelerates their AI journey.

The Rapid Adoption of AI

Adoption of AI across federal agencies continues to gain steam, as agencies want smarter and faster ways to gain value from growing data sets.



73% of organizations currently **have AI projects in flight that utilize specialized infrastructure** to handle their AI workloads.

Federal Use Cases:

- Developing autonomous ground vehicles to improve the safety of soldiers and other personnel.
- Maximizing government fleet and asset uptime by using predictive maintenance from sensor data and NLP to analyze technician records.
- Increasing drug safety through the application of deep learning on package labeling.
- Detecting anomalies in radiology and pathology images with clinical-grade accuracy.
- Addressing environmental disasters by appropriately directing first responders in scenarios like wildfires.
- Delivering real-time conversational speed transcription and translation services for international government affairs.

Challenges Preventing AI Success

Forging a new frontier with AI requires organizations to quickly assess their infrastructure, application stack, data management capabilities, and desired outcomes with a goal of adopting new approaches that yield better, faster outcomes.

Biggest challenges or barriers with AI projects

- 35%** Cost of IT infrastructure required
- 29%** Need better IT infrastructure capabilities
- 21%** Insufficient data quality
- 21%** Need better app dev environments
- 21%** Inexperienced staff delaying results

Infrastructure Considerations to Deliver AI at Scale

Agencies want the simplicity and agility of the public cloud for rapid AI development and the performance and power of an on-premises AI platform delivered as-a-Service.

Top infrastructure considerations to support AI initiatives

- Hybrid/multi-cloud capability
- Maximizing hardware/infrastructure utilization
- Integrated development environment (IDE)
- Model management and monitoring
- Integration with GPU
- Speed of deployment/provisioning
- Management simplicity
- Lowest possible latency
- Data movement
- Highest possible bandwidth

The Bigger Truth

Agencies strive to move from AI development to production quickly and at scale. AI Infrastructure-as-a-Service is proving to be the answer. VION and NVIDIA are enabling federal agencies to iterate on their AI use cases rapidly with an on-premises, cloud-like infrastructure that delivers right-sized resources as-a-Service. Optimized for all AI workloads, agencies gain peace of mind knowing testing, scaling, and production capabilities are handled through a flexible and agile infrastructure across the complete AI lifecycle, from development and experimentation to training and inference.

[LEARN MORE](#)



The NVIDIA DGX™ A100 system features eight NVIDIA GPUs and two second-generation AMD EPYC™ processors.