Speech-to-Text

cloud.google.com/speech-to-text

Accurately convert speech into text with an API powered by the best of Google's AI research and technology.

New customers get $300 in free credits to spend on Speech-to-Text. All customers get 60 minutes for transcribing and analyzing audio free per month, not charged against your credits.

- ✓ Transcribe your content with accurate captions
- ✓ Enable the power of voice to create better user experiences
- ✓ Improve your service with insights from customer interactions
- ✓ Get started with in-console tutorials in Java, Go, Python, and Node.js

Gartner names Google a Leader in the 2022 Gartner® Magic Quadrant™ for Cloud AI Developer Services report.

Register to download the report

Benefits

State-of-the-art accuracy

Leverage Google's most advanced deep learning neural network algorithms for automatic speech recognition (ASR).

Easy model customization

Experiment with, create, and manage custom resources with the Speech-to-Text UI.

Flexible model deployment

Deploy ASR wherever you need it, whether in the cloud with the API or on-premises with Speech-to-Text On-Prem.

Demo

Put Speech-to-Text into action

As in this demo, you can easily infuse speech transcription into your applications with the Speech-to-Text API.

Key features

Key features

Speech adaptation

Provide hints to boost the transcription accuracy of rare and domain-specific words or phrases. Use classes to automatically convert spoken numbers into addresses, years, currencies, and more.

Domain-specific models

Choose from a selection of trained models for voice control, phone call, and video transcription optimized for domain-specific quality requirements.

Easily compare quality

Experiment on your speech audio with our easy-to-use user interface. Try different configurations to optimize quality and accuracy.

Speech-to-Text On-Prem

Maintain control over your infrastructure and protected speech data by leveraging Google's speech recognition technology on-premises, in your own private data centers. Contact sales to get started.

View all features
BLOG

Unveiling a new user interface for Google Cloud's STT API

Documentation

Tutorials and getting started resources

Tutorial

Speech-to-Text in-console tutorials

Learn how to add Speech-to-Text to your existing apps. Get started quickly with in-console tutorials in Java, Go, Python, and Node.js.

Learn more
Google Cloud Basics

Speech-to-Text basics

Learn the fundamental concepts in Speech-to-Text.

Learn more
Quickstart

Quickstart: Using the Speech-to-Text-UI

Learn how to set up your project and authorization, create and refine a transcription, and learn how to use this configuration in the console.

Learn more
Quickstart

Quickstart: Using the gcloud tool

Send an audio transcription request to Speech-to-Text using the gcloud tool from the command line.

Learn more
Best Practice

Best practices

Review the best practices for transcribing audio with Speech-to-Text.

Learn more
Tutorial

ML onramp

Explore Speech-to-Text tutorials, codelabs, and more.

Learn more
Google Cloud Basics

Supported languages

Learn which languages are available for Speech-to-Text, plus the features and recognition models available for each.

Learn more
Google Cloud Basics

Speech-to-Text On-Prem

Learn more about Speech-to-Text On-Prem, which enables easy integration of Google speech recognition technology into your on-premises solutions.

Learn more
Not seeing what you're looking for?

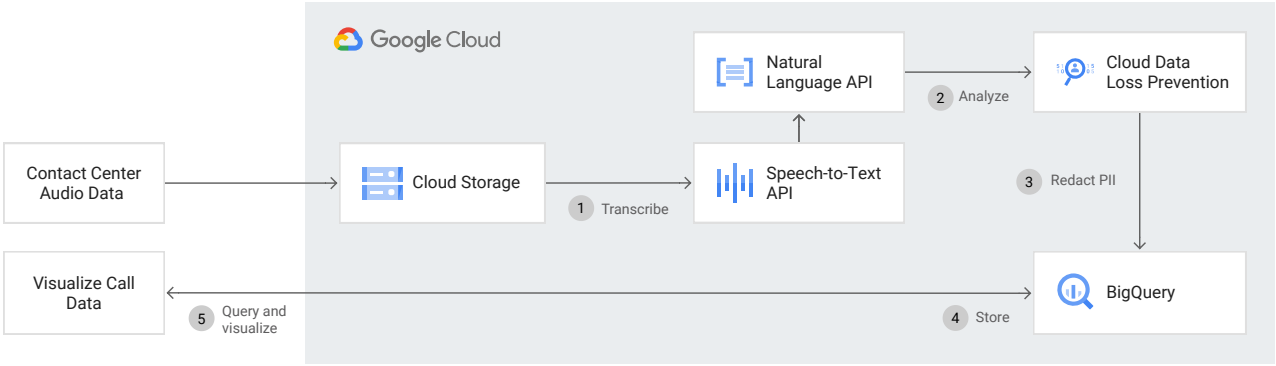
View all product documentation
Use cases

