aws

How technology leaders can get started with AI/ML in AWS GovCloud (US)

Technology leaders in the public sector want to use the benefits of artificial intelligence (AI), machine learning (ML), and generative AI to enhance mission impact and achieve business goals. At the same time, they are also obligated to safeguard sensitive government data through alignment with strict regulatory compliance programs and standards. They must also align emerging government policies such as the White House Executive Orders on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Improving the Nation's Cybersecurity, and OMB Policy to Advance Governance, Innovation, and Risk Management in Federal Agencies' Use of Artificial Intelligence.

How can these technology leaders rapidly deliver responsible AI/ML-based innovation while aligning with the broad spectrum of strict regulatory requirements? The answer is simple: Amazon Web Services (AWS) GovCloud (US) provides the technology that underpins a solid foundation for securely and compliantly building and deploying AI capabilities.

AWS is a long-term AI/ML innovator in highly regulated cloud

More than a decade ago, AWS realized that technology leaders with US data residency, ITAR, FedRAMP High, or DoD SRG IL-4/5 compliance requirements needed secure and compliant access to AI/ML capabilities with supporting infrastructure to promote innovation at scale responsibly. We learned that our AWS GovCloud (US) customers must deploy AI/ML solutions that include classifying images, conducting enterprise searches, processing and analyzing documents, transcribing audio, translating languages, and enabling computer-generated speech and conversation chat. AWS answered this need many years ago by deploying a suite of AI/ML capabilities, including Amazon SageMaker, Amazon Rekognition, Amazon Kendra, Amazon Textract, Amazon Transcribe, Amazon Translate, Amazon Polly, and Amazon Lex. All these AI/ML capabilities are available and authorized at FedRAMP High and DoD SRG IL-4/5 for AWS GovCloud (US) customers. Their long-term positive impact on the mission is prolific.

AWS has believed for years that technology leaders need the power of choice in selecting AI/ML solutions while benefiting from the security and compliance posture of AWS GovCloud (US). This is why AWS is focused on democratizing AI/ML technologies and promoting industry partnerships so technology leaders using AWS GovCloud (US) can achieve a broad spectrum of mission and business benefits. Benefits are available to the national industrial base and government agencies and include improved patient care, transformed citizen experience, enhanced warfighter mission impact, more impactful cyber analytics, more accurate forecasting models, more precise digital engineering, and accelerated scientific research at colleges.

AI is ushering in a new era

The rapid pace of generative AI advancement in the early 2020s, coupled with 2023's Executive Order 14110, elevated the visibility of the benefits of AI/ML solutions. Many technology leaders within highly regulated environments began 2024 by developing a strategy to responsibly take advantage of the vast opportunities to innovate with generative AI technologies. This is because they realized this new technology era provides excellent possibilities to elevate business and mission outcomes in ways that previously seemed impossible or uneconomical to achieve.

Shortly after Executive Order 14110, AWS introduced managed generative AI capabilities to AWS GovCloud (US) with the availability of Amazon Bedrock. Amazon Bedrock in AWS GovCloud (US) is now available with Guardrails, and Llama and Claude 3 Sonnet/Haiku models to complement Amazon Titan. Amazon SageMaker is now available with Studio v2 and, which adds support for Mistral, Llama, Falcon, GPT, Mixtral, plus additional features, including Inference, Feature Store, Model Registry, Evaluations, and Management.

As part of Amazon's commitment to responsible AI, we have initiated a 4 billion dollar strategic partnership with Anthropic, one of the world's leading foundation model providers and a

leading advocate for the responsible deployment of generative AI.

The value of these AI solutions to technology leaders operating highly regulated environments is extreme: AWS GovCloud (US) customers can now responsibly advance their generative AI experimentation and development journey in a compliant, secure, resilient, and isolated hyperscale cloud environment. Watch the "What's new with AWS GovCloud (US)" page for more updates.

What's possible with AI?

AWS AI/ML capabilities, infrastructure, and compliance programs position technology leaders to responsibly deploy AI/ML to accelerate their mission and business outcomes using AWS GovCloud (US). Here are some examples:

Transform and enhance citizen experience AI/ML-enabled Amazon Lex chatbots in webpages and contact centers powered by Amazon Connect can increase customer satisfaction and call velocity. Conversation analytics and language translation during customer interactions may be used to reduce time to resolution, improve customer experience, and create tailored customer experiences.

Elevate patient care quality and outcomes AI/ML-enabled Amazon Textract and Amazon Comprehend underpin an AI/ML-based intelligent document processing (IDP) solution that will not only extract text and structured data from documents but will also employ AI/ML capabilities to provide business insights and determine relationships between patient datasets. The velocity of IDP helps healthcare providers efficiently manage patient care and automatically reduce the potential for information errors so providers can focus on positive patient health outcomes rather than on managing patient records and bills.

Improve employee experience and impact
AI/ML-enabled Amazon Kendra provides employers with solutions to help employees spend less searching intranets and data repositories and opening countless documents to find answers to their questions. This results in increased productivity and decreased stress. AI/ML-enabled Amazon Transcribe and Amazon Comprehend enable employees to summarize meetings and dialogues through automated audio note collection, which saves employees time and improves information accuracy. This empowers leaders to make more informed business decisions and increases decision-making velocity.

