UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF TEXAS
WACO DIVISION

CELLULAR SOUTH, INC.	§	
Plaintiff,	§	
v.	§	**Civil Action No.** 6:24-cv-00245
GOOGLE, LLC	§	**JURY TRIAL DEMANDED**
Defendant.	§	

COMPLAINT FOR PATENT INFRINGEMENT DEMAND FOR JURY TRIAL

Plaintiff Cellular South Inc. ("CSI" or "Plaintiff") files this Complaint against defendant Google, LLC ("Google" or "Defendant") and alleges as follows:

NATURE OF THIS ACTION

1. This complaint alleges patent infringement. CSI alleges that Google has infringed and continues to infringe three patents: U.S. Patent Nos. 10,218,954 ("the '954 Patent"), 9,940,972 ("the '972 Patent"), and 11,126,853 ("the '853 Patent"). Copies of these patents (collectively, the "Patents-in-Suit") are attached hereto as **Exhibits A–C**.

2. The Patents-in-Suit cover foundational technologies for organizing unstructured data within video that allows content providers to gain a better understanding of the context and value of content and present it back to users in a highly personalized manner.

3. Google directly infringes the Patents-in-Suit by making, using, offering to sell, selling, and/or importing into the United States video classification and recognition technology, software, and services that practice the inventions claimed in the Patents-in-Suit. Google directs and controls each relevant aspect of the accused technology discussed herein, and benefits from the use of each feature that infringes the Patents-in-Suit.

4. CSI seeks damages and other relief for Google's infringement of the Patents-in-Suit.

THE PARTIES

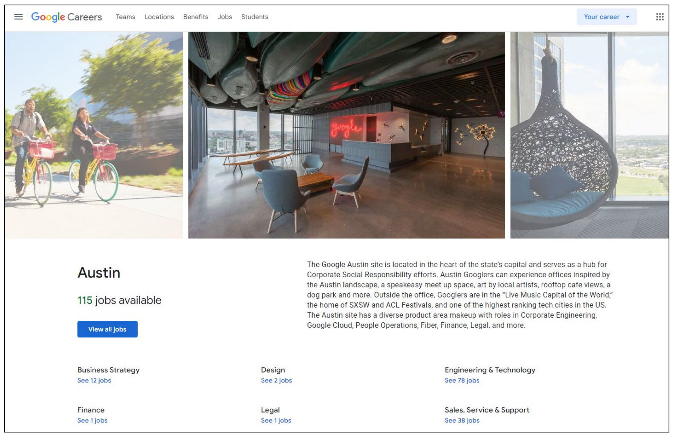
5. Plaintiff CSI is a corporation formed under the laws of Mississippi with a principal place of business at 1018 Highland Colony Parkway, Suite 300, Ridgeland, Mississippi 39157.

6. Defendant Google LLC is a corporation formed under the laws of Delaware with a principal place of business at 1600 Amphitheatre Parkway, Mountain View, California 94043. Google also has an office located at 500 West 2nd Street in Austin, Texas.

JURISDICTION AND VENUE

7. CSI brings this civil action for patent infringement under the Patent Laws of the United States, 35 U.S.C. § 1 *et. seq.*, including 35 U.S.C. §§ 271, 281-285. This Court has subject matter jurisdiction over this action under 28 U.S.C. §§ 1331 and 1338.

8. Venue is proper in this judicial district pursuant to 28 U.S.C. § 1400(b). Google maintains an established place of business in the state of Texas, and the Western District of Texas specifically, including its office at 500 West 2nd Street, Austin, Texas 78701.



Austin

115 jobs available

The Google Austin site is located in the heart of the state's capital and serves as a hub for Corporate Social Responsibility efforts. Austin Googlers can experience offices inspired by the Austin landscape, a speakeasy meet up space, art by local artists, rooftop cafe views, a dog park and more. Outside the office, Googlers are in the "Live Music Capital of the World," the home of SXSW and ACL Festivals, and one of the highest ranking tech cities in the US. The Austin site has a diverse product area makeup with roles in Corporate Engineering, Google Cloud, People Operations, Fiber, Finance, Legal, and more.

Business Strategy
See 12 jobs

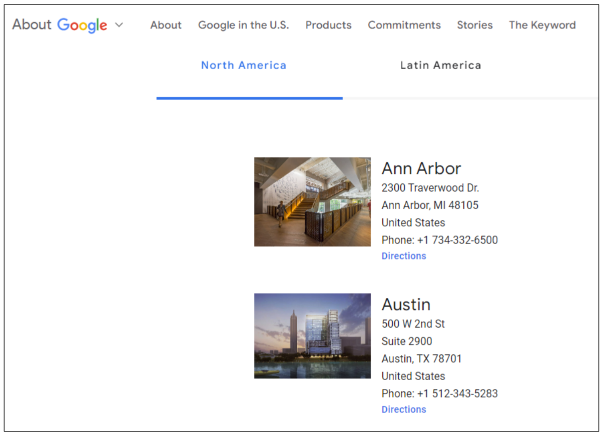
Design
See 2 jobs

Engineering & Technology
See 78 jobs

Finance
See 1 jobs

Legal
See 1 jobs

Sales, Service & Support
See 38 jobs



Ann Arbor
2300 Traverwood Dr.
Ann Arbor, MI 48105
United States
Phone: +1 734-332-6500
Directions

Austin
500 W 2nd St
Suite 2900
Austin, TX 78701
United States
Phone: +1 512-343-5283
Directions

Source:https://www.google.com/about/careers/applications/locations/austin/ (last accessed Mar. 1, 2024); https://about.google/locations/?region=north-america (last accessed Mar. 1, 2024).

Additionally, on information and belief, Google has committed acts of infringement in this judicial district, and has purposefully transacted business in this judicial district.

9. Google sells and offers for sale the accused technology in this state, including within the Western District of Texas. More specifically, Google offers the ability for anyone in the state to purchase its Cloud Video Intelligence and Video AI software platform that includes the accused technology. Additionally, upon information and belief, Google employs personnel who work on and use the accused technology in their Austin, Texas office. Google, therefore, obtains the benefits and protections of the laws of the State of Texas. This dispute arises out of and has a substantial connection with Google's contacts within this state and its infringement in this state, resulting in the exercise of jurisdiction being fair and reasonable.

10. Google is subject to this Court's specific and general personal jurisdiction pursuant to due process and/or the Texas Long Arm Statute because Google conducts substantial business in this forum, including: (i) making, using, selling, importing, and/or offering for sale the accused technology all throughout the District; and (ii) upon information and belief, employs engineers and other personnel who work on, sell, and use the accused technology.

FACTUAL BACKGROUND

11. CSI is a diversified telecommunications and technology company based in Ridgeland, Mississippi. CSI's mission is to "engage the exceptional and embrace operational excellence to best deliver connectivity and technology solutions that advance our communities and customers' lives."[1] CSI employs approximately 1,850 people who work and live in Mississippi, Alabama, and Tennessee. It has three lines of business—Wireless, Home Fiber and Business.

12. CSI is part of a family of operating companies owned by a privately held company, Telapex, Inc. Telapex, Inc. is owned by the Creekmore family, which has been in the telecommunications business in Mississippi since 1948.

13. Starting in the 1948, Wade Creekmore, Sr. began to acquire small rural telephone exchanges in areas of Mississippi which Southern Bell had no interest in serving. In the late 1950's, Mr. Creekmore divided the properties into two operating landline telephone companies—

[1] Cellular South Corporate Mission Statement. *See* https://www.cspire.com/web/company/about.

Franklin Telephone Company, Inc. and Delta Telephone Company, Inc.—and put his sons Wade, Jr. and Jimmy in charge of each, respectively. As incumbent telephone operators, Franklin and Delta were eligible to participate in the first lotteries conducted by the Federal Communications Commission, and in 1986, they acquired their first wireless licenses in the State of Mississippi. Two years later, CSI launched the first 1G wireless service on the Mississippi Gulf Coast. Within the next decade, CSI brought wireless telephone services to forty-two (42) Mississippi counties.

