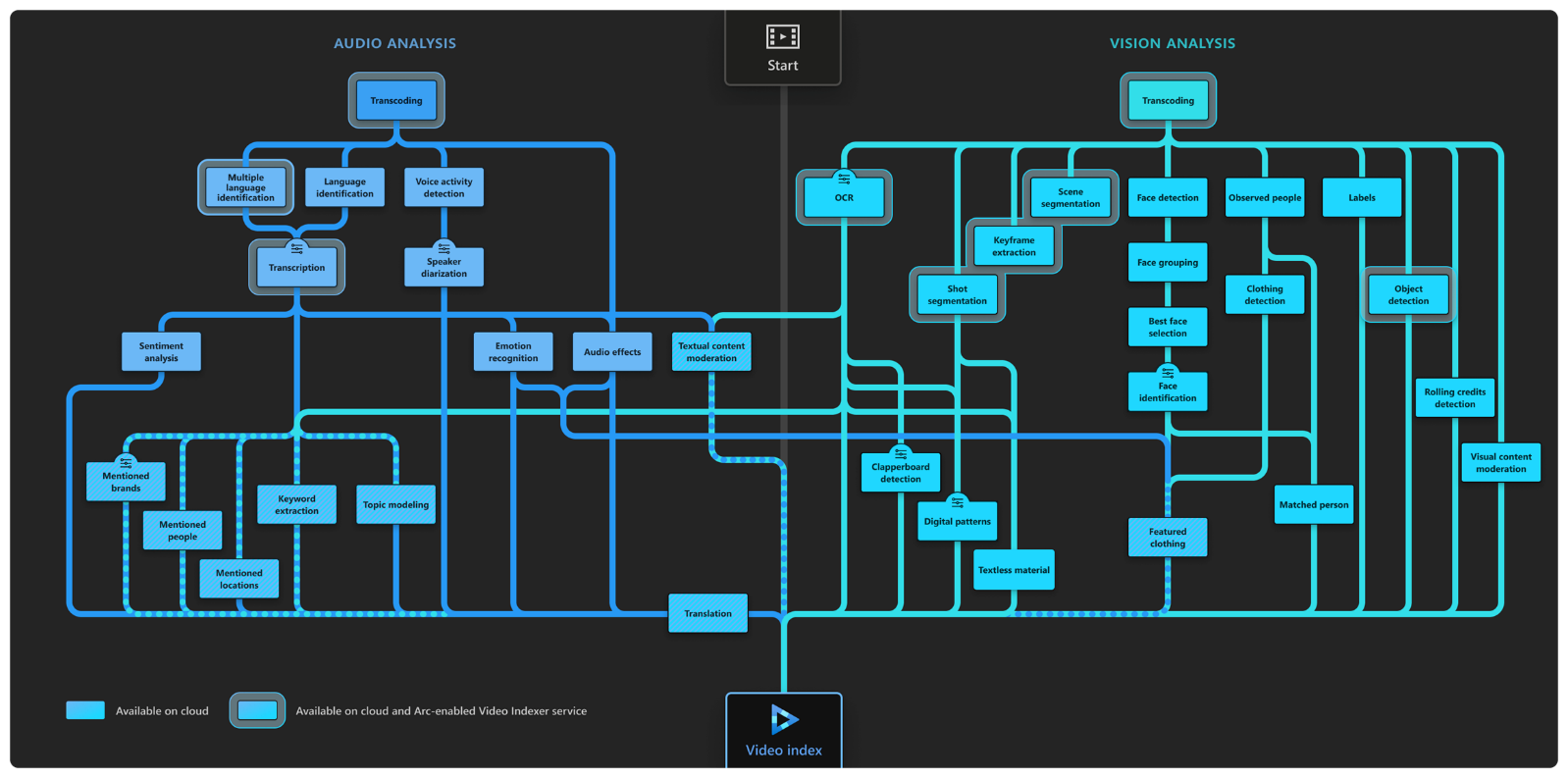
**Use Azure AI Video Indexer to extract insights from a video or live stream**

Azure AI Video Indexer is a cloud application, part of Azure AI services, built on Azure AI services (such as the Face, Translator, Azure AI Vision, and Speech). It enables you to extract the insights from your videos using Azure AI Video Indexer video and audio models.

