

The Ultimate Guide to Speech-to-Text Technology

2022



Table of Contents

- 1 [Executive Summary](#)
- 2 [Why Choose Speechmatics?](#)
- 4 [How Our Technology Makes a Difference](#)
- 6 [We Speak Your Language](#)
- 8 [Unrivalled Accuracy](#)
- 10 [Robust, Scalable, and Flexible Deployment](#)
- 12 [Capabilities and Features](#)
- 13 [How it Works](#)
- 14 [Inside the Speechmatics Engine](#)
- 17 [Summary](#)



Executive Summary

Unlocking the value of your data is crucial for business success in today's competitive global marketplace. And audio data is a key ingredient. Analyzing contact center calls, for example, reveals hidden insights that can help improve the customer experience.

But that's not all. From subtitling and closed captioning to web conferencing transcription, speech recognition technology is increasingly at the heart of business today. That's why Speechmatics is on a mission to harness world-leading machine learning innovation to make speech recognition accessible to everyone, regardless of language or accent.

We pioneered the approach of applying neural networks to speech recognition back in the 1980s. The huge rise in computing power, graphics processing and cloud computing since then means speech-to-text technology is now poised to transform the way companies work. Tedious and laborious tasks can be automated, and new value can be extracted from both live and recorded media.

Speechmatics is spun out of Cambridge – a world-renowned hub of excellence and deep technology. We're on a mission to make voice inclusive and accessible to everyone in the world, regardless of gender, ethnic background, orientation or nationality.

Speechmatics believes in excellence. We believe in innovation, breaking the mould and trying new things, new approaches. Our expertise in machine learning is unlocking some truly world-beating applications underpinned by speech-to-text.

Want to read more content like this?

[Sign up to our newsletter!](#)



Why Choose Speechmatics?

Speechmatics is different to other speech-to-text providers.



Accuracy & Speed

Consistently low word error rate across all languages and use cases.



Language & Coverage

Industry-leading languages, accents, and dialects coverage.



Deployment & Scalability

Use speech-to-text technology in real time or with pre-recorded files, in the cloud or securely on-premises.



Data Privacy & Compliance

Efficient and effective indexing of interactions at scale to enhance record keeping.