14. Beginning in the early 2000s, CSI expanded its footprint to include parts of Tennessee and northern Mississippi, upgraded to 4G service, and then to 5G. CSI remains a true regional wireless provider and remains committed to its community through the continued expansion of telecommunications technology and its investment into the education and development of the Mississippi community.

15. In addition to its wireless business, CSI through its subsidiaries owns or manages more than 20,000 miles of optical fiber networks primarily in the states of Mississippi and Alabama. CSI is also a leading value-added reseller for commercial hardware and software applications offered by major technology companies and solutions providers. Using its fiber and commercial technology solutions, CSI connects businesses and government entities with a suite of world-class IT solutions.

16. CSI announced its Fiber to the Home initiative in 2013, when there were only four cities in the United States where fiber to the home was available. In 2014, CSI launched its first market in the small Mississippi town of Quitman, which became one of the first communities in the United States to have buried fiber providing speeds of 1 Gigabit per second (Gig Internet Service) to the home. Today CSI has residential fiber customers in about 115 communities in Mississippi and Alabama, and it has recently developed new "fiberhoods" in Tennessee as well.

17. While CSI is a relatively small company compared to its competitors in the telecommunications industry, it has always been an innovator. Beginning with its founder's willingness to offer telephone services in rural areas where the Bell System would not, CSI continues to lead through its commitment to offer the best wireless service using the latest

technologies, as exemplified by its position as one of the first providers to offer Gig Internet Service over fiber to residential areas.

18. CSI has also worked to diversify its product offerings to remain innovative and competitive in the telecommunications and technology marketplace. Through one of these efforts, CSI created and offered "Video-to-Data" (or "V2D"), a cloud-based digital profiling and analytics system that could scan and turn video into searchable data faster than the real-time viewing of the video by using multiple servers that run simultaneously and in parallel. CSI offered V2D commercially through a wholly-owned subsidiary named Vū Digital, LLC ("Vu Digital" or "Vu").

19. V2D breaks down a video frame-by-frame and identifies the objects—for example, text, audio, images, faces, locations, logos—within each frame. V2D then creates a chronological record of all the objects identified within the video. V2D was a transformative product that made video content as easy to search as text, thereby providing users with unprecedented video classification and clustering capabilities, as well as significantly enhanced search engine indexing, content personalization, and targeted advertising capabilities. As Vu Digital's spokesperson explained: "We [Vu] offer the only single, comprehensive automated solution on the market today capable of organized the unstructured data within video," that will help content providers "unlock the power of video."[2]

20. CSI believed V2D could be an incredibly useful technology for video data processing and content personalization and set out to market it to a broad variety of potential customers. By at least 2016, Vu Digital outlined how V2D could be used by law enforcement to process body camera footage which was becoming more widely adopted around this same time. Vu Digital also marketed V2D to a variety of entertainment, technology, and sports vendors, many

[2] https://www.cspire.com/cms/news/wireless/25500006/V%C5%AB%20Digital%20provides%20PGA%20TOUR%20Entertainment%20Division%20with%20Video-to-Data%20(V2D)%20Analytics%20Services (last accessed Mar. 5, 2024).

of whom expressed interest in the product. For example, in or around 2016, Vu entered a multi-year agreement with the PGA Tour to help process the PGA's video and audio content.[3]

21. The V2D product debuted in May 2015 and mere months later was recognized by TCM—a global integrated media conglomerate and leading source of news and information for the communications and technology industries worldwide—as the "best-of-the-best technology solutions available on the market today."[4] V2D received TCM's coveted Communications Solutions 2015 Product of the Year Award that recognizes exceptional voice, data, and video communications products and services. Other 2015 recipients for this honor included prominent industry players such as AT&T, Alcatel-Lucent, Comcast, Dell, Hewlett Packard, and Logitech.[5]

22. Vu's V2D product also drew praise from Innovate Mississippi, a statewide group that helps innovation-based startup companies by connecting entrepreneurs with investors. As Innovate's then-President and CEO Tony Jeff explained, "It's not every day that a Mississippi company is mentioned in the same breath alongside some of the world's technology giants, but thanks to [CSI] and Vū Digital, it's becoming more common."[6]

23. In addition to the TMC award, Vu Digital was selected in 2015 to be part of the National Association of Broadcasters' ("NAB") year-long SPROCKIT accelerator program for innovative startups in media and entertainment.[7] SPROCKIT is a global innovation platform created to help large media, entertainment, and technology companies meet with emerging tech start-ups to fast-track investment, acquisition, and partnerships between the two groups on new

[3] *Id.*

[4] https://www.cspire.com/cms/news/wireless/24600006/V%C5%AB%20Digital%20Wins%20Coveted%20TMC%20Communications%20Solutions%202015%20Product%20of%20the%20Year%20Award (last accessed Mar. 5, 2024).

[5] *Id.*

[6] *Id.*

[7] *Id.*

products and services. On information and belief, Vu Digital's Wade Smith presented at the SPROCKIT NAB Show, which was held April 11-16, 2015 in Las Vegas.[8]

24. SPROCKIT Sync is invitation-only event put on three times a year by SPROCKIT that entails a series of private meetings between selected startups and corporate players in the media and entertainment industry. Following the launch of its V2D product, Vu Digital was invited to attend the SPROCKIT Sync conference that was held on June 18, 2015 at Google Tech Corners in Sunnyvale, California. Vu Digital's Project Specialist, Gregory Sandifier, presented its then "patent pending" V2D technology at the conference. On information and belief, when Vu Digital presented at this conference in 2015, the corporate partners participating included Disney, Fox, Samsung, and Google, among others. Vu Digital presented V2D to these corporate partners, including Google.

FIRST CLAIM FOR RELIEF

(Infringement of U.S. Patent. No. 10,218,954)

25. CSI realleges and incorporates by reference the allegations of the foregoing paragraphs.

26. On February 26, 2019, the United States Patent and Trademark Office ("USPTO") duly and legally issued, after a full and fair examination, United States Patent No. 10,218,954 (the "'954 Patent") entitled "Video to Data" to inventors Naeem Lakhani, Bartlett Wade Smith, IV, and Allison A. Talley. A true and correct copy of the '954 patent is attached as **Exhibit A** to this Complaint.

27. The '954 Patent was assigned to CSI, which currently holds all substantial rights, title, and interest in and to the '954 Patent.

28. Pursuant to 35 U.S.C. § 282, the '954 Patent is presumed valid.

29. The '954 Patent is presumed to be patent eligible under 35 U.S.C. § 101.

[8] https://www.businesswire.com/news/home/20150402006447/en/NAB-Show-Unveils-Final-10-Participants-Selected-for-SPROCKIT-2015 (last accessed Mar. 5, 2024).

30. The ’954 Patent is directed to an improvement in the functionality of machine learned video recognition and classification systems, particularly with regard to specific techniques for improving the accuracy of predictions made using image, audio, text, and other video data. More specifically, the ’954 Patent is directed to techniques that utilize contextual information such as the arrangement of certain objects in a series of still images in order to improve computer visual recognition in a video classification system.

31. The ’954 Patent addresses specific technological challenges that arose in video classification systems when attempting to extract meaningful metadata from video content.