Azure AI Video Indexer analyzes the video and audio content by running 30+ AI models, generating rich insights. Here is an illustration of the audio and video analysis performed by Azure AI Video Indexer in the background:

[](https://learn.microsoft.com/en-us/azure/azure-video-indexer/media/video-indexer-overview/model-chart.png#lightbox)

**What can I do with Azure AI Video Indexer?**

Azure AI Video Indexer's insights can be applied to many scenarios:

* Deep search: Use the insights extracted from the video to enhance the search experience across a video library. For example, indexing spoken words and faces can enable the search experience of finding moments in a video where a person spoke certain words or when two people were seen together. Search based on such insights from videos is applicable to news agencies, educational institutes, broadcasters, entertainment content owners, enterprise LOB apps, and in general to any industry that has a video library that users need to search against.
* Content creation: Create trailers, highlight reels, social media content, or news clips based on the insights Azure AI Video Indexer extracts from your content. Keyframes, scenes markers, and timestamps of the people and label appearances make the creation process smoother and easier, enabling you to easily get to the parts of the video you need when creating content.
* Accessibility: Whether you want to make your content available for people with disabilities or if you want your content to be distributed to different regions using different languages, you can use the transcription and translation provided by Azure AI Video Indexer in multiple languages.
* Monetization: Azure AI Video Indexer can help increase the value of videos. For example, industries that rely on ad revenue (news media, social media, and so on) can deliver relevant ads by using the extracted insights as additional signals to the ad server.
* Content moderation: Use textual and visual content moderation models to keep your users safe from inappropriate content and validate that the content you publish matches your organization's values. You can automatically block certain videos or alert your users about the content.
* Recommendations: Video insights can be used to improve user engagement by highlighting the relevant video moments to users. By tagging each video with additional metadata, you can recommend to users the most relevant videos and highlight the parts of the video that matches their needs.

**Video/audio AI features**

* The following list shows the insights you can retrieve from your video/audio files using Azure AI Video Indexer video and audio AI features (models).

**Video models**

* **Face detection**: Detects and groups faces appearing in the video.
* **Celebrity identification**: Identifies over 1 million celebrities—like world leaders, actors, artists, athletes, researchers, business, and tech leaders across the globe. The data about these celebrities can also be found on various websites (IMDB, Wikipedia, and so on).
* **Account-based face identification**: Trains a model for a specific account. It then recognizes faces in the video based on the trained model. For more information, see [Customize a Person model from the Azure AI Video Indexer website](https://learn.microsoft.com/en-us/azure/azure-video-indexer/customize-person-model-with-website) and [Customize a Person model with the Azure AI Video Indexer API](https://learn.microsoft.com/en-us/azure/azure-video-indexer/customize-person-model-with-api).
* **Thumbnail extraction for faces**: Identifies the best captured face in each group of faces (based on quality, size, and frontal position) and extracts it as an image asset.
* **Optical character recognition (OCR)**: Extracts text from images like pictures, street signs and products in media files to create insights.
* **Visual content moderation**: Detects adult and/or racy visuals.
* **Labels identification**: Identifies visual objects and actions displayed.
* **Scene segmentation**: Determines when a scene changes in video based on visual cues. A scene depicts a single event, and it's composed by a series of consecutive shots, which are semantically related.
* **Shot detection**: Determines when a shot changes in video based on visual cues. A shot is a series of frames taken from the same motion-picture camera. For more information, see [Scenes, shots, and keyframes](https://learn.microsoft.com/en-us/azure/azure-video-indexer/scene-shot-keyframe-detection-insight).
* **Black frame detection**: Identifies black frames presented in the video.
* **Keyframe extraction**: Detects stable keyframes in a video.
* **Rolling credits**: Identifies the beginning and end of the rolling credits in the end of TV shows and movies.
* **Editorial shot type detection**: Tags shots based on their type (like wide shot, medium shot, close up, extreme close up, two shot, multiple people, outdoor and indoor, and so on). For more information, see [Editorial shot type detection](https://learn.microsoft.com/en-us/azure/azure-video-indexer/scene-shot-keyframe-detection-insight#keyframe-editorial-shot-type-detection).
* **Observed people detection**: Detects observed people in videos and provides information such as the location of the person in the video frame (using bounding boxes) and the exact timestamp (start, end) and confidence when a person appears. For more information, see [Trace observed people in a video](https://learn.microsoft.com/en-us/azure/azure-video-indexer/observed-matched-people-insight).
  + **Matched person**: Matches people that were observed in the video with the corresponding faces detected. The matching between the observed people and the faces contain a confidence level.
  + **Detected clothing**: Detects the clothing types of people appearing in the video and provides information such as long or short sleeves, long or short pants and skirt or dress. The detected clothing is associated with the people wearing it and the exact timestamp (start, end) along with a confidence level for the detection are provided.
  + **Featured clothing**: Captures featured clothing images appearing in a video. You can improve your targeted ads by using the featured clothing insight. For information on how the featured clothing images are ranked and how to get the insights, see [featured clothing](https://learn.microsoft.com/en-us/azure/azure-video-indexer/observed-people-featured-clothing).
* **Object detection** Detects unique objects that are also tracked so that if they return to the frame they are recognized. See [Azure AI Video Indexer object detection](https://learn.microsoft.com/en-us/azure/azure-video-indexer/object-detection)
* **Slate detection**: Identifies the following movie post-production insights when indexing a video using the advanced indexing option:
  + Clapperboard detection with metadata extraction.
  + Digital patterns detection, including color bars.
  + Textless slate detection, including scene matching.