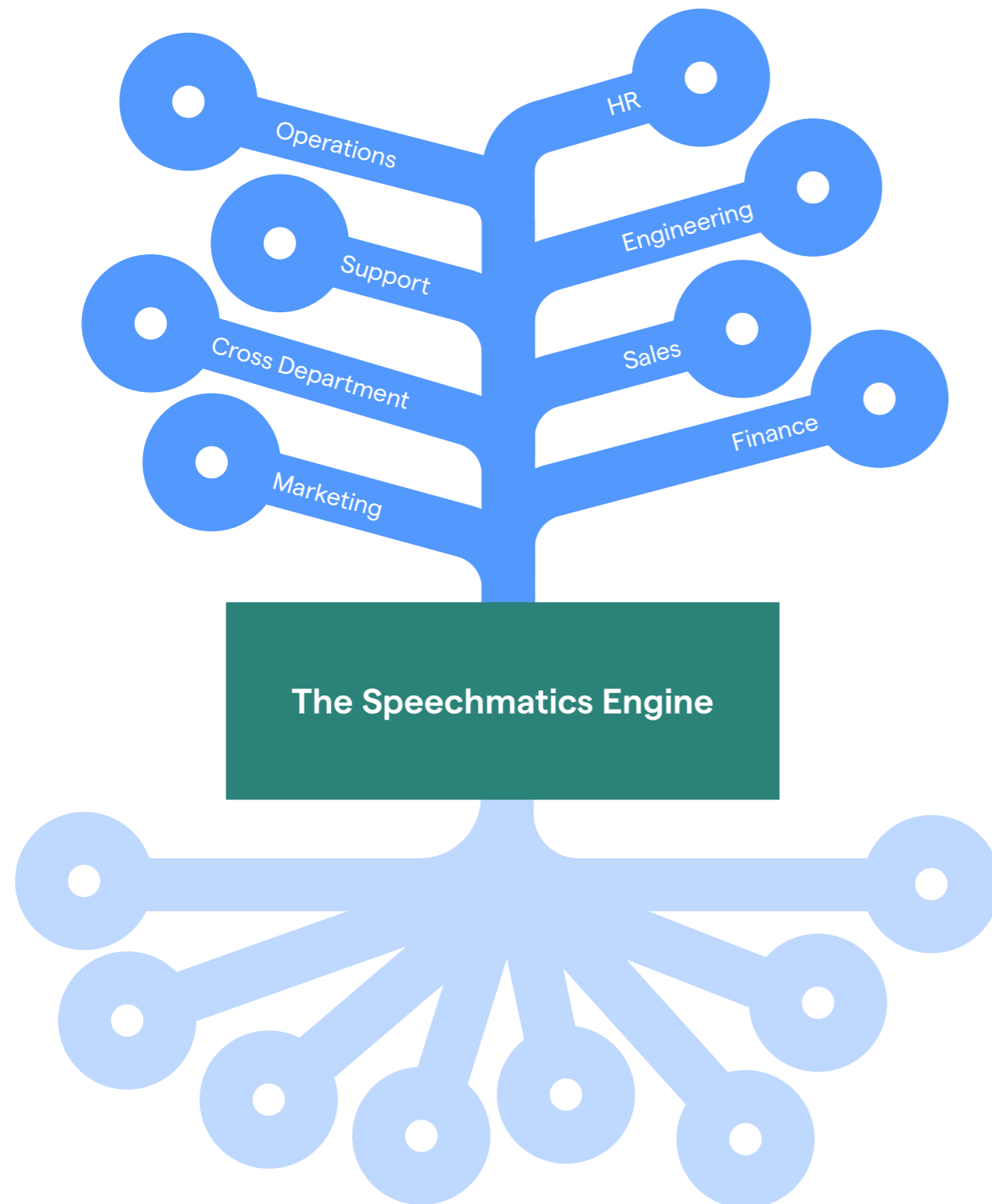
Innovation & Insight

Use machine learning and AI to deliver value to your products and platforms.



How our Technology Makes a Difference

Where we sit in your technology stack



Your spoken data wants to be understood. It's time to use accurate, easy-to-integrate speech recognition technology to unlock the value in your voice data. Speech-to-text is just the beginning – integrating it into your workflows and systems is easy and leads to accurate indexing, analysis and keyword detection, as well as better overall management of your voice data. See how our technology is making a difference:

Media Broadcast

The global media broadcast market is adopting automatic speech recognition technology for live and archived content. Keyword triggers can be set for media monitoring, audio recordings are transformed into searchable transcriptions for media asset management, and live or pre-recorded subtitling can be used in broadcast scenarios. Voice-to-text is bringing automation to media workflows.

Compliance

Legislation is increasing the need to keep data secure. Businesses are using our any-context speech recognition engine to help with compliance and risk management, regulatory intelligence and reporting, and identity and fraud management. Creating transcriptions of call recordings provides searchable content for auditing and compliance, as well as yielding valuable business insights, saving companies time and money while protecting brand reputation.

Contact Centers

Gathering insights from contact center calls has become crucial. Converting call recordings into text enables analysis of audio content to understand the mood, tone and overall sentiment of customers – supporting continuous improvements in customer experience. The searchable content generated can also be used for dispute resolution, compliance, quality management, and event reconstruction.

Transcription

We've made transcribing an interview, a conference or a corporate video as easy as uploading an audio file and receiving an accurate transcript in minutes. For companies providing transcription services, our world-leading automatic speech recognition technology also enables the provision of features such as speaker identification, adjustable timestamps and a customizable dictionary to their customers.