32. The ’954 Patent describes the challenges in the field of computer visual recognition systems and explains the advantages of the claimed inventions:

> Identifying various objects in an image can be a difficult task. For example, locating (segmenting) and positively identifying an object in a given frame or image can yield false positives-locating but wrongfully identifying an object. Therefore, present embodiments can be utilized to eliminate false positives, for example, by using context. As one example, if the audio soundtrack of a video is an announcer calling a football game, then identification of ball in a given frame as basketball can be assigned a reduced probability or weighting. As another example of using context, if a given series of image frames from a video is positively or strongly identified as a horse race, then identifying an object to be a mule or donkey can be given a reduced weight.[9]

33. The ’954 Patent provides additional examples illustrating the improvements that the described methodologies provide to computer visual recognition systems:

> In certain instances, identification of an individual can be a difficult task. For example, facial recognition can become difficult when an individual’s face is obstructed by another object like a football, a baseball helmet, a musical instrument, or other obstructions. An advantage of some embodiments described herein can include the ability to identify an individual without identification of the individual’s face. Embodiments can use contextual information such as association of objects, text, and/or other context within an

[9] ’954 Patent, 5:3–16.

> image or video. As one example, a football player scores a touchdown but rather than identifying the player using facial recognition, the play can be identified by object recognition of, for example, the player's team's logo, text recognition of the player's jersey number, and by cross referencing this data with that team's roster (as oppose to another team, which is an example of why the logo recognition can be important). Such embodiments can further learn to identify that player more readily and save his image as data.[10]

34. The '954 Patent explains that the described methodologies for deriving contextual information can be helpful in identifying and correcting or eliminating false positives.[11] Moreover, these methodologies may make it possible to identify unknown objects in a given image by narrowing a large number of practically infinite possibilities to a relatively small number of object possibilities, thereby allowing positive object recognition where identification previously could not be achieved.[12]

35. As the examiner acknowledged during prosecution, none of the prior art cited during examination disclosed the claimed techniques which recite using audio and video semantic information to generate "contextual topics" and "generating a contextual text, an image, or an animation" based on the determined contextual topics.

36. The technological solutions described above are recited in the '954 Patent claims, including, for example, independent claims 1 and 13 (and their corresponding dependent claims).

37. Claim 1 of the '954 Patent reads as follows:

> 1. A method to generate video data from a video comprising:
>
> generating audio files and image files from the video;
>
> distributing the audio files and the image files across a plurality of processors and processing the audio files and the image files in parallel;

[10] *Id*. at 5:35–52.

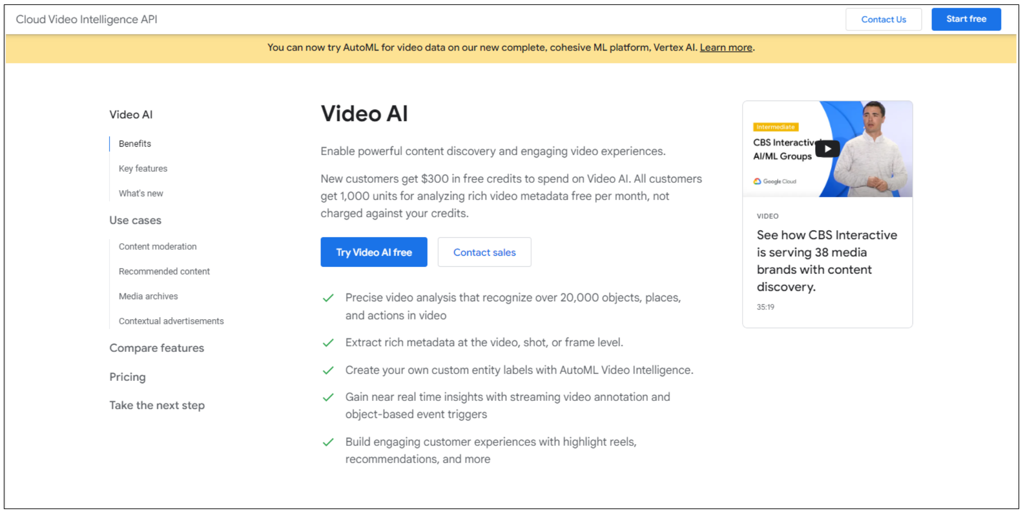
[11] *Id*. at 5:53–63.

[12] *Id*. at 5:64–6:14.

> converting audio files associated with the video to text;
>
> identifying an object in the image files;
>
> determining a contextual topic from the image files;
>
> assigning a probability of accuracy to the identified object based on the contextual topic;
>
> converting the image files associated with the video to video data, wherein the video data comprises the object, the probability, and the contextual topic;
>
> cross-referencing the text and the video data with the video to determine contextual topics;
>
> generating a contextual text, an image, or an animation based on the determined contextual topics;
>
> generating a content-rich video based on the generated text, image, or animation.

38. Upon information and belief, Google directly infringes at least claim 1, at least in the exemplary manner described below.

39. Google directly infringes the ’954 Patent by making, using, offering to sell, and/or selling in the United States its Cloud Video Intelligence platform (“Video Intelligence” or the “Accused Product”), which relies on machine learning (“ML”) models and natural language processing (“NLP”) techniques to identify and classify videos using contextual information. Google directs and controls each relevant aspect of the accused technology discussed herein, and benefits from the use of each feature that infringes the ’954 Patent. On information and belief Google uses the ’954 patented inventions and the Accused Product to classify videos uploaded to its online video sharing platform, YouTube (www.youtube.com) and to its Google Photos iOS and Android apps.



You can now try AutoML for video data on our new complete, cohesive ML platform, Vertex AI. Learn more.

Video AI

Enable powerful content discovery and engaging video experiences.

New customers get $300 in free credits to spend on Video AI. All customers get 1,000 units for analyzing rich video metadata free per month, not charged against your credits.

- ✓ Precise video analysis that recognize over 20,000 objects, places, and actions in video
- ✓ Extract rich metadata at the video, shot, or frame level.
- ✓ Create your own custom entity labels with AutoML Video Intelligence.
- ✓ Gain near real time insights with streaming video annotation and object-based event triggers
- ✓ Build engaging customer experiences with highlight reels, recommendations, and more

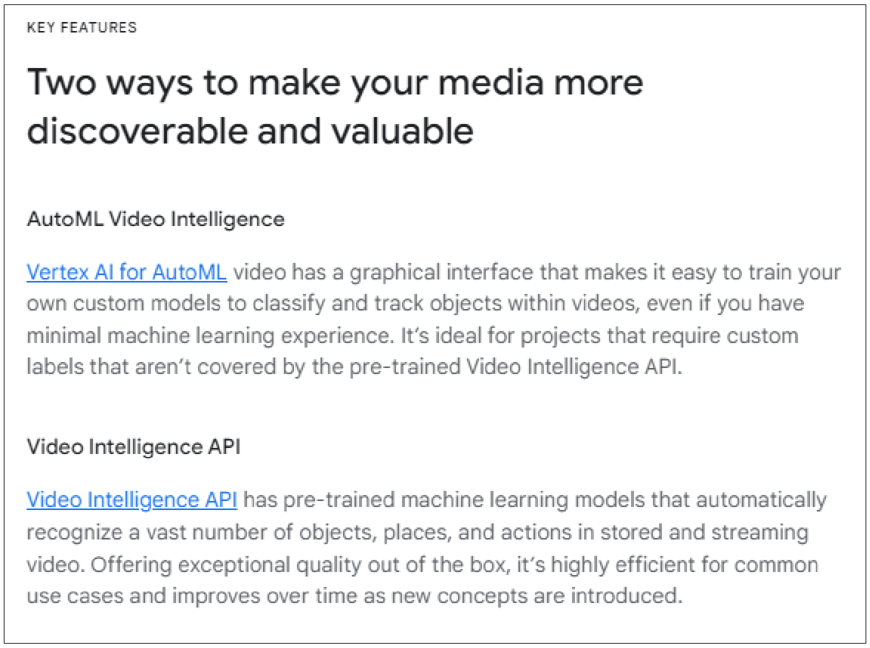
See how CBS Interactive is serving 38 media brands with content discovery.

