



AWS Polly is a cloud service from Amazon Web Services that converts written text into lifelike speech, enabling developers to integrate voice interaction capabilities into their applications. Using deep learning models, Polly produces high-quality speech in a variety of languages and voice styles, offering both standard and advanced neural text-to-speech (NTTS) options.

Key Features of AWS Polly:

1. Multiple Languages and Voices:

- Polly supports over 60 voices across more than 30 languages. This includes male, female, and childlike voices, catering to global user bases.

2. Neural Text-to-Speech (NTTS):

- NTTS is an advanced speech synthesis technique that offers more natural-sounding voices, reducing robotic tones common in traditional text-to-speech systems. This makes Polly's speech output more engaging and humanlike.

3. Real-Time Speech Generation:

- Polly can convert text to speech in real-time, allowing applications to generate speech dynamically, ideal for use cases like customer service chatbots, voice-activated assistants, or news readers.

4. Customization with SSML (Speech Synthesis Markup Language):

- Developers can control aspects like pronunciation, intonation, and pacing using SSML tags. For instance, you can adjust the pitch, volume, and speech speed to create more personalized and suitable voice outputs.

5. Lexicon Support:

- Polly allows users to upload lexicons, which are custom dictionaries that define the correct pronunciation of specific words, names, or jargon, ensuring speech accuracy for specialized vocabulary.

6. Speech Marks for Synchronization:

- Polly provides speech marks (timestamps of words, sentences, and phonemes) that help synchronize speech with visual media. This is useful for applications involving video narration, subtitles, or lip-syncing animations.

7. Audio File Storage:

- Polly enables the saving of speech output as audio files in formats such as MP3, OGG, and PCM. These files can be stored for later use, for example, in podcasts or e-learning modules.

8. Cost-Effective and Scalable:

- Being a fully managed AWS service, Polly automatically scales to handle large volumes of text-to-speech requests, making it cost-effective for high-demand applications.

Use Cases for AWS Polly:

- **Voice Assistants and Chatbots:** Polly powers conversational AI systems in virtual assistants, providing natural voice interactions for customer support.
- **Audiobook and Podcast Creation:** Polly can read articles, stories, or even books, automating the process of content-to-audio conversion.
- **E-Learning Platforms:** Polly narrates educational material, making content more accessible for auditory learners and visually impaired users.
- **Smart Devices:** Internet of Things (IoT) devices use Polly for voice output, enhancing user interaction with devices like smart speakers and home automation systems.
- **Content Localization:** With its multi-language support, Polly helps companies localize content by converting text into speech in various languages.

How AWS Polly Works:

1. **Text Input:** You provide the input text via an API request or a web-based console.
2. **Text Processing:** Polly processes the text and applies speech synthesis models to convert it into lifelike speech.
3. **Voice Output:** The generated speech can be streamed in real time or saved as an audio file for later use.

Advantages of AWS Polly:

- **Engagement:** Converts written content into spoken words, enhancing user experience and accessibility.
- **Customization:** Developers can fine-tune speech characteristics like intonation, emotion, and pace.
- **Flexibility:** Polly's API makes it easy to integrate with a wide range of applications, from mobile apps to multimedia content.
- **Global Reach:** Extensive language support enables businesses to reach diverse, global audiences.

AWS Polly is widely used for creating interactive voice experiences, making digital content more accessible, and providing natural-sounding speech across a wide array of industries.

To begin with the Lab:

1. Login to AWS Console and search for Amazon Polly and from its dashboard choose Try Polly.

Machine Learning

Amazon Polly

Turn text into lifelike speech using deep learning

Amazon Polly is a service that turns text into lifelike speech, enabling you to create applications that talk and build entirely new categories of speech-enabled products. Amazon Polly is a Text-to-Speech (TTS) service that uses advanced deep learning technologies to synthesize speech that sounds like a human voice.

Get started with Amazon Polly

Explore high-quality spoken output with dozens of lifelike voices across a broad set of languages.

[Try Polly](#)

How it works

Amazon Polly's Text-to-Speech (TTS) service turns text into lifelike speech using advanced deep learning technologies. With dozens of lifelike voices across a broad set of languages, you can build speech-enabled applications that work in many different countries. The latest Polly Generative engine is able to synthesize the most expressive and adaptive speech that can be used in any context or use-case. The

Pricing

With Amazon Polly, you only pay for what you use. You are charged based on the number of characters of text that you convert either to speech or to Speech Marks metadata.

[Amazon Polly Pricing](#)

2. This is how the editor would look like. So, this is text to speech IDE.
3. Here you can see that we have to choose the Engine Neural or Standard based on your preference.