We Speak Your Language

We recognize that you need us to support the languages you do business in. That's why we cover **48 different languages** – from Arabic and Bulgarian through Japanese and Mandarin to Swedish and Turkish.

There are many challenges when you use automatic speech recognition in the real world – including noisy environments, different dialects, accents and context specifics. Our industry-leading expertise ensures our voice-to-text technology works effectively, regardless of the environmental conditions.

We make sure you don't have to adjust the way you speak to be understood. Australian English, American English, Jamaican English, African English – with our speech recognition technology, you don't have to choose. Simply select our [Global English language pack](#) which supports all major English accents. It's been trained on thousands of hours of spoken data from more than 40 countries and tens of billions of words drawn from global sources.

With approximately 500 million speakers globally, Spanish is the second most natively spoken language in the world and fourth most spoken language overall. So, we've also developed a [Global Spanish language pack](#) which supports all major Spanish accents.

The underpinning technology that enables us to build and adapt languages at scale is the Automatic Linguist (AL). AL is an innovative machine learning framework which was accredited with a [Queen's Award for Enterprise](#) in 2019 and helped us build a functional Hindi language pack in just one week.



- | | | | | |
|------------|-----------|-------------|------------|------------|
| Arabic | Dutch | Hungarian | Marathi | Swedish |
| Bashkir | English | Indonesian | Mongolian | Tamil |
| Basque | Esperanto | Interlingua | Norwegian | Thai |
| Belarusian | Estonian | Italian | Polish | Turkish |
| Bulgarian | Finnish | Japanese | Portuguese | Ukrainian |
| Cantonese | French | Korean | Romanian | Uyghur |
| Catalan | Galician | Latvian | Russian | Vietnamese |
| Croatian | German | Lithuanian | Slovakian | Welsh |
| Czech | Greek | Malay | Slovenian | |
| Danish | Hindi | Mandarin | Spanish | |

We are continually innovating to ensure our language offering is both relevant and far-reaching. We already have several more languages in research – and, if you have a need for a language that isn't listed here, [we'd love to speak to you](#) about our roadmap.

Unrivaled Accuracy

We're continually pushing the boundaries of speech recognition technology to stay leaps and bounds ahead of our competitors. These are the results we've seen:

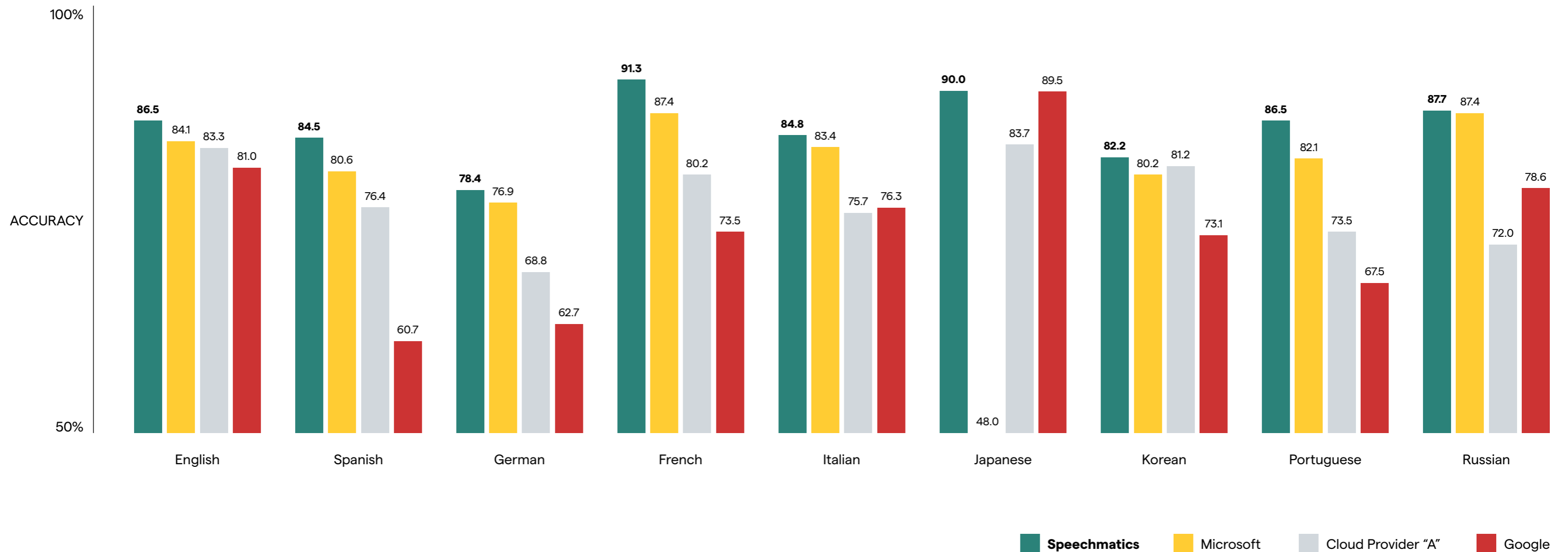
Have a Speech-to-Text Project?
Test our accuracy for yourself!

