

Accelerate AI in the global public sector

Discover how Red Hat helps government agencies transition from pilot programs to production-scale AI with a platform built to:

- Modernize citizen services by using predictive, generative, and agentic AI capabilities.
- Enforce data sovereignty across hybrid, edge, and air-gapped environments.
- Significantly cut AI operating costs by maximizing hardware efficiency.

Resolving the paradox of public sector AI

Public sector organizations worldwide face a critical mandate: do more with less while meeting rising citizen expectations for digital-native experiences. Governments are accelerating the adoption of artificial intelligence (AI) to increase productivity, reduce manual errors, and support decision-making. However, scaling these technologies presents significant structural and operational hurdles.

To effectively modernize, agencies must navigate the transition from pilot programs to production-scale AI. This requires a clear understanding of the evolving AI landscape and a strategy that prioritizes data sovereignty, cost control, and transparency.

Defining the AI landscape for government

As agencies modernize, understanding the specific capabilities of different AI technologies is essential for aligning tools with mission outcomes. For example, predictive AI technology analyzes historical data to identify patterns and forecast future events. For government agencies, predictive AI is vital for mitigating risk. Common use cases include detecting tax fraud, forecasting disease outbreaks, predicting maintenance needs for critical infrastructure, and assessing cybersecurity risks. Generative AI (gen AI) goes beyond analysis to produce, translate, or transform original content by learning from vast quantities of data. It is transforming public sector productivity through knowledge retrieval, semantic search, and the automation of routine tasks like summarizing documents, drafting correspondence, and refactoring existing code. Lastly, agentic AI represents the next evolution in automation. Agentic AI consists of autonomous systems capable of reasoning, making decisions, and executing multistep tasks within predefined parameters. Unlike a chatbot that waits for a prompt, an AI agent can initiate actions to achieve a goal, such as resolving customer service issues across multiple platforms or automating IT remediation. This allows agencies to move from simple task automation to autonomous operations that can adapt to changing conditions.

Navigating structural hurdles to innovation

Despite the promise of AI, public sector leaders face distinct barriers to adoption.

- ▶ Aging infrastructure and disconnected data: A major portion of IT budgets is often allocated to maintaining legacy systems, leaving little room for innovation. Furthermore, disconnected data held in isolation prevents agencies from training AI on comprehensive, real-time information, hindering automation at scale. Without a unified view of data, agencies struggle to deploy the automated workflows necessary for modern service delivery.
- ▶ Cost and scalability: The computational demands of gen AI can push cloud budgets far beyond planned expenditure. As agencies automate processes, such as citizen support chatbots, the costs associated with inference—the process of generating a response—can escalate rapidly. Agencies face a paradox where the tools meant to reduce manual effort create staggering infrastructure bills, forcing trade-offs with other essential programs.

- ▶ Data sovereignty and compliance: Regulatory frameworks, such as the EU Artificial Intelligence Act, classify many public sector use cases as high risk, requiring strict technical documentation, systematic bias testing, and tamper-proof audit logs. Agencies must navigate complex data sovereignty barriers to ensure sensitive information remains within specific jurisdictions or organizational boundaries. This is particularly critical for healthcare, law enforcement, and judicial data, where privacy cannot be compromised.
- ▶ The skills gap: As AI becomes embedded in public services, the demand for skilled talent often outpaces supply. Public sector organizations frequently struggle to compete with the private sector for highly qualified AI and data science talent due to salary constraints. Agencies need tools that lower the barrier to entry, allowing existing staff to contribute to AI initiatives without requiring deep specialization in data science.

Red Hat's approach: Open, scalable, and security-focused AI

Red Hat provides a flexible, open platform designed to help government organizations accelerate AI innovation while controlling costs and mitigating risk. By using open source technologies, agencies can avoid vendor dependence, maintain flexibility across open hybrid cloud environments, and benefit from community-powered innovation.

Technological solutions for public sector challenges

- ▶ Hybrid cloud flexibility and sovereignty: Red Hat® OpenShift® AI helps agencies build, train, and deploy models across various infrastructures—on site, in the cloud, or at the edge of the network. This flexibility allows data to remain in security-focused environments to meet strict data privacy and sovereignty requirements, such as those found in air-gapped or disconnected environments.
- ▶ Cost-effective inference: To address escalating costs, the Red Hat AI Inference Server optimizes model performance. It uses technologies like virtual large language model (vLLM) for advanced memory sharing and dynamic batching, alongside model compression techniques, to maximize hardware efficiency. This approach can significantly reduce the cost per token compared to traditional serving methods.
- ▶ Privacy-preserving collaboration: Red Hat supports federated learning, allowing models to be trained across different departments or regions without moving sensitive data. For example, agencies can use privacy-preserving techniques to predict risk locally without exposing sensitive records. Additionally, support for confidential computing (secure enclaves) ensures data remains protected even during processing in shared cloud environments.
- ▶ Automated governance and security: Red Hat Trusted Application Pipeline embeds security controls directly into the development lifecycle. It automates the generation of Software Bills of Materials (SBOMs), signature attestation, and vulnerability scanning, ensuring that each model component is verified and compliant before deployment.
- ▶ Bridging the skills gap: Tools like Red Hat Ansible® Lightspeed use gen AI to translate plain-language commands into production-ready automation code. This helps multiple teams with varying skill levels work consistently and efficiently, reducing the burden of manual scripting and allowing staff to focus on higher-value tasks.

The Red Hat AI portfolio

- ▶ [Red Hat Enterprise Linux® AI](#): A foundation model platform that enables agencies to develop, test, and deploy gen AI models using the open source Granite family of large language models (LLM) and InstructLab tooling for domain-specific alignment.
- ▶ [Red Hat OpenShift AI](#): A flexible machine learning operations (MLOps) platform that allows data scientists and developers to build, train, and deploy predictive and generative models at scale across open hybrid cloud environments.
- ▶ [Red Hat AI Inference Server](#): A scalable engine that helps agencies run various gen AI models on various hardware accelerators to achieve consistent, cost-effective inference in less time.
- ▶ [Red Hat Ansible Automation Platform](#): An intelligent automation solution that integrates gen AI via Ansible Lightspeed to streamline infrastructure management and accelerate IT modernization.

Empower your mission with Red Hat AI

Transition from pilot programs to production-scale AI with a platform built for the specific needs of the global public sector. Red Hat offers the tools to automate strategically, spend wisely, and serve citizens effectively.

To learn more about how Red Hat can help you build your AI-empowered future, visit us at [Red Hat AI](#) or [contact](#) our experts.



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.

f facebook.com/redhatinc
x [@RedHat](https://twitter.com/@RedHat)
in linkedin.com/company/red-hat

North America

1888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa

00800 7334 2835
europe@redhat.com

Asia Pacific

+65 6490 4200
apac@redhat.com

Latin America

+54 11 4329 7300
info-latam@redhat.com