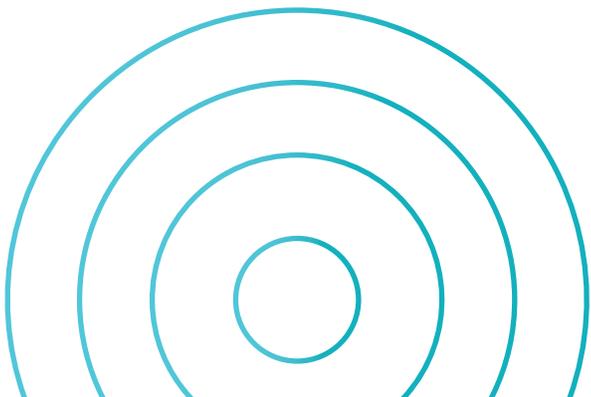




Kore.ai Automated Speech Recognition Engine

Equip Chatbots with Robust Speech-to-Text Capabilities for User Interactions that Result in Complete and Satisfying Experiences



kore.ai

A superior user experience is more than a “nice to have.” It’s a necessity in today’s competitive digital climate. UX impacts revenue (gains and losses) across all industries. Delivering less than satisfying interactions can lead to abandonment of shopping carts, banking transactions, subscription upgrades, and more. Kore.ai has worked hand-in-hand with some of the world’s top enterprises to gauge the pulse of the modern UX and where it’s headed, and we’ve infused that knowledge and expertise into everything we do. We’ve gone to great lengths to ensure that you can build and deploy chatbots that users can interact with in the most human way possible – speech – and help drive lasting and satisfying adoption.

Kore.ai’s Approach to Speech Recognition

Our automated speech recognition (ASR) engine enables voice-driven interactions with chatbots and is a key component of the Kore.ai Bots Platform. It also allows your chatbot to communicate outside of traditional text interfaces or messaging applications, including IoT integrations like IP phones, wearables, and speech-enabled gadgets like Amazon’s Echo and more. This extends your chatbot’s capability to understand voice commands in channels like a website or mobile app, where speech-to-text functionality isn’t typical.

Kore.ai’s unique approach to speech recognition was designed to offer enterprise clients a speech engine that gives them complete control.

- **Complete control over speech engine training** - enterprises control what they feed into the engine and how they handle ongoing training. For example, the engine can be trained with customer names rather than a lengthy list of “common names” that may or may not be applicable to the users.
- **Optional on premise hosting of Kore.ai’s ASR** - access to an on-prem speech solution gives enterprises more flexibility and security when needed.
- **Secure conversations between users and businesses** - control over historical conversations that happen via voice, including personal or financial information.
- **Complete auditability and compliance** - enterprises have access to pull records for any compliance or regulatory issues without having to rely on third party vendors to access their speech data.
- **Direct and ongoing linkage between bot training data and speech engine data** - synonyms, patterns, and other NL training data is automatically fed into the speech engine for its training, no need for separate training on an ongoing basis.

How it works

The Kore.ai ASR engine is built using an open-source speech recognition engine. Like any powerful ASR engine, it requires proper training, which is essential to preventing confusion and a cumbersome UX that can lead to user frustration. The accompanying training corpus includes a collection of spoken texts which give chatbots access to an expansive vocabulary and pronunciation set.

We train the engine with thousands of hours of conversational, broadcast, and read speech. The engine is trained using multiple graphics processing units (GPUs) and CPU's in a distributed setup; this method yields expedited training using large datasets (and more training data leads to better recognition capabilities). More specifically, the engine's extensive vocabulary includes words well-known to enterprise chatbots such as "Salesforce", "Asana" etc., whose pronunciations and context may be unique. It also provides flexibility to clients to impart knowledge of unknown/new vocabulary.

Key Features

1. **Streaming Recognition** Converted text is returned as a stream, while the user is still speaking
2. **Noise Robustness** Works well even with noisy speech input
3. **Domain specific vocabulary** Engine can be customized to recognize domain specific terms
4. **Domain specific phrases** Engine can be improved for recognizing domain specific phrases, by providing a custom training
5. **Mobile and Web SDK** Kore ASR engine has mobile and web SDKs can get you started without having to write code for integration. SDK supports audio streaming and continuous display of converted text as the user is speaking.

Put to the Test

Kore.ai's approach to speech recognition is forward thinking and uses deep neural networks and some of the latest technology to build our ASR. The Kore.ai ASR engine undergoes ongoing training, testing and benchmarking against standard datasets, in order to continually adjust and enhance the engine. Our latest developments are based on state-of-the-art DNN architectures, viz. time-delay neural nets (TDNNs) and long short-term memory nets (LSTMs).

We present the performance of Kore.ai ASR engine in terms of Word Error Rate (WER) on NIST 2000 HUB5 English conversational speech for both Switchboard (SWB) and CallHome (CH) tasks. On a common note, a lower word error rate shows better accuracy in recognition of speech.

Table: WER (%) of different acoustic models

Model	3-gram LM		With LM rescoring	
	CH	SWB	CH	SWB
TDNN-LSTM	16.8	8.5	16.5	8.2

While calibrating the Kore.ai ASR engine with real users invited to talk to the bot, we collected a custom chatbot specific speech corpus for several bots, and found that, on average across multiple domains, the WER was 12%.

[Explore more](#) about Kore.ai's approach to speech recognition and other essential must-haves for building robust, AI-rich chatbots your customers, business partners, and employees will love.