

In-House Transcription

AI Transparency Technical Note

April 2024

Contents

Feature Overview _____ 4

Model Overview _____ 4

 Introduction _____ 4

 Model Architecture _____ 4

 Usage Guidelines _____ 6

 Model Inputs and Outputs _____ 6

Data Sources for Training and Evaluation _____ 6

Model Evaluation and Performance _____ 6

Safety and Ethical Considerations _____ 6

Fairness _____ 7

Privacy and Security _____ 7

Updates and Maintenance _____ 7

License _____ 7

References _____ 8



At Cisco, we appreciate that Artificial Intelligence (AI) can be leveraged to power an inclusive future for all. We also recognize that by applying this technology, we have a responsibility to mitigate potential harm. That is why Cisco adheres to our [Responsible AI Framework](#) (the “Framework”), which is based on six principles of Transparency, Fairness, Accountability, Privacy, Security and Reliability. Cisco translates these principles into product development requirements, which ultimately form part of the product development lifecycle alongside our Security by Design, Privacy by Design, and Human Rights by Design processes.

Accordingly, Webex by Cisco (Webex) features that leverage AI are built with transparency, fairness, accountability, privacy, security, and reliability at their core. Each feature powered by AI undergoes an AI Impact (AI) Assessment – a best-in-class review of how the technical underpinnings of the functionality measure against the Framework Principles.

Webex provides closed captions and transcriptions, documented [here](#), built in-house offered in English, French, German, Spanish, and Italian. Webex uses third parties to deliver closed captions and transcriptions with real-time translation add-on in Chinese (Traditional and Simplified), Dutch, Hindi, Japanese, Korean, Polish, and Portuguese. The document scope is in-house transcriptions.

Webex In-House Transcription was built with the Framework Principles at the center of how we deliver the AI-powered technology. This Technical Note describes more information about the feature and how Cisco responsibly leverages AI deliver the functionality.

Feature Overview

Webex offers closed captions and transcription in English, French, German, Spanish, and Italian to provide a more inclusive experience with Webex Meetings, Webex Events, and Vidcast. The feature is available as closed captioning (live) transcription or offline (batch) transcription. Closed Captioning (Live) Transcription provides live, real-time closed captioning functionalities for the Webex meeting or event experience in five spoken languages of English, French, German, Spanish, and Italian. Offline (Batch) Transcription, provided after the meeting or event, whereby a user receives a full, searchable transcription of the recording. Offline transcription allows users to jump to any portion of the video recording from the transcript; however, is currently only available in English. Offline Transcription is not available for Webex Events or Webinars.

Meeting hosts, participants, and Vidcast creators who have view/edit privileges can review, edit, and share the post-meeting transcript. In addition, users can review, edit, and share the resulting transcript and download the transcript to a text format like txt or vtt. In Webex Meetings users can also view a word-cloud that provides a navigation engine for the main topics of the meeting (located in the review tab of the of the post meeting page).

In-House Transcription is available within Webex Meetings, Webex Events, Webex Contact Center, Webex Calling, Cisco Devices, and Vidcast.

Model Overview

Introduction

In-House Transcription is a Cisco-built machine learning model that leverages Automatic Speech Recognition to provide closed captions and transcription features. In-House Transcription first takes speech audio input, performs feature extraction, decodes with the use of acoustic, language, and other models, and finally produces the text output.

Model Architecture

The In-House Transcription model architecture is inspired by concepts from state-of-the-art research around Automatic Speech Recognition. The model is trained with Cisco unique datasets curated for diverse demographics that reflect customers and users. Webex further

fine tunes the model for specific feature use with Webex Meetings, Webex Events, Webex Contact Center, Webex Calling, Cisco Devices, and Vidcast.

The figure below shows the data flow for In-House Transcription.