Source: https://cloud.google.com/video-intelligence

40. According to Google, Video Intelligence may be used with custom (via Vertex AI for AutoML)[13] or pretrained ("Video Intelligence API") machine learning models in order to classify and annotate videos so that video data may be more easily parsed or indexed and to enable users to implement, for example, contextual-based advertising and/or content recommendation for their video content libraries.[14]

[13] On information and belief, as of January 23, 2024, Google no longer supports use of the "AutoML Video Intelligence" product to allow customers to classify video content using custom ML models. Instead, AutoML has been replaced and/or consolidated with Google's "Vertex AI" product which continues to provide this capability to customers. *See, e.g.*, https://cloud.google.com/video-intelligence/automl/pricing (last accessed March 4, 2024). As Google explains "[a]ll of the functionality of legacy AutoML Video Intelligence and new features are available on the Vertex AI platform." *Id.*

[14] *See* https://cloud.google.com/video-intelligence (last accessed Mar. 5, 2024).



KEY FEATURES

Two ways to make your media more discoverable and valuable

AutoML Video Intelligence

Vertex AI for AutoML video has a graphical interface that makes it easy to train your own custom models to classify and track objects within videos, even if you have minimal machine learning experience. It's ideal for projects that require custom labels that aren't covered by the pre-trained Video Intelligence API.

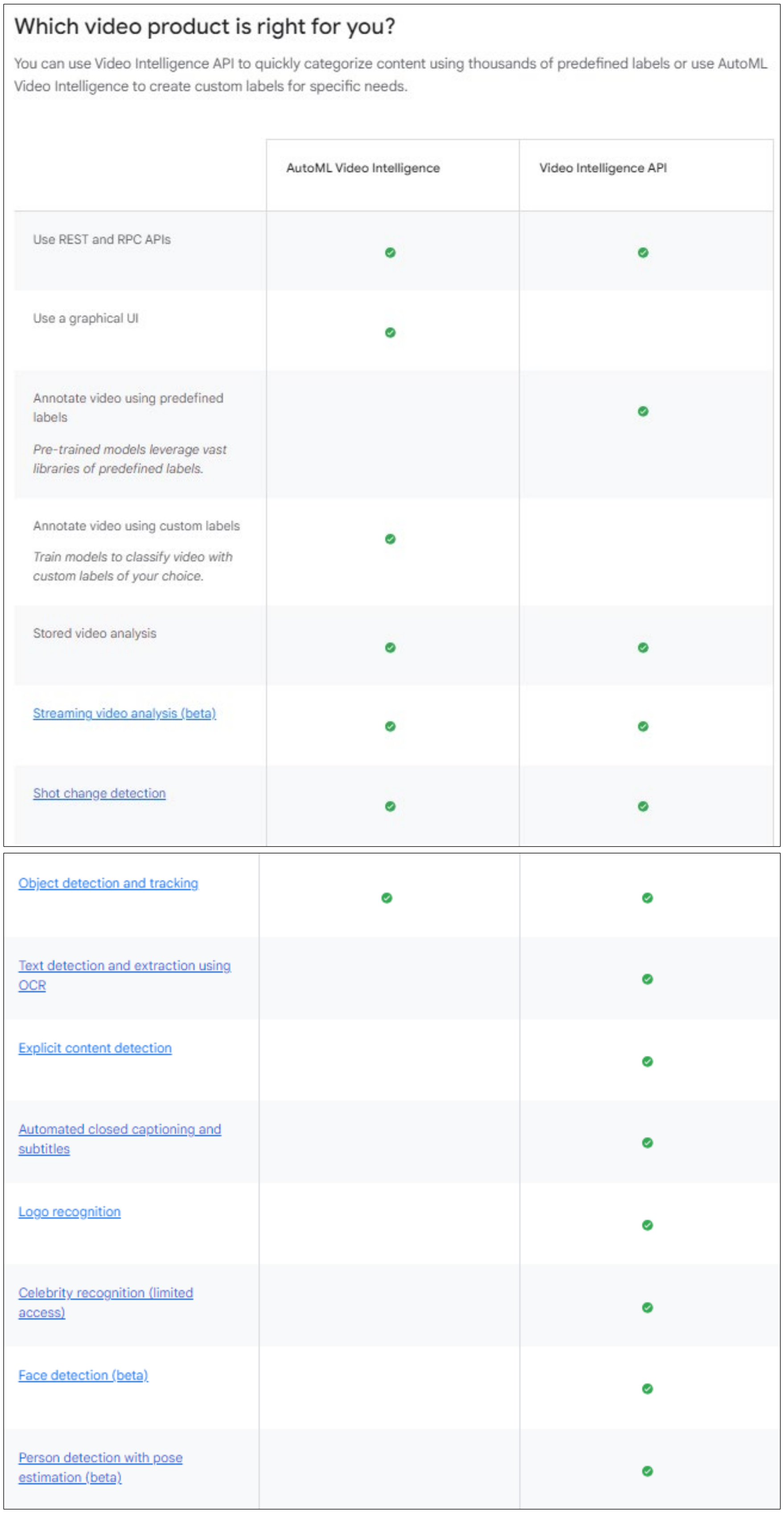
Video Intelligence API

Video Intelligence API has pre-trained machine learning models that automatically recognize a vast number of objects, places, and actions in stored and streaming video. Offering exceptional quality out of the box, it's highly efficient for common use cases and improves over time as new concepts are introduced.

Source: https://cloud.google.com/video-intelligence (last accessed Mar. 5, 2024)

41. As Google explains, Video Intelligence may be used to add custom or predefined labels for both stored video content library as well as for live-streaming video applications.[15]

[15] *Id.*

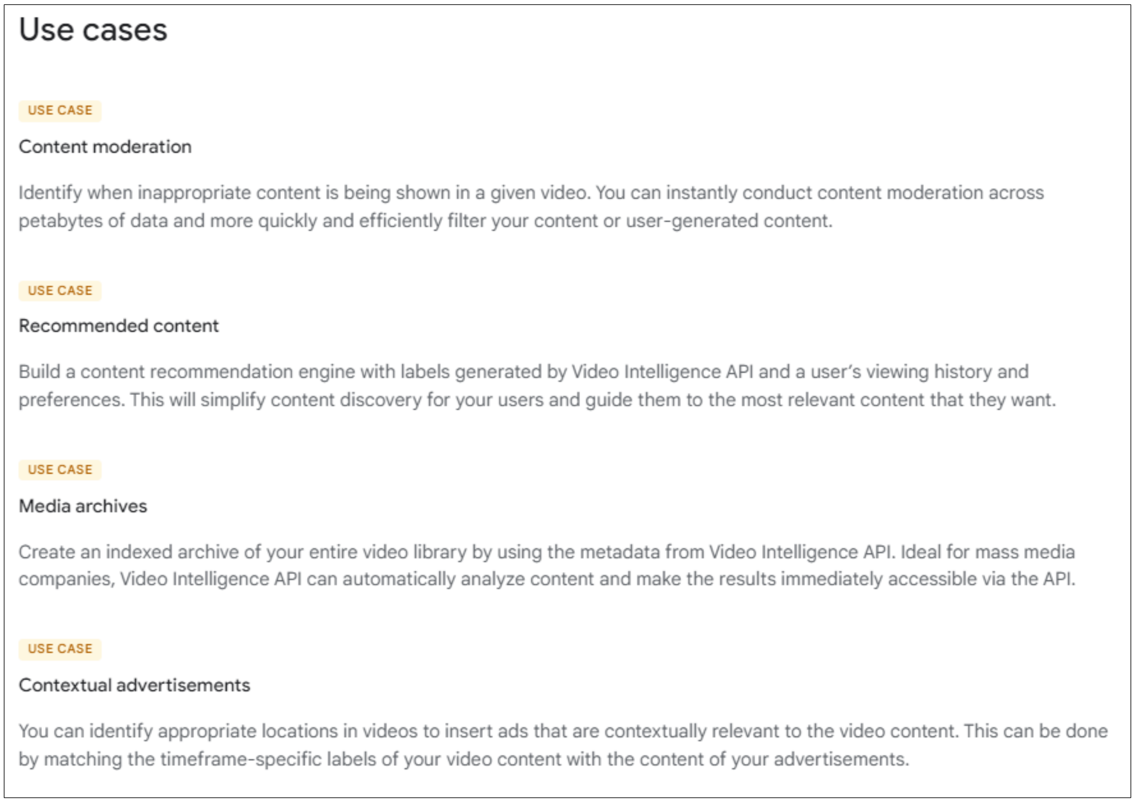


Which video product is right for you?