[Try for Free >](#)



Accuracy % comparison

At the time of writing, we support 48 different languages from Arabic to Welsh. While we won't stop until we cover every country, we're extremely pleased with our results on some of the most frequently spoken.



Robust, Scalable, and Flexible Deployment

Our speech-to-text technology has the flexibility to be deployed anywhere. It can be used on-premises – ensuring data remains within your private environment – with your choice of cloud provider or using Speechmatics' cloud offering.

You'll be using a robust and scalable platform that allows for growth as your business expands. And our extensively tested APIs enable simplified integration into any existing solution architecture – significantly reducing your time to market.

Our deployment options are available for either pre-recorded or real-time media files:



Pre-recorded Speech Recognition

Transcribe pre-recorded media files whenever you want.

Schedule a transcription at a time that suits you – to optimize your available resources.

Real-time Speech Recognition

Get a transcription in real time to gather actionable data instantly.

Our proprietary Low Latency Finals deliver best-in-class accuracy through automatic word correction rescoring. The configurable latency setting is designed for low-latency use cases, including captioning live broadcasts, media monitoring, and agent assist workflows.

Deployment Options

Speechmatics' Cloud Offering

Speechmatics' cloud offering is a fully managed service delivering all the benefits of our speech recognition technology without the complexities of deploying with your own team and environment. Accelerate your time to market with a fully supported and secure service with instant access to all new features, languages, and updates.

The cloud offering is an easy to use, headless application with no user interface required. It is accessed using a [well-documented API integration](#) into your workflow. We support:

- All available features
- All available languages
- Pre-recorded (batch) media files
- Open and extensively documented APIs for operational simplicity
- Cloud instances available in the United States and Europe

On-premises

Our on-premises option enables the transcription of latency or security sensitive media files in your own secure environment or within public cloud environments as part of an existing cloud strategy. Our on-premises deployment option supports:

- Pre-recorded (batch) and real-time audio or video files via Docker containers or virtual appliances
- All available features
- Virtual appliances support flexible language pack deployment options to optimize footprint
- Flexible language selection through batch containers

Public Cloud

Speechmatics' public cloud is agnostic to any cloud environment and supports virtual appliances for both pre-recorded and real-time use cases. This provides the flexibility for our voice-to-text technology to be integrated into your existing cloud strategy, as well as an on-premises deployment model.

Hybrid

Our hybrid deployment supports both real-time requirements and pre-recorded files that need transcribing. It also supports a mixture of data requirements that need some cloud and some on-premises processing.

Capabilities & Features



Precise Timecodes

Precise timecodes provided in an audio or video file enable accurate timestamps - or tags - on a transcript. This feature makes searching a transcript considerably faster.