Text-to-Speech [Info](#)

[Save to S3](#) [Download](#) [Listen](#)

[Engine](#) [Info](#)
The Long-Form engine isn't supported in every region. For a list of supported regions, see [Feature and Region Compatibility](#).
The Generative engine isn't supported in every region. For a list of supported regions, see [Feature and Region Compatibility](#).

Generative
Produces the most expressive and adaptive speech using Generative AI.

Long-Form
Produces the most natural sounding speech for longer content.

Neural
Produces more natural and human-like speech than Standard Engine.

Standard
Produces natural-sounding speech.

[Language](#) [Info](#)
[English, US](#)

[Voice](#) [Info](#)
[Joanna, Female](#)

[Input text](#) [Info](#)
Hi! My name is Joanna. I will read any text you type here.

[SSML](#) [Info](#)

58 characters used

[Restore default text](#) [Clear text](#)

4. Now what you can do is in the Input text area write a statement of your choice and click on Listen.
5. You can choose either Neural or Standard engine type based on your choice and they will sound different.
6. Also, you have the option to change the language type and the voice character.

- If you click on download then you can download this input text in audio format to listen to it later.

The screenshot shows the Amazon Polly Text-to-Speech configuration page. At the top right, there are three buttons: "Save to S3", "Download", and "Listen". The "Listen" button is highlighted with a red box. Below these buttons, there's a section titled "Engine Info" with a note about engine compatibility. There are four engine options: "Generative" (selected), "Long-Form", "Neural" (selected), and "Standard". Under "Language", "English, US" is selected. Under "Voice", "Joanna, Female" is selected. In the "Input text" section, the text "Hi! In this Lab you are going to Learn about Amazon Polly and convert your text to Speech." is entered. A "SSML Info" link is also present.

- If you click on Save to S3 option then here you have the ability to save this audio to your desired S3 bucket.

The screenshot shows the "Save to S3" configuration dialog. It has fields for "S3 output bucket" (containing "my-unique-bucket-name"), "S3 key prefix - optional" (containing "myFolder/myFile"), and "SNS topic for notifications - optional" (containing "arn:aws:sns:us-east-1:123456789012:task-status-updates"). At the bottom right are "Cancel" and "Save to S3" buttons, with "Save to S3" being highlighted.

- You can also turn on SSML (Speech Synthesis Markup Language) to check difference between different options.

The screenshot shows the Amazon Polly Text-to-Speech interface. On the right side, there is a sidebar titled "SSML" which contains information about Speech Synthesis Markup Language (SSML). It explains that SSML tags allow you to modify speech output, such as selecting a voice or changing pronunciation. Below this, there are links to "Learn more" and "Generating speech with SSML".

10. Below you can see that you have the code for it. I'll be providing multiple codes to try out SSML.
11. If you get any errors saying not invalid syntax just change the engine or voice.

This screenshot shows the same Amazon Polly interface but with different settings. The "Engine" dropdown is set to "Long-Form". The "Voice" dropdown is set to "Ivy, Female". The "SSML" toggle switch is turned on. The input text field contains the following SSML code:

```
<speak>
  I will speak <prosody rate="slow">slowly</prosody> and
  <prosody rate="fast">quickly</prosody>.
  I can also speak with a <prosody pitch="+10%">higher pitch</prosody>
  or <prosody pitch="-10%">lower pitch</prosody>.
</speak>
```

<speak>

Hello! Welcome to our service.

<break time="500ms"/>

We are here to assist you.

<emphasis level="strong">How can we help you today?</emphasis>

</speak>

<speak>

I will speak **<prosody rate="slow">slowly</prosody>** and
<prosody rate="fast">quickly</prosody>.

I can also speak with a **<prosody pitch="+10%">higher pitch</prosody>**
or **<prosody pitch="-10%">lower pitch</prosody>**.

</speak>

<speak>

Welcome to **<phoneme alphabet="ipa" ph="'æməzən'">Amazon</phoneme>** Polly.

</speak>

<speak>

I can speak **<prosody volume="loud">loudly</prosody>** or
<prosody volume="soft">softly</prosody>.

<amazon:effect name="whispered">I can also whisper like this.</amazon:effect>

</speak>

<speak>

The next step is **<break time="1s"/>** a critical decision.

Please **<break strength="strong"/>** take a moment to think.

</speak>

<speak>

Today is <say-as interpret-as="date" format="mdy">10/20/2024</say-as>.

The meeting starts at <say-as interpret-as="time">3:00 PM</say-as>.

</speak>