You can use Video Intelligence API to quickly categorize content using thousands of predefined labels or use AutoML Video Intelligence to create custom labels for specific needs.

	AutoML Video Intelligence	Video Intelligence API
Use REST and RPC APIs	✓	✓
Use a graphical UI	✓	
Annotate video using predefined labels *Pre-trained models leverage vast libraries of predefined labels.*		✓
Annotate video using custom labels *Train models to classify video with custom labels of your choice.*	✓	
Stored video analysis	✓	✓
Streaming video analysis (beta)	✓	✓
Shot change detection	✓	✓
Object detection and tracking	✓	✓
Text detection and extraction using OCR		✓
Explicit content detection		✓
Automated closed captioning and subtitles		✓
Logo recognition		✓
Celebrity recognition (limited access)		✓
Face detection (beta)		✓
Person detection with pose estimation (beta)		✓

Source: https://cloud.google.com/video-intelligence



Use cases

USE CASE

Content moderation

Identify when inappropriate content is being shown in a given video. You can instantly conduct content moderation across petabytes of data and more quickly and efficiently filter your content or user-generated content.

USE CASE

Recommended content

Build a content recommendation engine with labels generated by Video Intelligence API and a user's viewing history and preferences. This will simplify content discovery for your users and guide them to the most relevant content that they want.

USE CASE

Media archives

Create an indexed archive of your entire video library by using the metadata from Video Intelligence API. Ideal for mass media companies, Video Intelligence API can automatically analyze content and make the results immediately accessible via the API.

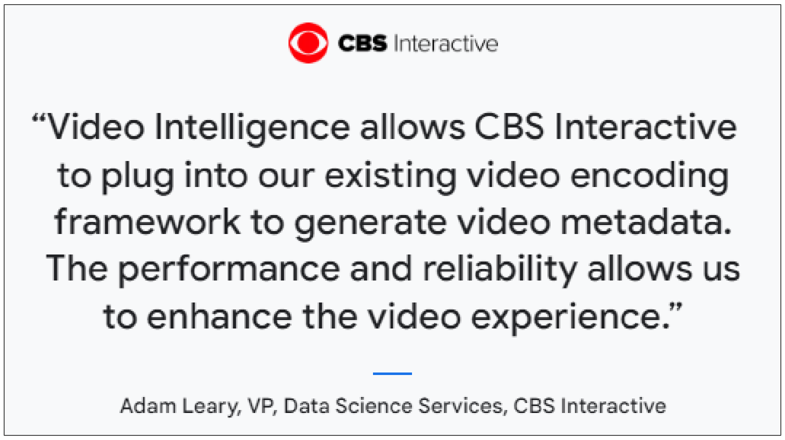
USE CASE

Contextual advertisements

You can identify appropriate locations in videos to insert ads that are contextually relevant to the video content. This can be done by matching the timeframe-specific labels of your video content with the content of your advertisements.

Source: https://cloud.google.com/video-intelligence

42. On information and belief, Google's customers use Video Intelligence to generate video metadata using the context-based techniques of the '954 Patent in order to "enhance the video experience."



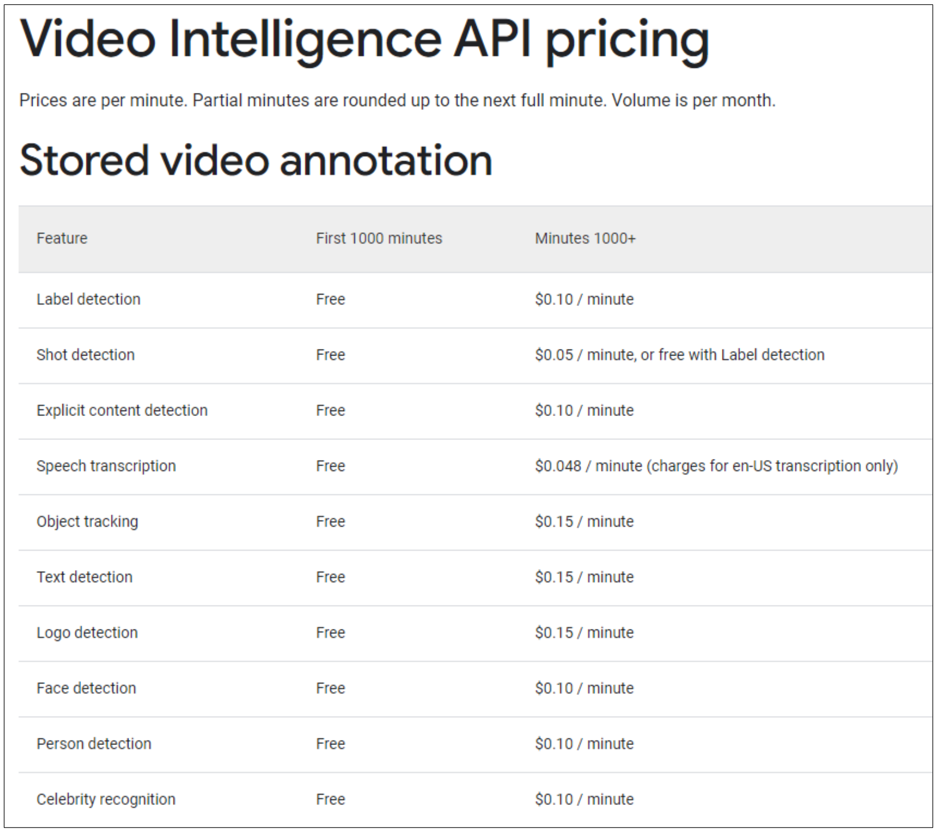
CBS Interactive

"Video Intelligence allows CBS Interactive to plug into our existing video encoding framework to generate video metadata. The performance and reliability allows us to enhance the video experience."

Adam Leary, VP, Data Science Services, CBS Interactive

Source : https://cloud.google.com/video-intelligence

43. Google provides the Video Intelligence product to users through its Google Cloud Platform ("GCP") and charges users for the amount of resources (*e.g.*, cloud server virtual CPUs) consumed for video processing on a per minute basis.



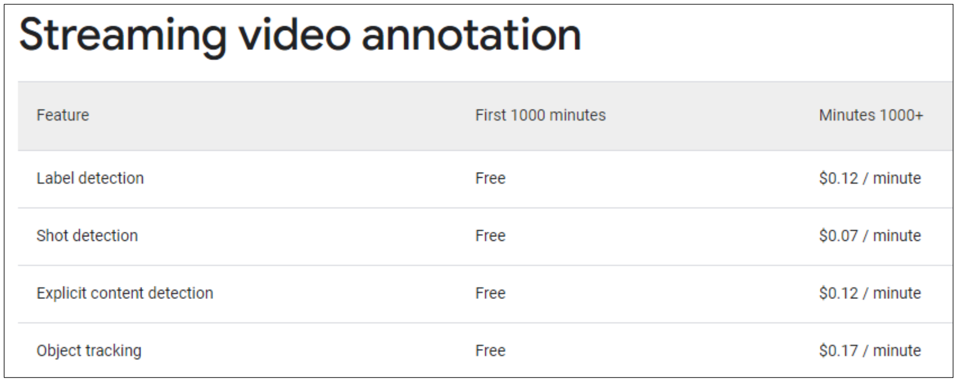
Video Intelligence API pricing

Prices are per minute. Partial minutes are rounded up to the next full minute. Volume is per month.

