

Service Name

Amazon Polly

Service Overview

Amazon Polly is a cloud-based text-to-speech (TTS) service offered by AWS that converts text into lifelike speech using advanced machine learning technology. The service enables developers to create engaging and accessible applications across multiple platforms including mobile apps, eLearning platforms, accessibility applications, and IoT devices. Amazon Polly supports over 60 voices in multiple languages and offers four voice engines: Standard, Neural, Long-form, and Generative, each optimized for different use cases and speech quality requirements.

Key Use Cases

- **Mobile applications:** Adding voice capabilities to mobile apps for enhanced user experiences and accessibility
- **eLearning platforms:** Converting educational content to speech for auditory learners and accessibility compliance
- **Accessibility applications:** Creating inclusive experiences for visually impaired users through document-to-speech conversion
- **IoT devices:** Enabling voice responses and notifications in smart home and embedded systems
- **Contact centers:** Integrating with Amazon Connect for interactive voice response (IVR) systems
- **Content creation:** Generating audio narration for videos, podcasts, and marketing materials

Features

- **Multiple Voice Engines:** Standard, Neural, Long-form, and Generative voices with varying quality levels and pricing
- **SSML Support:** Speech Synthesis Markup Language for controlling speech rate, pitch, emphasis, and pronunciation
- **Speech Marks:** Metadata providing timing information for synchronizing speech with visual elements
- **Custom Lexicons:** Ability to modify pronunciation of specific words and acronyms
- **Brand Voice:** Custom Neural Text-to-Speech voices for organizations
- **Multiple Audio Formats:** Support for MP3, OGG Vorbis, and raw PCM output formats

- **Newscaster Speaking Style:** Professional news-style delivery for US English, British English, and US Spanish voices
- **Time-driven Prosody:** Ability to define maximum speech duration for localization and dubbing

Value Proposition

Amazon Polly provides superior value through its pay-as-you-go pricing model eliminating upfront costs and setup fees, allowing organizations to start small and scale as needed. The service delivers high-quality, natural-sounding speech using advanced neural technology with fast response times suitable for real-time applications:

- Cloud-based deployment reduces local resource requirements while providing instant access to all voices and languages with automatic improvements
- Amazon Polly offers extensive customization options including custom lexicons and brand voices, supports caching and replay at no additional cost, and maintains compliance certifications for HIPAA and PCI DSS regulated workloads

Service Integration

Amazon Polly integrates natively with Amazon Connect for contact center solutions and Amazon Chime SDK for communication applications. The service works seamlessly with AWS Lambda for serverless processing workflows, Amazon S3 for audio storage and retrieval, and AWS Step Functions for orchestrating complex text-to-speech pipelines:

- Additional integrations include Amazon Textract for document-to-speech conversion, Amazon Translate for multilingual workflows, Amazon API Gateway for REST API access, and Amazon CloudWatch for monitoring and analytics
- Amazon Polly also supports integration with third-party contact center platforms like Genesys Cloud CX

Onboarding and Offboarding

Getting started with Amazon Polly requires signing up for an AWS account, which automatically provides access to the service with charges only for actual usage. Customers can begin immediately using the AWS Management Console, AWS CLI, or SDKs in multiple programming languages including Java, Python, Node.js, .NET, PHP, Ruby, Go, and C++:

- The service offers comprehensive getting started documentation, sample code, and tutorials for various platforms including iOS and Android mobile development
- New AWS customers receive free tier benefits with varying character limits for each voice type during the first 12 months

Getting Started Guide: <https://docs.aws.amazon.com/polly/latest/dg/getting-started.html>

Data Removal

Amazon Polly does not retain text inputs processed by the service, ensuring automatic data removal upon processing completion. Users maintain full ownership of their content and generated audio files. For offboarding, customers simply stop using the service as there are no long-term commitments or stored data requiring manual deletion. Generated speech can be cached and replayed indefinitely without additional charges, but storage and management of audio files remain the customer's responsibility.

Backup/Restore and Disaster Recovery

Amazon Polly operates as a fully managed service with built-in resilience across multiple AWS Availability Zones. The service does not store customer text inputs, eliminating the need for traditional backup of processed content. Customers are responsible for backing up their audio files stored in services like Amazon S3. The service maintains high availability through AWS's global infrastructure with automatic failover capabilities built into the platform architecture.

Backup Capabilities

Amazon Polly does not provide built-in backup capabilities as it does not retain customer text inputs or generated audio files. The service is stateless, processing text requests and returning audio streams without persistent storage:

- Customers must implement their own backup strategies for generated audio content using services like Amazon S3 with versioning and cross-region replication
- Amazon Polly does not integrate with AWS Backup service due to its stateless nature

Disaster Recovery

Amazon Polly supports disaster recovery through its multi-region availability and stateless architecture. The service operates across multiple AWS regions, allowing customers to redirect traffic to alternative regions during outages:

- Amazon Polly does not integrate with AWS Elastic Disaster Recovery as it does not maintain persistent infrastructure requiring replication
- The service's cloud-native design provides inherent resilience with automatic geographic distribution of processing capabilities

Recovery Objectives

Amazon Polly helps customers achieve low Recovery Point Objectives (RPO) and Recovery Time Objectives (RTO) through its stateless design and multi-region availability. Since no customer data is retained, RPO is essentially zero. RTO is minimized through the service's ability to process requests in multiple regions simultaneously. Customers can implement

automated failover using Amazon Route 53 health checks to redirect traffic to healthy regions, typically achieving RTOs of minutes rather than hours.

Service Constraints

Service Quotas: <https://docs.aws.amazon.com/polly/latest/dg/limits.html>

FAQ: <https://aws.amazon.com/polly/faqs/>

Technical Requirements

Service Documentation: <https://docs.aws.amazon.com/polly/latest/dg/what-is.html>

User Guide: <https://docs.aws.amazon.com/polly/latest/dg/getting-started.html>

API Reference: https://docs.aws.amazon.com/polly/latest/dg/API_Reference.html

Pricing Overview

Please see the AWS UK G Cloud 15 Pricing Document accompanying this service in the Digital Marketplace.

Public Pricing: <https://aws.amazon.com/polly/pricing/>

Cost Optimization

Amazon Polly offers several cost optimization strategies through its pay-per-character pricing model. Customers can optimize costs by selecting appropriate voice engines based on quality requirements - Standard voices at \$4.00 per million characters for basic needs, versus Neural voices at \$16.00 for higher quality:

- Caching generated audio eliminates repeated synthesis costs
- Using asynchronous processing for long-form content can be more cost-effective than real-time synthesis
- Implementing efficient text preprocessing to remove unnecessary characters and using SSML judiciously helps minimize billable character counts while maintaining speech quality

Savings Considerations

Primary cost savings come from the pay-as-you-go model eliminating upfront licensing fees and infrastructure costs. Caching synthesized speech allows unlimited replay without additional charges, significantly reducing costs for frequently accessed content:

- Choosing the most cost-effective voice engine for specific use cases - Standard voices for basic applications versus premium Neural voices only when necessary - can reduce costs by up to 75%

- Efficient application design with proper error handling and retry logic minimizes unnecessary API calls
- Batch processing longer content using asynchronous APIs can be more economical than multiple real-time requests