For details, see [Slate detection](https://learn.microsoft.com/en-us/azure/azure-video-indexer/slate-detection-insight).

* **Textual logo detection**: Matches a specific predefined text using Azure AI Video Indexer OCR. For example, if a user created a textual logo: "Microsoft", different appearances of the word *Microsoft* will be detected as the "Microsoft" logo. For more information, see [Detect textual logo](https://learn.microsoft.com/en-us/azure/azure-video-indexer/detect-textual-logo).

**Audio models**

* **Audio transcription**: Converts speech to text over 50 languages and allows extensions. For more information, see [Azure AI Video Indexer language support](https://learn.microsoft.com/en-us/azure/azure-video-indexer/language-support).
* **Automatic language detection**: Identifies the dominant spoken language. For more information, see [Azure AI Video Indexer language support](https://learn.microsoft.com/en-us/azure/azure-video-indexer/language-support). If the language can't be identified with confidence, Azure AI Video Indexer assumes the spoken language is English.
* **Multi-language speech identification and transcription**: Identifies the spoken language in different segments from audio. It sends each segment of the media file to be transcribed and then combines the transcription back to one unified transcription. For more information about transcription see [Transcription](https://learn.microsoft.com/en-us/azure/azure-video-indexer/transcription-translation-lid-insight)
* **Closed captioning**: Creates closed captioning in three formats: VTT, TTML, SRT.
* **Two channel processing**: Auto detects separate transcript and merges to single timeline.
* **Noise reduction**: Clears up telephony audio or noisy recordings (based on Skype filters).
* **Transcript customization** (CRIS): Trains custom speech to text models to create industry-specific transcripts. For more information, see [Customize a Language model](https://learn.microsoft.com/en-us/azure/azure-video-indexer/customize-language-model-how-to).
* **Speaker enumeration**: Maps and understands which speaker spoke which words and when. Sixteen speakers can be detected in a single audio-file.
* **Speaker statistics**: Provides statistics for speakers' speech ratios.
* **Textual content moderation**: Detects explicit text in the audio transcript.
* **Text-based emotion detection**: Emotions such as joy, sadness, anger, and fear that were detected via transcript analysis.
* **Translation**: Creates translations of the audio transcript to many different languages. For more information, see [Azure AI Video Indexer language support](https://learn.microsoft.com/en-us/azure/azure-video-indexer/language-support).
* **Audio effects detection**: Detects the following audio effects in the non-speech segments of the content: alarm or siren, dog barking, crowd reactions (cheering, clapping, and booing), gunshot or explosion, laughter, breaking glass, and silence.