Stored video annotation

Feature	First 1000 minutes	Minutes 1000+
Label detection	Free	$0.10 / minute
Shot detection	Free	$0.05 / minute, or free with Label detection
Explicit content detection	Free	$0.10 / minute
Speech transcription	Free	$0.048 / minute (charges for en-US transcription only)
Object tracking	Free	$0.15 / minute
Text detection	Free	$0.15 / minute
Logo detection	Free	$0.15 / minute
Face detection	Free	$0.10 / minute
Person detection	Free	$0.10 / minute
Celebrity recognition	Free	$0.10 / minute

Source: https://cloud.google.com/video-intelligence/pricing



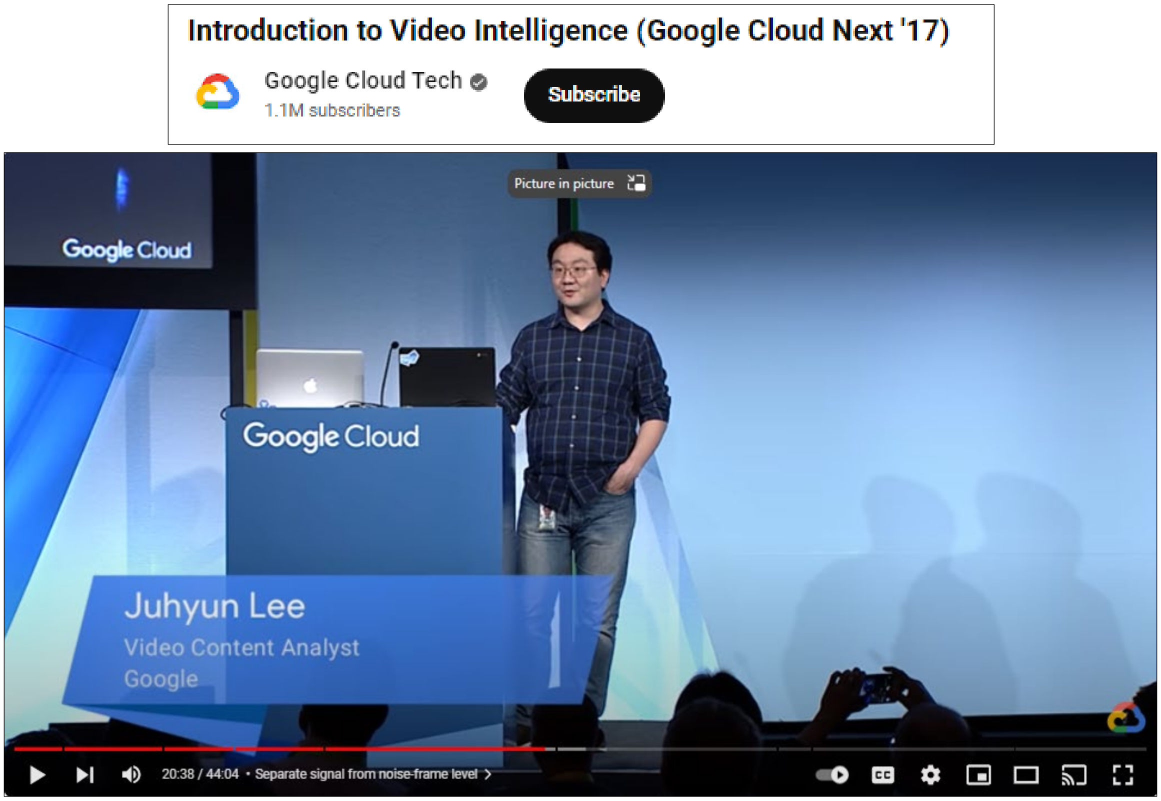
Streaming video annotation

Feature	First 1000 minutes	Minutes 1000+
Label detection	Free	$0.12 / minute
Shot detection	Free	$0.07 / minute
Explicit content detection	Free	$0.12 / minute
Object tracking	Free	$0.17 / minute

Source: https://cloud.google.com/video-intelligence/pricing#streaming_video_annotation

44. With regard to claim 1 of the ’954 Patent, Google practices “[a] method to generate video data from a video comprising: generating audio files and image files from the video.” For

example, at the Google Next 2017 developer conference, Google explained that Video Intelligence begins by decoding video into video data, audio data, subtitle data, and various other metadata.[16]



Introduction to Video Intelligence (Google Cloud Next '17)

Source: https://www.youtube.com/watch?v=y-k8oelbmGc

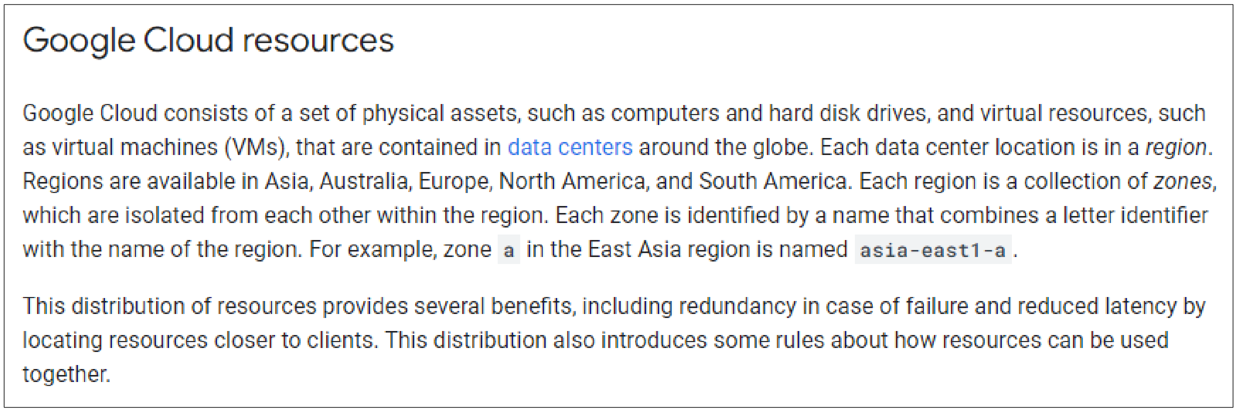
> So given a video in your GCS [Google Cloud Storage] bucket, **the first thing that we do is we decode the video**. And usually a video container contains like **audio streams, video streams, subtitles and all of the metadata**.[17]

45. The Accused Product performs the step of "distributing the audio files and the image files across a plurality of processors and processing the audio and image files in parallel." On information and belief, the Video Intelligence product executes on a distributed network of one or more Google Cloud servers. As Google explains, Video Intelligence uses the Google Cloud

[16] Introduction to Video Intelligence (Google Cloud Next '17) (available at https://www.youtube.com/watch?v=y-k8oelbmGc) (last accessed Mar. 5, 2024).

[17] *Id*. at 21:00 to 22:14 (emphasis added).