SaaS and ISV offerings
The regulated software as a service (SaaS) and independent software vendor (ISV) industrial base can also take advantage of the innovative possibilities AI/ML technology offers to enable sustainable growth, expand market presence, strengthen brand quality, modernize legacy systems, increase security, drive efficiencies, and reduce environmental impact.

These are just a few examples of how technology leaders can take advantage of the AI/ML capabilities available in AWS GovCloud (US) today, but how should technology leaders start their AI/ML journey in AWS GovCloud (US)?

What should I do first?

Here are three pieces of advice to help technology leaders dive into this fast-paced AI/ML era safely and responsibly:

1. Start experimenting today, but don't try to boil the ocean
Start with a small yet well-defined AI/ML project with short-term measurable outcomes, broad executive support, and minimal risk. Demonstrate the success of this first project, and then kick off a slightly larger project with more impactful measures. Keep repeating this cycle in time-

bound iterations and continue to demonstrate increasing incremental mission impact. This is how Amazon started over two decades ago when we launched our Amazon.com e-commerce recommendations engine underpinned with AI/ML technology. Technology leaders will benefit more from this approach than endless planning and waiting for the "perfect time" to start their journey.

2. Implement a data strategy that supports AI/ML innovation
Creating a robust enterprise data strategy is critical to implementing successful AI/ML solutions. Generative AI and foundation models (FMs) have elevated the need for clean training data. Quality and clean data are necessary to fine-tune generative AI models to unlock and maximize business and mission value. Technology leaders sometimes face situations where their data strategy does not provide quality training data, which causes the results from AI/ML solutions to produce inferior outcomes due to the "garbage in, garbage out" phenomena. In fact, "Training models with poor data quality will lead to poor results," says Werner Vogels, VP & CTO at Amazon.com. "You will need to filter out bias, hate speech, and toxicity. You'll need to ensure that the data is free of personally identifiable information (PII) or sensitive data…and make sure your data is deduplicated, balanced, and doesn't lead to oversampling."

A well-defined data strategy will pay dividends to organizations embarking on their AI/ML journey. Whether you are building your model or customizing one, all leaders need a data strategy that ensures relevant, high-quality data is available. Some data may be more than 20 years old residing on a mainframe, while others might be massive unstructured datasets living in a legacy storage system. Data must be up-to-date, complete, accurate, discoverable, and available when needed. AWS can help accomplish these goals with Amazon Redshift for data warehouses, Amazon Simple Storage Service (Amazon S3) for data lakes, and Amazon EMR for big data. In turn, these services can increase the value of data as a component of an enterprise AI/ML solution.

> ❝ **Training models with poor data quality will lead to poor results."**
>
> **Werner Vogels | VP & CTO at Amazon.com**

3. Partner with AWS to promote positive, mission-driven outcomes

AWS offers our customers access to AI/ML-focused AWS GovCloud (US) specialist solutions architects and AI/ML-focused professional services. It also has a broad partner community of consultants and integrators. Visit the [AWS AI/ML solutions library](#) to learn more about AI/ML use cases and reference architectures.

AWS GovCloud (US) cloud technology underpins the AI/ML journey

Providing a performant, secure, and compliant technology platform to host AI/ML workloads in an on-premises environment can be prohibitively costly for a typical technology leader. With AWS, you pay as you go. The benefits of AWS GovCloud (US) as a platform for AI/ML solutions are extensive:

AWS GovCloud (US) supports critical sovereignty, security, and compliance programs that support AI/ML workloads. AWS GovCloud (US) Regions are inside an ITAR, FedRAMP High, and DoD SRG IL-4/5 compliance boundary. AWS GovCloud (US) cloud services are authorized at FedRAMP High and DoD SRG IL4 and IL5 or are on a path to authorization. This compliance posture provides customers with a trusted, isolated, and proven path to receive the authority to operate. The simple, pay-as-you-go, cloud consumption model enables customers to consume AI/ML services, database services, GPUs, CPUs, storage, and networking for finite amounts of time in an on-demand, compliant, and scalable model. No longer are there requirements to build and deploy AI/ML on-premises platforms that will sit idle for intermittent periods awaiting consumption or for users to wait for resources to become available due to resource scheduling constraints.