The detected acoustic events are in the closed captions file. The file can be downloaded from the Azure AI Video Indexer website.

**Audio and video models (multi-channels)**

When indexing by one channel, partial results for those models are available.

* **Keywords extraction**: Extracts keywords from speech and visual text.
* **Named entities extraction**: Extracts brands, locations, and people from speech and visual text via natural language processing (NLP).
* **Topic inference**: Extracts topics based on various keywords (that is, keywords 'Stock Exchange', 'Wall Street' produces the topic 'Economics'). The model uses three different ontologies ([IPTC](https://iptc.org/standards/media-topics/), [Wikipedia](https://www.wikipedia.org/) and the Video Indexer hierarchical topic ontology). The model uses transcription (spoken words), OCR content (visual text), and celebrities recognized in the video using the Video Indexer facial recognition model.
* **Artifacts**: Extracts rich set of "next level of details" artifacts for each of the models.
* **Sentiment analysis**: Identifies positive, negative, and neutral sentiments from speech and visual text.

**Compliance, privacy and security**

You must comply with all applicable laws in your use of Azure AI Video Indexer, and you may not use Azure AI Video Indexer or any Azure service in a manner that violates the rights of others, or that may be harmful to others.

Before uploading any video/image to Azure AI Video Indexer, You must have all the appropriate and legal rights to use the video/image, including, where required by law, all the necessary consents from individuals (if any) in the video/image, for the use, processing, and storage of their data in Azure AI Video Indexer and Azure. Some jurisdictions may impose special legal requirements for the collection, online processing and storage of certain categories of data, such as biometric data. Before using Azure AI Video Indexer and Azure for the processing and storage of any data subject to special legal requirements, you must ensure your use complies with all such legal requirements that may apply to You and your intended use.

**Get a trial account**

Go to the [web portal](https://www.videoindexer.ai/) and sign in with an Entra account, a personal Microsoft account, or a Google account. You're automatically assigned a trial account.

**Upload a video**

Upload and index the video using the default settings. The activity of indexing a video is called a "job."

On the home page of the web portal:

1. Select the **Upload** button. The upload screen appears.
2. Select **Browse for files**.
3. Select the video, then select **Open**.
4. The **video name** field is auto populated, but you can change it if you want to.
5. Leave the rest of the settings the way they are except for the source language.
6. Select the source language from the **Source language** drop-down menu.
7. Select **Review + upload**. The review screen appears.
8. Select the **rights certification** checkbox, then select **Upload + index**. The video starts uploading. You can close the screen by selecting **Run in background**.
9. If you closed the uploading window, select the **notification** (bell) icon to check on the status of the upload.

**View the timeline and insights**

The timeline and Insights are the resulting information returned by the service after indexing the video.

While your video is uploading, view some of the sample videos.

**View the timeline**

1. Select the **Timeline** tab. The video transcript appears.
2. Scan the transcript of the spoken audio of the video to get an idea of what was covered in the video.

**View the insights**

1. Select the **Samples** tab.
2. Select any one of the videos from the Samples library.
3. Select **View** from the menu.
4. Select **Monitoring** to deselect most of the insights.
5. Uncheck the remaining insight checkboxes. The insights area of the screen should be empty.
6. Select and unselect the checkboxes one at a time to see the resulting insights.

**View your video timeline and insights**

1. Select **Library.** Your indexed videos are located here. If your video is still uploading, it's represented by a thumbnail with the percentage of the video upload indicated.
2. If the video indexing is complete, select **your video** to see the insights.
3. Compare the insights you received to the ones that you viewed in the Samples library.

**View the JSON**

You can access the API JSON response of your indexing job.

1. Select **Download**.
2. Select **Insights (JSON)**. The JSON file opens in a new browser window or tab.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a video chat

Description automatically generated

A screenshot of a video chat

Description automatically generated

A person talking to a friend

Description automatically generated with medium confidence

A screenshot of a video chat

Description automatically generated

A screenshot of a video chat

Description automatically generated

A person with a cigarette in his mouth

Description automatically generated

A screenshot of a video chat

Description automatically generated

A person in a purple shirt

Description automatically generated