Platform ("GCP"), which comprises a number of computer resources housed in Google's data centers around the world.[18]

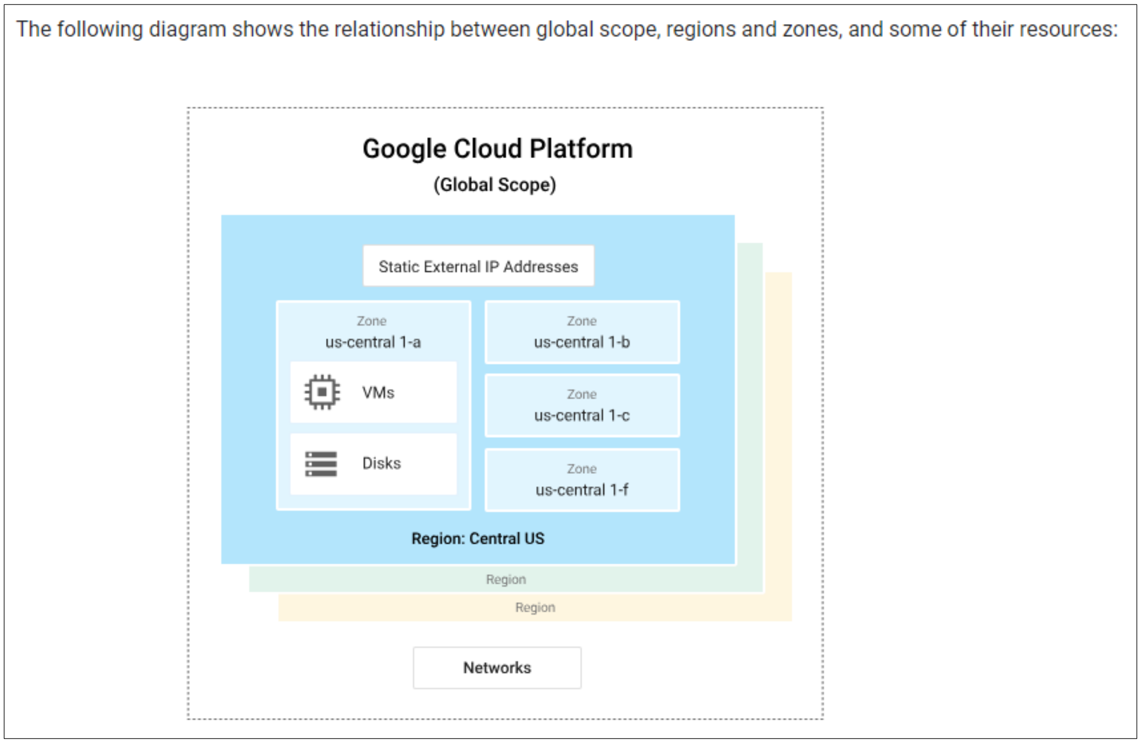


Google Cloud resources

Google Cloud consists of a set of physical assets, such as computers and hard disk drives, and virtual resources, such as virtual machines (VMs), that are contained in data centers around the globe. Each data center location is in a *region*. Regions are available in Asia, Australia, Europe, North America, and South America. Each region is a collection of *zones*, which are isolated from each other within the region. Each zone is identified by a name that combines a letter identifier with the name of the region. For example, zone `a` in the East Asia region is named `asia-east1-a`.

This distribution of resources provides several benefits, including redundancy in case of failure and reduced latency by locating resources closer to clients. This distribution also introduces some rules about how resources can be used together.

Source: https://cloud.google.com/docs/overview

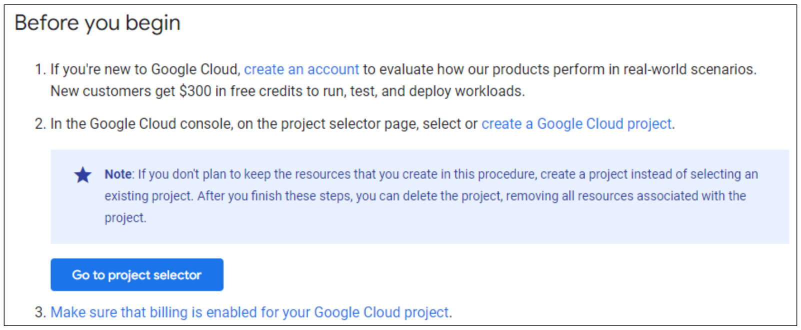


The following diagram shows the relationship between global scope, regions and zones, and some of their resources:

Source: https://cloud.google.com/docs/overview

46. In addition, Google explains that a GCP account and project are required in order to use the Video Intelligence API.

[18] *See* https://cloud.google.com/docs/overview.



Before you begin

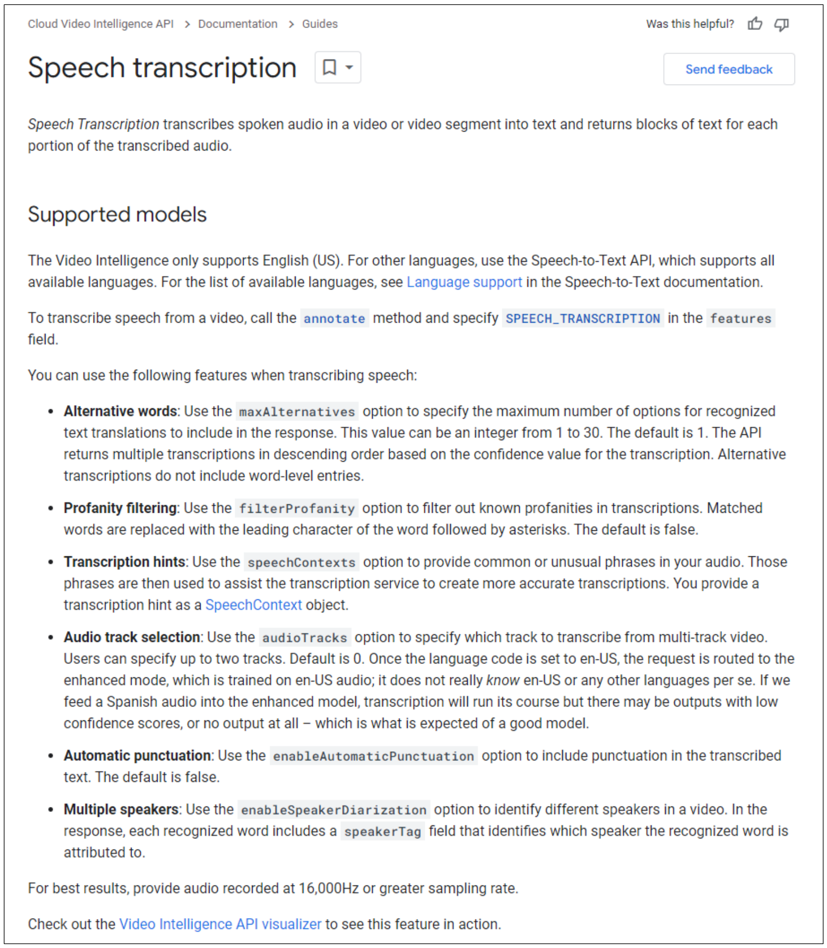
1. If you're new to Google Cloud, create an account to evaluate how our products perform in real-world scenarios. New customers get $300 in free credits to run, test, and deploy workloads.
2. In the Google Cloud console, on the project selector page, select or create a Google Cloud project.

★ **Note**: If you don't plan to keep the resources that you create in this procedure, create a project instead of selecting an existing project. After you finish these steps, you can delete the project, removing all resources associated with the project.

3. Make sure that billing is enabled for your Google Cloud project.

Source: https://cloud.google.com/video-intelligence/docs/annotate-video-command-line

47. The Accused Product also practices the step of "converting audio files associated with the video to text." For example, Google explains that Video Intelligence extracts audio data and converts that data into text using the "SPEECH_TRANSCRIPTION" feature.[19]



Speech transcription

Speech Transcription transcribes spoken audio in a video or video segment into text and returns blocks of text for each portion of the transcribed audio.

Supported models

The Video Intelligence only supports English (US). For other languages, use the Speech-to-Text API, which supports all available languages. For the list of available languages, see Language support in the Speech-to-Text documentation.

To transcribe speech from a video, call the `annotate` method and specify `SPEECH_TRANSCRIPTION` in the `features` field.

