Transform Webex with AI-Driven Processes

IBM Watson Media

> 42 percent of web collaboration users strongly agree that their organization needs to do more to capitalize on currently-available video technologies.
>
> Vonder Haar, S., Wainhouse Research

Video conferencing is growing. In fact, Gartner forecasts a compound annual growth rate of 9 percent for cloud-based web/videoconference users from 2017 to 2022[1]. On-demand video use is also increasing, with 46 percent of executives who use web collaboration sessions at least once a week reporting that they also watch on-demand video for business purposes every day[2].

But videoconference users are eager to expand how they utilize this technology. In a survey of over 2,000 executives, Wainhouse Research found that 42 percent of web collaboration users strongly agree that their organization needs to do more to capitalize on currently-available video technologies[3].

One method is to introduce artificial intelligence (AI) to this process to enhance how users interact with recorded sessions. This white paper details how an integration between IBM® Enterprise Video Streaming and Webex Meetings can help make that possible.

Broadcasting and recording for AI analysis

With this integration, Webex Meetings users can click a "Start Live Streaming" option to push their videoconferencing session to a passive audience who won't actively participate in the conference.
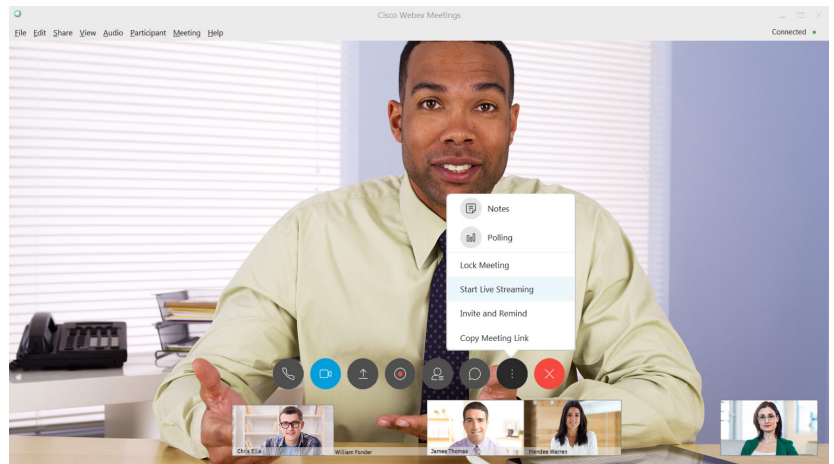
Webex Meetings with IBM Video Streaming integration

Live transcoding is applied as part of this process to help enable access to a wide range of devices, such as computers and mobile phones. Adaptive bitrates are also created and served to viewers based on their connection speed. This feature is designed to help ensure that while viewers with a faster connection will receive a better-quality video, those on a slower connection can still watch the content.

Once the conference is over, the session is automatically recorded and archived for use by the organization that owns the account. At this point, IBM Watson® is introduced, analyzing the content to increase how users can interact with these archived sessions.

...85 percent of the social platform's video content is watched without sound.

Patel, S., Digiday

Using AI to help enhance accessibility

One way AI helps improve stored assets is by broadening accessibility. In this case, AI automates the development of closed captions to help improve accessibility for those who have suffered a level of hearing loss.

AI can also help address changing viewing habits by recognizing when people may desire to watch without volume due to either personal preferences or their current circumstances, such as sitting in an office or commuting on public transportation. Facebook has seen profound impacts from these changes as a projected 85 percent of the social platform's video content is watched without sound[4].

The workflow starts by designating a language. This can be done per video or on a channel basis, which is often organized by category. Keep in mind that channels are defined by the enterprise. For example, a channel can be designated for engineering meetings, town halls or another purpose that makes sense for the use case.