AWS GovCloud (US) comprises two isolated and US Sovereign Regions that promote resilient architectures. While they are available from the public internet, they are isola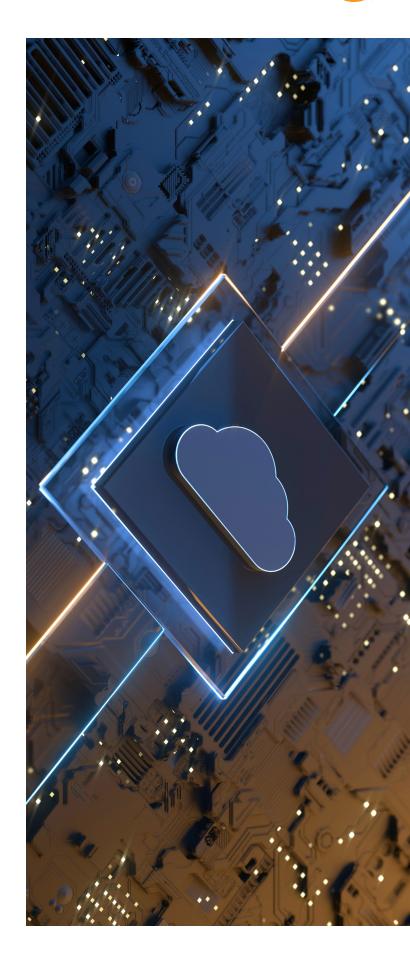ted from the balance of the AWS global footprint and operated independently by US citizens. Each Region contains three Availability Zones composed of data centers with redundant power, networking, and connectivity. Availability Zones are located far enough apart to reduce the risk of a single event impacting availability yet near enough to enable synchronous replication, rapid failover, and low latency. This Region design helps protect applications against disruptions, such as human mistakes, unexpected traffic spikes, utility failures, earthquakes, and weather events. The resulting business and mission value is that technology leaders can architect for the level of availability and resiliency their AI/ML application requires and their budget allows.

The bottom line on AI/ML and regulated cloud solutions

The intersection of GPU-based cloud computing, AI/ML services, networking, and data storage with security and compliance enables highly regulated customers to confidently and responsibly deploy AI/ML solutions on AWS GovCloud (US). AWS encourages technology leaders to think big, experiment, responsibly innovate and take advantage of AWS GovCloud (US) as a strategic technology enabler to achieve their AI/ML-based mission outcomes.

For further reading, we encourage you to download the AWS World-Wide Public Sector AI Solutions Guide to learn more. If you need additional assistance, contact your AWS representative for more information or visit AWS in the Public Sector.

About the authors

David Schatzman

David is a technical business development manager for Amazon Web Services (AWS), which is focused on serving public sector civilian and financial customers using the AWS GovCloud (US) Regions. In this role, David works closely with customers to ensure alignment of their mission goals and technology strategies with the capabilities of the AWS GovCloud (US) Regions. David is also interested in global economics, digital assets, fog computing, and cloud resiliency and is the lead for the AWS GovCloud (US) digital assets and resiliency product strategies. Before working at AWS, David served as a public sector business development manager for Dell Technologies focused on civilian markets, as a principal engineer at Sierra Nevada Corporation (SNC) focused on DoD special projects, and began a career at Sun Microsystems as an account manager for financial and DoD customers. David earned a BS, MA, MBA at George Mason University, and holds Project Management Institute PfMP certification.

Aaron Sengstacken

Aaron is a machine learning specialist solutions architect at Amazon Web Services. Aaron works closely with public sector customers of all sizes to develop and deploy production machine learning and generative AI applications. He is interested in machine learning, technology, and space exploration. Aaron earned a BS in mechanical engineering from the University of Missouri and an MS in aeronautics and astronautics from Purdue University.

Michael Greenwald

Michael is the global head of financial innovation and digital assets at Amazon Web Services (AWS) and leads AWS Global Executive Relations. He works with U.S. and international governments on cloud computing and is responsible for emerging technology innovation and implementation. In 2023, he was appointed to represent Amazon on the U.S. Commodity Futures Trading Commission's (CFTC) Technology Advisory Committee. He previously served as the first U.S. Treasury attaché to Qatar and Kuwait and has served in senior roles with two presidential administrations and under three Treasury Secretaries. He is a graduate of Harvard Business School, has a Juris Doctor graduate from Boston University School of Law, a Master's graduate from Boston University's Frederick S. Pardee School of Global Studies, and a B.A. from George Washington University.

Scott Bourn

Scott is a technical business development manager for Amazon Web Services (AWS) driving ISV and SaaS partner success in the public sector using the AWS GovCloud (US) regions. In this role, Scott works closely with partners to break through market barriers and drive campaign fulfillment aligned with their strategic plan and goals, all with the capabilities of the AWS GovCloud (US) regions. Scott is also interested in privacy, cyber security, network virtualization, cloud FinOps, edge computing, and cloud resiliency. Scott is also the lead for the AWS GovCloud (US) telecommunication and network provider partnerships. Before AWS, Scott served as a Senior Manager for Deloitte Consulting's public sector focused cloud managed services and as a Program Manager for Akamai Technologies' DoD managed CDN. Scott earned an MBA at University of Maryland, and holds PMP, CISSP, ITIL, and DevOps certifications.

Shawn Asfeld

Shawn is a senior solutions architect for AWS GovCloud (US). He has extensive experience working with federal and civilian agencies to build a large variety of secure and compliant workloads on premises and in the cloud. Shawn's current focus is on helping customers and partners to build solutions on AWS GovCloud (US) to meet various levels of compliance, including FedRamp, CMMC and DoD authorization. Shawn earned a BS degree from Texas A&M University and is an AWS-certified solutions architect professional.