Speaker Change

Easily identify a change of speaker within your transcript with Speaker Change. A token is automatically added to the transcript each time a speaker change is detected. This enables easy modification of the transcript to improve its readability.



Flexible Deployment

Consume Speechmatics' speech-to-text technology in real time or with pre-recorded files, in the cloud or on-premises. Speechmatics has the flexibility to be deployed according to your needs.



Channel Diarization

Detect and label speakers on up to six streams or channels.



Speaker Diarization

Detect and label different speakers within the same channel.



Advanced Punctuation

Speechmatics' Advanced Punctuation is built on over 2.5 billion words and has an industry-leading set of supported punctuation marks. This use of punctuation optimizes the speed and ease of consuming a transcript for human users.



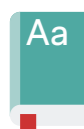
Confidence Scores

Visualize the confidence of every word within the transcript.



Extensive File Format Support

Accelerate your transcript turnaround time by reducing the time it takes for you to prepare your audio or video files. Speechmatics supports an extensive set of file formats so you don't have to worry about converting files to suit our requirements.

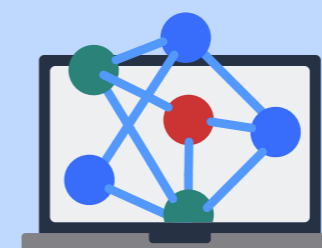


Custom Dictionary & Sounds Feature

Add a set of context-specific words to the dictionary to enhance your transcription accuracy.

How it Works

Innovation is at the heart of what we do. Our team is made up of numerous Cambridge PhDs working at the cutting edge of artificial intelligence, neural networks, machine learning, and language networks. It means our speech-to-text is constantly evolving to provide industry-leading accuracy and performance. And our deep learning expertise ensures our algorithms remain at the forefront of automatic speech recognition development.



Research

Speechmatics is continuously developing new algorithms to unlock rich meaning from speech and text data. We focus on vanguard unsupervised learning research and use the results as the foundation for Speechmatics' ASR.



Train

For each new generation of algorithms that Speechmatics develop, we then train two key neural networks: an Acoustic Network and a Language Network. The Acoustic Network is trained on millions of hand-labeled transcripts and the Language Network processes billions of words from across the Internet.