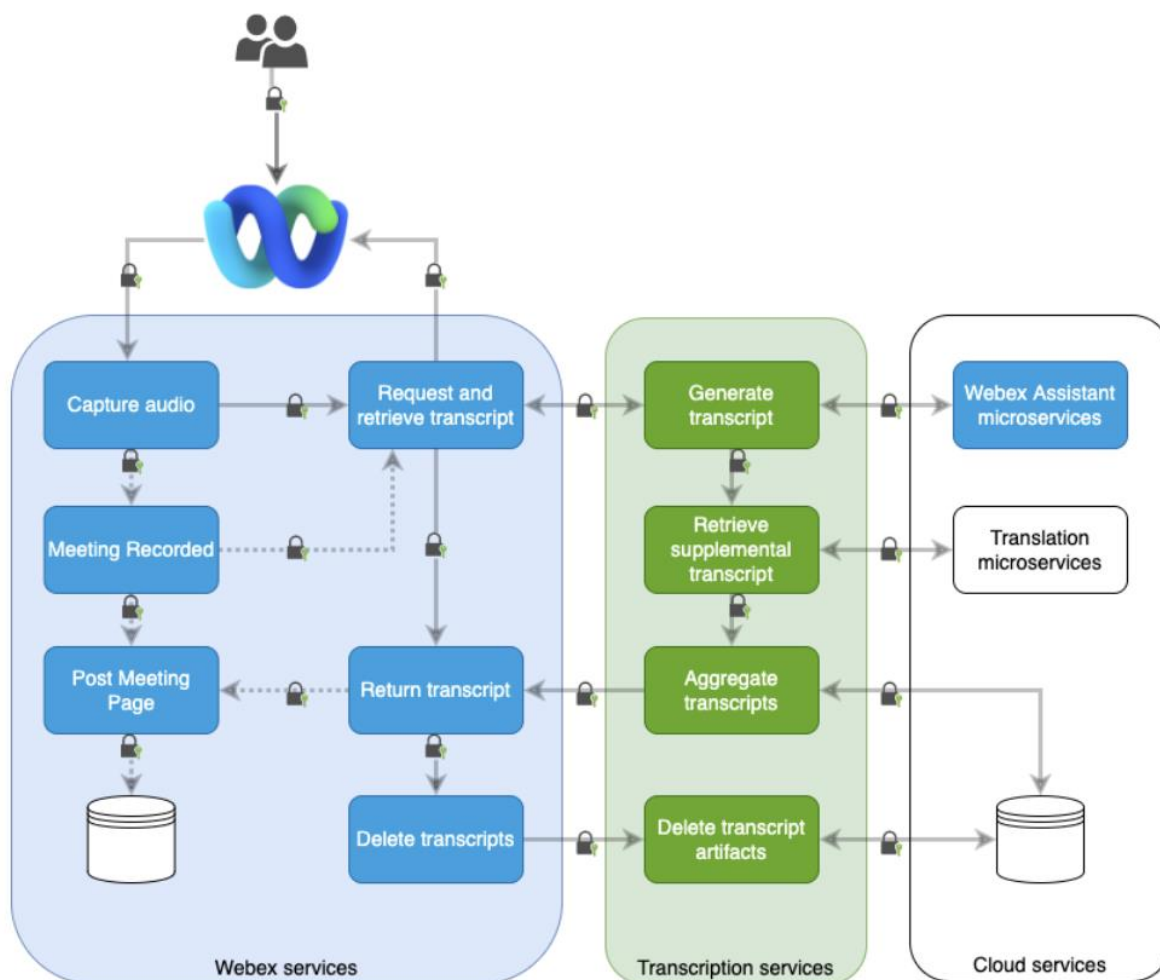


Figure 1: In-House Transcription data flow

When a recording is being processed, the audio is sent securely to Transcription services. The resulting transcript is returned and merged with the video recording, and the resulting data output transcript is deleted after merging. No other meeting or event content data besides the combined recording and transcript is retained to deliver the user with the functionality. The combined transcript recording is stored as user-generated information, with retention details according to the Webex Meetings Privacy Data Sheet, Addendum 3, the Webex Events (formerly “Socio”) Privacy Data Sheet, and the Vidcast Privacy Data Sheet found on the [Cisco Trust Portal](#).

Usage Guidelines

The In-House Transcription model cannot be used outside of Webex or Cisco Products.

Model Inputs and Outputs

Typical inputs to the In-House Transcription model include speech audio. Typical outputs of the model are a transcript of the audio in live or offline transcription.

Data Sources for Training and Evaluation

The In-House Transcription model is trained and evaluated on audio and transcription data. Sources of the data are off-the-shelf and open-source datasets, with appropriate licenses and permissions, and data collected in-product with specific consent from the user administrators. Information on Cisco's use of customer data is documented in the Webex Meetings Privacy Data Sheet, Addendum 3, the Webex Events (formerly "Socio") Privacy Data Sheet, and the Vidcast Privacy Data Sheet on the [Cisco Trust Portal](#).

Cisco leverages third-party vendors to ensure human review of the automated processes involved to provide In-House Transcription, including for training and evaluation. Cisco's third-party vendors undergo rigorous vendor reviews, including security and privacy assessments.

Model Evaluation and Performance

In-House Transcription is trained to work best on domain specific data such as business meetings, events, and webinars audio scenarios. The model is evaluated using the industry standard metric for transcription accuracy called Word Error Rate and benchmarked across other open-source models and third-party APIs such as Google, Microsoft, Amazon, and others.

Safety and Ethical Considerations

Cisco assessed the In-House Transcription feature and its use of AI against the Framework Principles by undertaking an AI Impact (AI) Assessment. The model is trained and evaluated on diverse datasets to ensure subsequent iterations continue to improve and remove bias.

Despite the rigorous safety and ethical scrutiny provided by Cisco in delivering the In-House Transcription, the model may output toxic, harmful, or unsafe content if such content exists in the input audio. Please contact Cisco if you have concerns or feedback about your experience with the In-House Transcription feature.

Fairness

In-House Transcription is trained and evaluated on diverse datasets that cover a wide range of demographics and audio inputs supported by Cisco, by sourcing data that is diverse in accent, dialect, age, pitch, and content.

Privacy and Security

Cisco addresses processing of and security around personal data, including retention periods for that data, in its Privacy Data Sheets, found on the [Cisco Trust Portal](#). Audio and transcript information is retained to provide users with the service and will be deleted once it is no longer necessary to provide the service. While Cisco operates the service, Cisco does not access or retain audio or transcript data unless the user opts into data collection used for product improvements.

Updates and Maintenance

In-House Transcription model is continuously evaluated using diverse datasets that would identify known biases, for example, a bias towards one accent over another can be identified by using these diverse datasets. User identified issues are also considered when developing subsequent iterations.

Ongoing Webex updates and product release documents include product updates and any improvements in user experience. Webex may also use customer email notifications and Webex Help Center documentation to inform customers of product changes, as well as communicate product updates in blogs and videos.

License

The In-House Transcription model is proprietary and not publicly available. Cisco uses it only to deliver this feature in the context of Webex applications, Cisco devices, or with Webex SDKs.

References

[RAI Principles](#)

[RAI Framework](#)

[Cisco Trust Portal](#)

[Real-time transcription and translation in meetings and webinars](#)