Assuming the selected language is supported by Watson, captions are automatically generated for the content. The speed at which these captions are generated is constantly decreasing, originally being the length of the content – so a 60-minute video would take 60 minutes to caption – to something less than the length now.
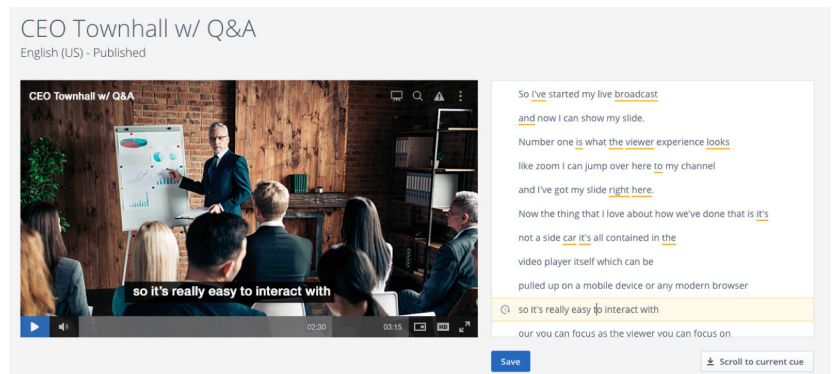
Caption editor

Once generated, captions can be set to either automatically publish or require manual review before release. The dashboard interface includes an editor that allows a content manager to watch the video with the captions. Users can skip to certain parts and click inside a line to edit. Certain words may be underlined as well to denote less confidence from the AI and encourage manual review. A variety of different elements can lead to reduced AI confidence, such as background noises, strong accents or slurred speech.

After being published, the captions may be set to display automatically or require the viewer to enable them inside the player.

IBM Watson Media
Transform Webex with AI-Driven Processes

79 percent of executives who use video archives agree to being frustrated with their inability to quickly find a piece of information when needed.

Vonder Haar, S., Wainhouse Research

## Using AI to address library growth through discoverability

Another improvement offered by AI is helping enhance asset discoverability. Traditionally, content structure was focused on user navigation and provided many categories from which viewers could choose.

But content libraries have grown. In a survey of 1,801 executives, 19 percent noted that their archives are growing by at least 25 hours' worth of video every month[5]. A separate study found that 71 percent of executives report that recording previously-live content makes up the majority of their video archive while 29 percent said their archive is exclusively recorded-live content[6].

This growth in libraries introduces a challenge for finding relevant content among a sea of video assets. 79 percent of executives who use video archives agree to being frustrated with their inability to quickly find a piece of information when needed[7]. Searchability can help address the issues created by these growing video libraries.

One method for powering that search is through manually classifying video. This process includes creating metadata that can be associated with the content, such as titles, descriptions and other tags that may be added. But these manual efforts become harder to scale as the speed of content creation grows.

AI can provide significant value by automating the process of analyzing video that is archived after a Webex session is over. IBM Watson AI reviews the content for insights by using speech-to-text to transcribe the audio into a searchable format. This allows users to search on topics, names and other aspects that were brought up during these archived Webex sessions. These searches can return relevant assets at the library level or find content inside the video, highlighting relevant moments.
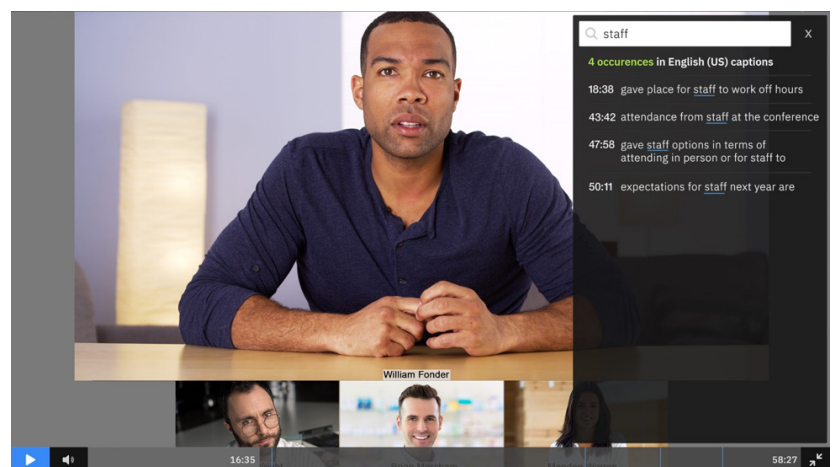
Search capability inside the video player

The value of better utilizing archives with this approach is noted by executives. 82 percent state that speech-to-text features that make it possible to search for – and jump directly to – specific video passages would encourage them to more frequently use on-demand libraries for reference[8].

AI, ASR and speech-to-text

Both the enhanced accessibility through closed captions and the improved discoverability lean on converting content into a machine-readable format, such as text. This process involves Automated Speech Recognition (ASR) while using the AI to improve accuracy.