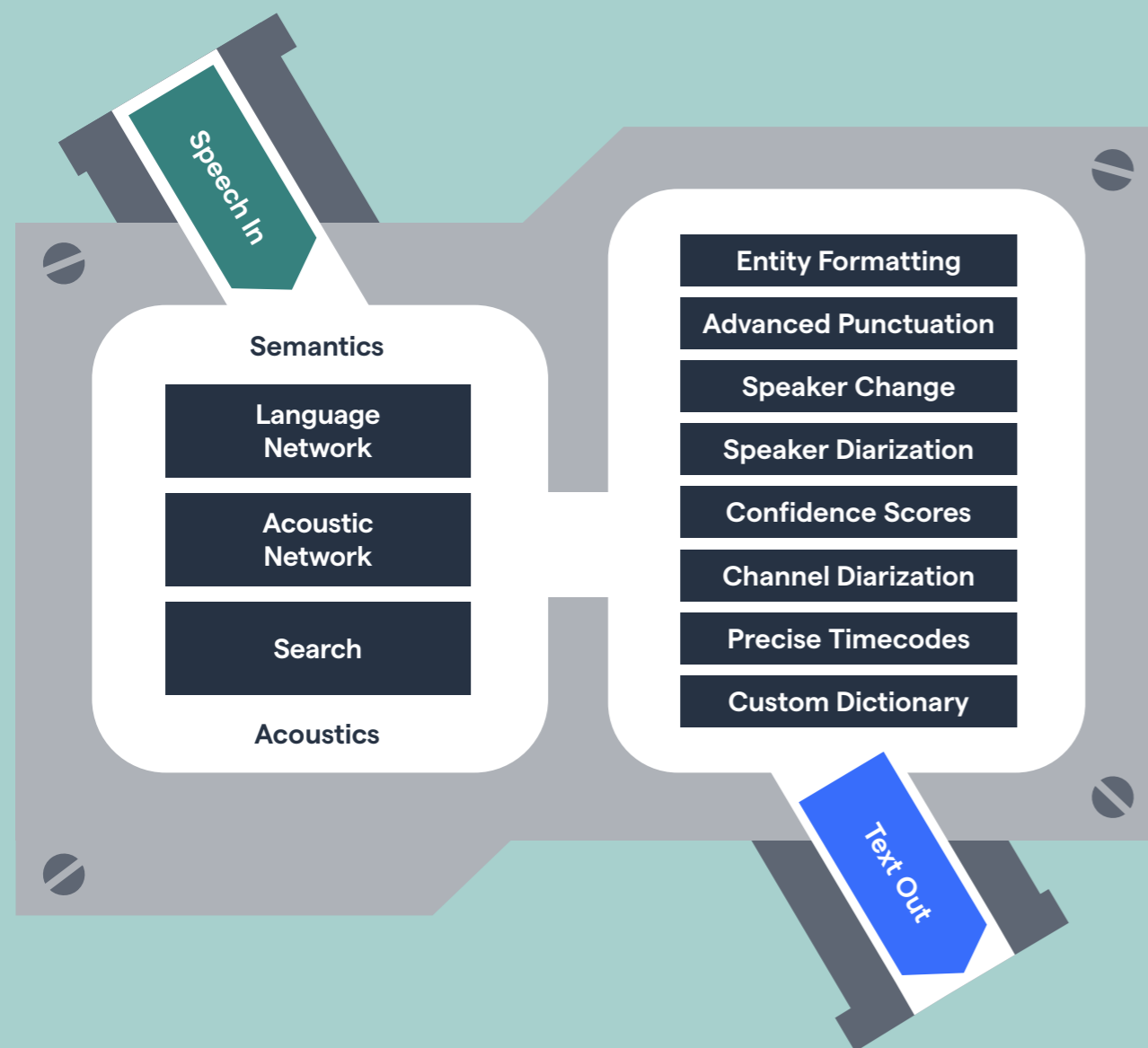


Deploy

Deploy Speechmatics' speech-to-text anywhere, anytime. On-premises, in the cloud, for pre-recorded files or in real time. Our flexible and scalable deployment options were designed for your ease of use, and our [online documentation](#) supports your build into production.

Inside the Speechmatics Engine

Our speech-to-text engine is more powerful and more capable than ever before.



See How it Works

Speech In

Audio is inputted into the Speechmatics engine. Speechmatics supports an extensive set of file formats so you don't have to worry about converting files to suit our requirements.

Custom Dictionary

Add customized wordlists to the Speechmatics dictionary to improve transcription accuracy.

Language Network

Recurrent neural networks (RNNs) estimate the probability of a given sequence of words within a sentence. They can then determine if a sequence of words forms a coherent and meaningful sentence. Speechmatics' Language Network takes the hypothesis transcripts from the Acoustic Network and sorts those that actually make sense from those that don't.

E.g. "That's the right answer" vs "That's ferrite ants are".

Acoustic Network

Convolutional neural networks (CNNs) learn the relationship between an audio signal and the phonemes and other linguistic units that make up speech. Speechmatics' Acoustic Network takes a piece of audio and produces multiple hypothesis transcripts, and corresponding confidences.