Use cases

Use case

Improve customer service

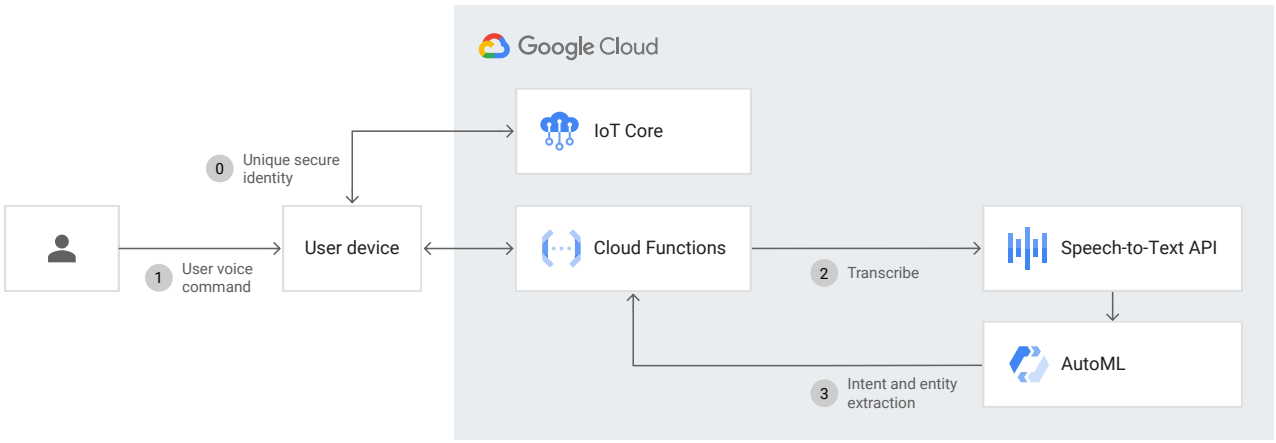
Empower your customer service system by adding IVR (interactive voice response) and agent conversations to your call centers. Perform analytics on your conversation data to gain more insights into the calls and your customers. Speech-to-Text and its enhanced phone call models are already powering Google Cloud's powerful solution, Contact Center AI.



Use case

Enable voice control

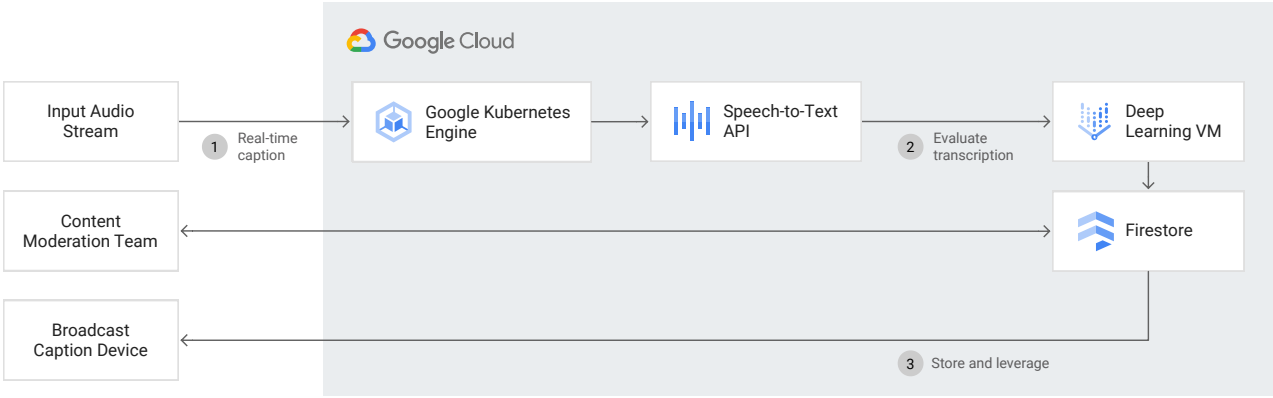
Implement voice commands such as "turn the volume up," and voice search such as saying "what is the temperature in Paris?" Combine this with the Text-to-Speech API to deliver voice-enabled experiences in IoT (Internet of Things) applications.



Use case

Transcribe multimedia content

Transcribe your audio and video to include captions and improve your audience reach and experience. Add subtitles to your content real time to your streaming content. Our video transcription model is ideal for indexing or subtitling video and/or multispeaker content and uses machine learning technology that is similar to video captioning on YouTube.



View all technical guides

All features

All features

Global vocabulary	Support your global user base with Speech-to-Text's extensive language support in over 125 languages and variants.
Streaming speech recognition	Receive real-time speech recognition results as the API processes the audio input streamed from your application's microphone or sent from a prerecorded audio file (inline or through Cloud Storage).
Speech adaptation	Customize speech recognition to transcribe domain-specific terms and rare words by providing hints and boost your transcription accuracy of specific words or phrases. Automatically convert spoken numbers into addresses, years, currencies, and more using classes.
Speech-to-Text On-Prem	Have full control over your infrastructure and protected speech data while leveraging Google's speech recognition technology on-premises, right in your own private data centers. Contact sales to get started.
Multichannel recognition	Speech-to-Text can recognize distinct channels in multichannel situations (e.g., video conference) and annotate the transcripts to preserve the order.
Noise robustness	Speech-to-Text can handle noisy audio from many environments without requiring additional noise cancellation.
Domain-specific models	Choose from a selection of trained models for voice control and phone call and video transcription optimized for domain-specific quality requirements. For example, our enhanced phone call model is tuned for audio originated from telephony, such as phone calls recorded at an 8khz sampling rate.
Content filtering	Profanity filter helps you detect inappropriate or unprofessional content in your audio data and filter out profane words in text results.
Transcription evaluation	Upload your own voice data and have it transcribed with no code. Evaluate quality by iterating on your configuration.

Automatic punctuation (beta)	Speech-to-Text accurately punctuates transcriptions (e.g., commas, question marks, and periods).
Speaker diarization (beta)	Know who said what by receiving automatic predictions about which of the speakers in a conversation spoke each utterance.

Pricing

Pricing

Speech-to-Text is priced per 15 seconds of audio processed after a 60-minute free tier.

View pricing details