This multifaceted conversion procedure includes:

– **Speech recognition**
  After receiving the audio, the AI matches the speech to known text. As the technology has advanced, so has its flexibility, moving toward the ability to decipher speech even when accents and dialects are present. However, accuracy will still not be as high as simple speech spoken plainly.

– **AI vocabulary**
  The process of matching the speech to text involves cross checking against a vocabulary list of known terms. The AI is not set up to invent words, meaning it will stick to the vocabulary list. If an unfamiliar term is mentioned, the solution will try to link it to a known term. For example, if your company is launching a new product called "Appssen," the solution might transcribe it as "app sun" due to a close proximity.

– **Audio recognition**
  Not all sounds are words. Coughs or squeaks from chairs during a meeting are noises, not speech. As a result, the AI can be trained to separate these and other noises from actual language. This can be helpful in avoiding the AI trying to match something that isn't a word to its vocabulary list.

– **Language identification**
  Multi-language content can be tricky. If a meeting is conducted in English but someone plays a video or message in Spanish, this can introduce a complex scenario.

  In these instances, it's beneficial for the technology to identify the different languages at a given time, preventing it from trying to match words from a language that is different than its vocabulary list. IBM Watson Media does limit organizations to picking a single language from which the AI can detect and utilize a vocabulary list.

IBM Enterprise Video Streaming enables the enterprise to retain ownership of their assets and related content.

– **Context**
Many phrases and words sound similar, causing issues for converting speech to text. Accurately transcribing homophones, such as "ate" and "eight," is often impossible without the contextual placement of these words.

AI can be trained to distinguish the difference between homophones by analyzing their placement in relation to other words. This helps AI interpret the intended meaning and choose the correct word.

Content ownership and security

Streaming content through IBM Enterprise Video Streaming enables the enterprise to retain ownership of their assets and related content, such as captions or transcripts generated by the AI.

Owners may also choose how they will restrict content, which is often done through linking the account to a corporate directory with support for Okta, OneLogin and others. This method enables employees to access these archived Webex sessions using login credentials that are familiar to them.

Note:
The integration mentioned in this white paper requires version 39.x of Webex Meetings. Please contact IBM Watson Media about enabling the feature. This Webex and live video streaming integration capability is provided and supported exclusively by IBM.

IBM Watson Media
Transform Webex with AI-Driven Processes

Footnotes

1. Fasciani, M., "How AI Is Transforming the Videoconference Room Experience (https://www.gartner.com/document/3893166)", Gartner, 2018

2. Vonder Haar, S., "Capitalizing on Enterprise Video Convergence (https://info.video.ibm.com/IBM-Enterprise-Nurture_Capitalizing-on-Enterprise-Video-Convergence-LP.html)," Wainhouse Research/ IBM, 2019

3. Vonder Haar, S., "Capitalizing on Enterprise Video Convergence (https://info.video.ibm.com/IBM-Enterprise-Nurture_Capitalizing-on-Enterprise-Video-Convergence-LP.html)," Wainhouse Research/ IBM, 2019

4. Patel, S., "85 Percent of Facebook Video is Watched Without Sound (https://digiday.com/media/silent-world-facebook-video/)," Digiday, May 2016

5. Vonder Haar, S., "Unlocking the Hidden Value of Business Video (https://info.video.ibm.com/unlocking-hidden-value-business-video.html)," Wainhouse Research/ IBM Cloud™ Video, 2017

6. Vonder Haar, S., "Capitalizing on Enterprise Video Convergence (https://info.video.ibm.com/IBM-Enterprise-Nurture_Capitalizing-on-Enterprise-Video-Convergence-LP.html)," Wainhouse Research/ IBM, 2019

7. Vonder Haar, S., "Unlocking the Hidden Value of Business Video (https://info.video.ibm.com/unlocking-hidden-value-business-video.html)," Wainhouse Research/ IBM Cloud Video, 2017

8. Vonder Haar, S., "Capitalizing on Enterprise Video Convergence (https://info.video.ibm.com/IBM-Enterprise-Nurture_Capitalizing-on-Enterprise-Video-Convergence-LP.html)," Wainhouse Research/ IBM, 2019

IBM **Watson Media**