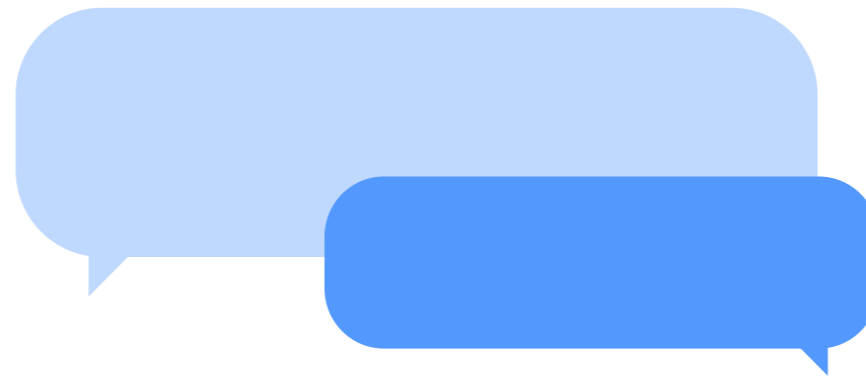
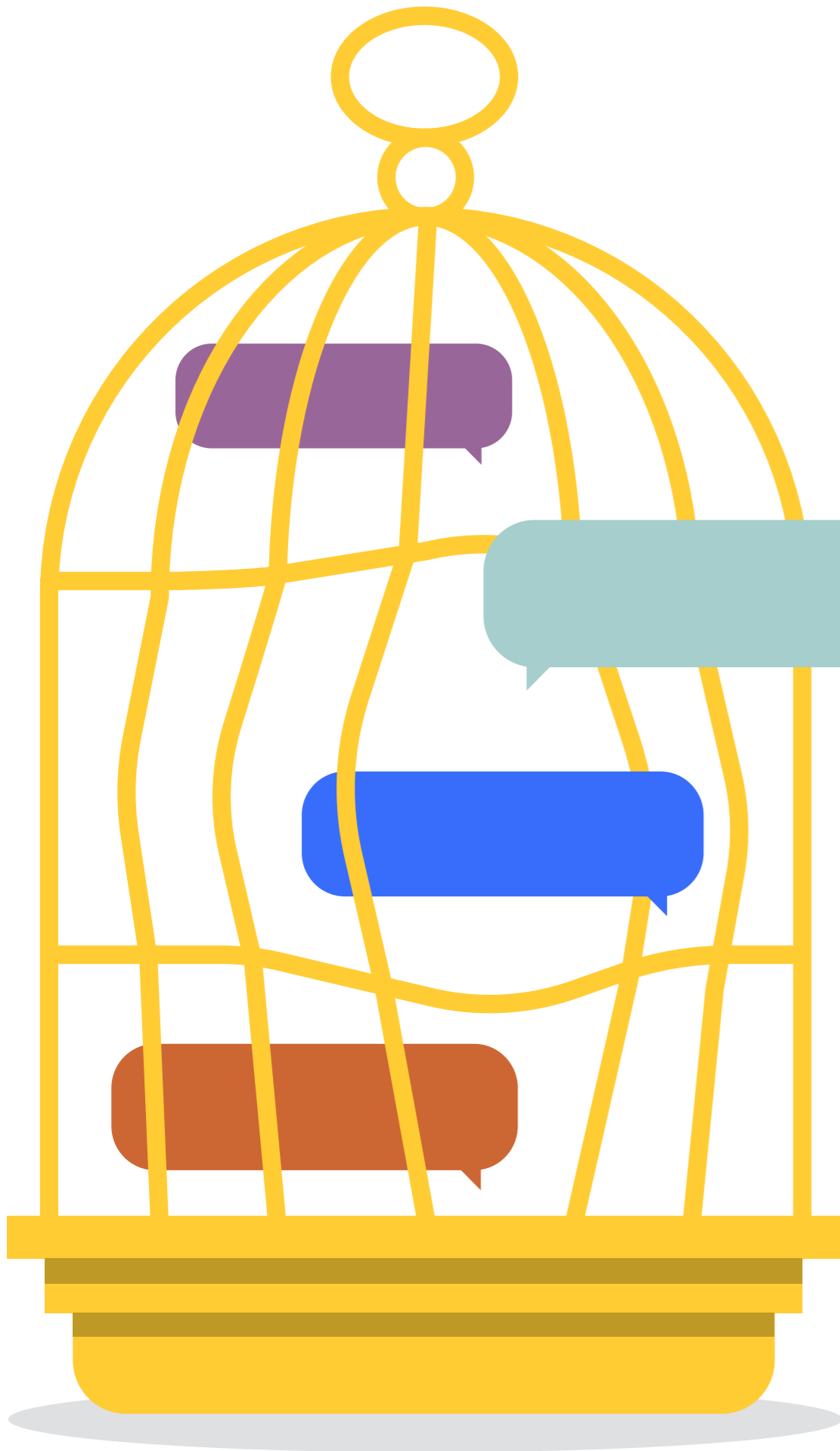
Search

Speechmatics' search module uses a decoder to combine the acoustic knowledge from the Acoustic Network and semantic knowledge from the Language Network to produce our final transcript.

Text Out

A textual transcript is produced with an excellent level of accuracy and understanding. The Speechmatics engine is so advanced that it adjusts to:

- Poor audio quality
- Background noise
- A range of accents and languages
- Varied file types
- Rates of speech
- Custom words



Summary

Speechmatics is harnessing machine learning and neural networks to tackle the challenges of speech recognition – powering applications that require mission-critical, accurate speech-to-text transcription.

We enable innovative applications driven by accurate transcription of any real-time or recorded media. Our technology unlocks meaning and insight from data at scale – we process millions of hours of transcription per month. And our any-context technology adapts as our customers change and grow.

We offer robust, scalable, and flexible control of your data. Our speech recognition engine has the flexibility to be deployed whenever and wherever your business needs it to so you can keep control over personal or sensitive data.

You also benefit from accurate speech recognition, regardless of your accent – with our Global English and Global Spanish language packs supporting all major accents in one model.

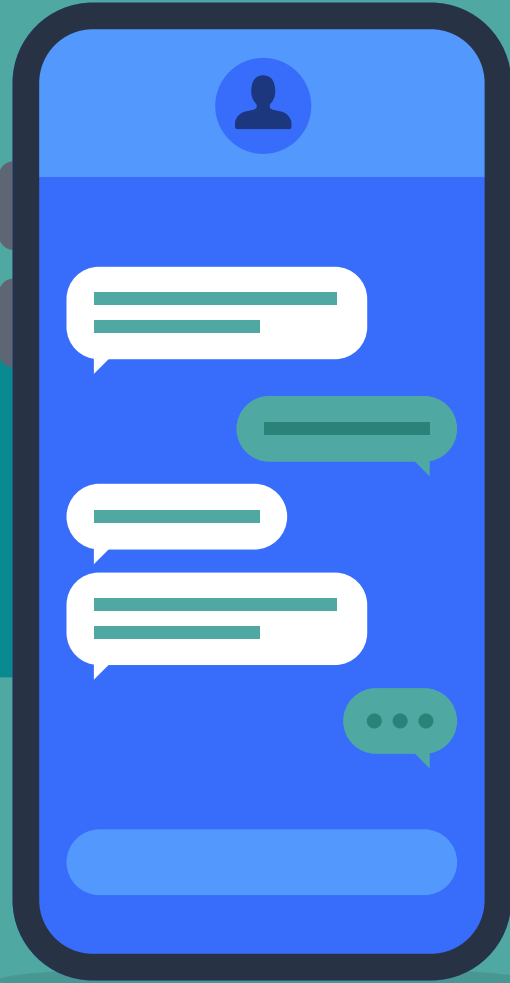
Speechmatics is on a mission to make voice-to-text technology inclusive and accessible to everyone in the world. If you're ready to take the next step in your voice-to-text journey – powered by machine learning – what are you waiting for?

Get in touch and let the power of speech unlock the hidden value in your business.



Speechmatics

Try for Free >



[Speechmatics.com](https://www.speechmatics.com)

296 Science Park
Milton Road, Cambridge
CB4 0WD United Kingdom