You can use the following features when transcribing speech:

- **Alternative words**: Use the `maxAlternatives` option to specify the maximum number of options for recognized text translations to include in the response. This value can be an integer from 1 to 30. The default is 1. The API returns multiple transcriptions in descending order based on the confidence value for the transcription. Alternative transcriptions do not include word-level entries.
- **Profanity filtering**: Use the `filterProfanity` option to filter out known profanities in transcriptions. Matched words are replaced with the leading character of the word followed by asterisks. The default is false.
- **Transcription hints**: Use the `speechContexts` option to provide common or unusual phrases in your audio. Those phrases are then used to assist the transcription service to create more accurate transcriptions. You provide a transcription hint as a SpeechContext object.
- **Audio track selection**: Use the `audioTracks` option to specify which track to transcribe from multi-track video. Users can specify up to two tracks. Default is 0. Once the language code is set to en-US, the request is routed to the enhanced mode, which is trained on en-US audio; it does not really *know* en-US or any other languages per se. If we feed a Spanish audio into the enhanced model, transcription will run its course but there may be outputs with low confidence scores, or no output at all – which is what is expected of a good model.
- **Automatic punctuation**: Use the `enableAutomaticPunctuation` option to include punctuation in the transcribed text. The default is false.
- **Multiple speakers**: Use the `enableSpeakerDiarization` option to identify different speakers in a video. In the response, each recognized word includes a `speakerTag` field that identifies which speaker the recognized word is attributed to.

For best results, provide audio recorded at 16,000Hz or greater sampling rate.

Check out the Video Intelligence API visualizer to see this feature in action.

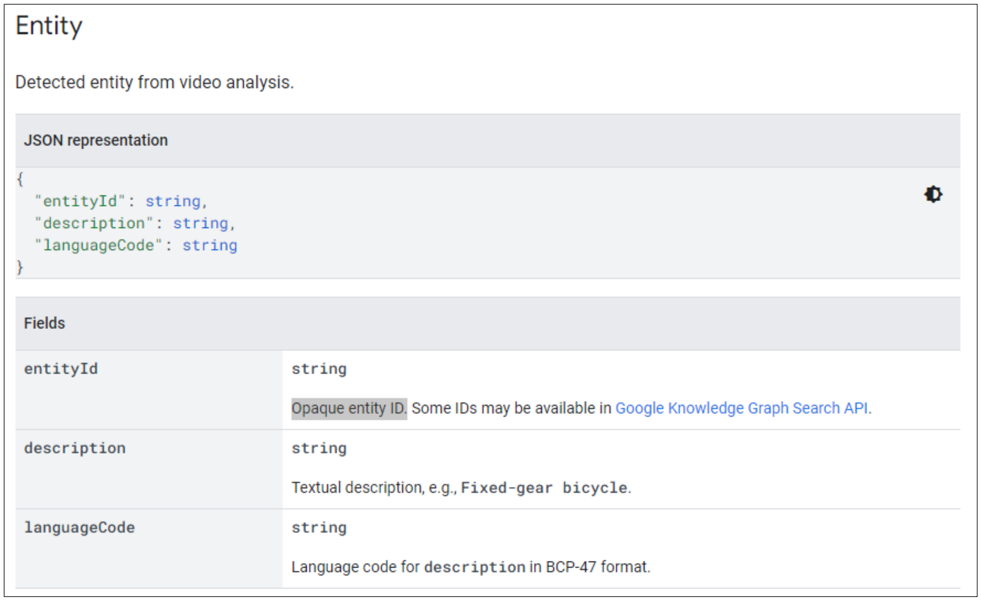
Source: https://cloud.google.com/video-intelligence/docs/feature-speech-transcription

[19] *See* https://cloud.google.com/video-intelligence/docs/feature-speech-transcription.

48. As Google explains, using audio data is important and necessary to ensure the accuracy of the Video Intelligence system:

> Assume that you have a video of an animated music video. It would be sad if we give you a label saying that it's an animated cartoon. **You definitely want your audio signal in there. So we are looking at various audio signals to improve the quality of the model.**[20]

49. The Accused Product also performs the step of "identifying an object in the image files." For example, Video Intelligence identifies and creates annotations for "entities" within a video, video segment, and/or frame.



Entity

Detected entity from video analysis.

JSON representation

```
{
  "entityId": string,
  "description": string,
  "languageCode": string
}
```

Fields

entityId	string Opaque entity ID. Some IDs may be available in Google Knowledge Graph Search API.
description	string Textual description, e.g., Fixed-gear bicycle.
languageCode	string Language code for description in BCP-47 format.

Source: https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse

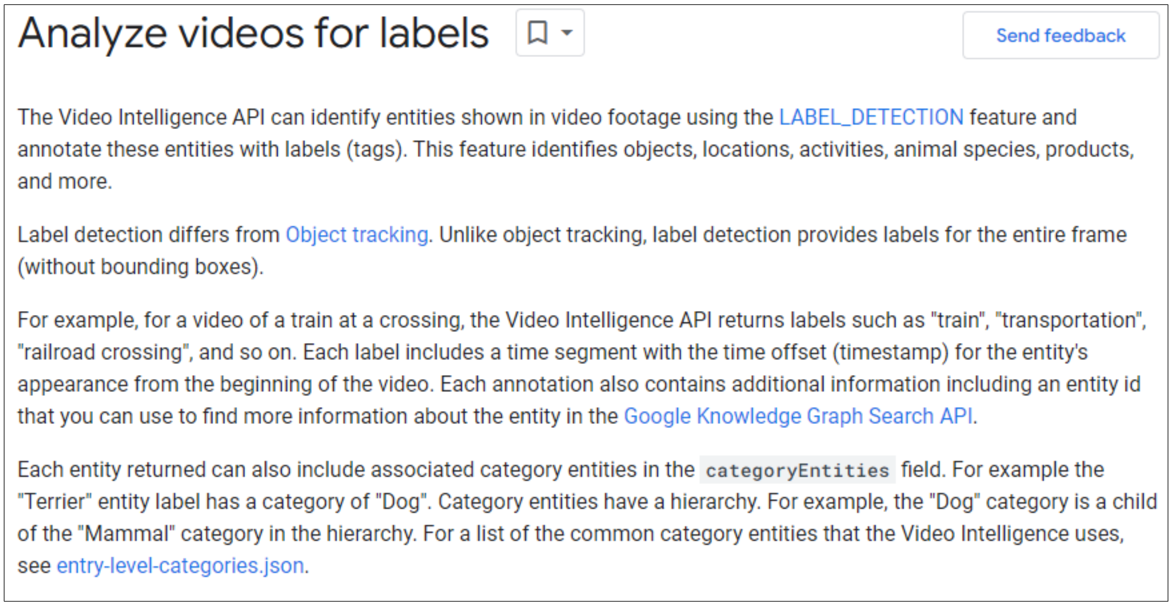
50. As Google explains, Video Intelligence first splits a video into a series of images and then applies an "image classifier" to identify objects within each of the images.

> And a video stream is often a sequence of images . . . a lot of them. For a five-minute video at 24 frames per second, you will have like 7,000, 8,000 images in them. And they will probably in sequence look a lot similar. So classifying each of those is not too efficient, let's say. So we shot sample the images, and say at one FPS. **So one frame per second, we apply an image classifier. An image classifier that is very similar to -- that you can get through the Google Cloud Vision API. And we do it for every image in the entire video. And for each frame, you will get image features such as this elephant with 97% accuracy, or**

[20] https://www.youtube.com/watch?v=y-k8oelbmGc at 23:44 to 24:07 (emphasis added).

> **animal, or river in that case. And, if you keep going on, you will get other labels for crocodiles and lions**.[21]

51. The Accused Product further performs the step of "determining a contextual topic from the image files." For instance, Video Intelligence may apply tags and/or "labels" to detected entities shown in an image extracted from a video by using the "*LABEL_DETECTION*" feature.



Analyze videos for labels

The Video Intelligence API can identify entities shown in video footage using the LABEL_DETECTION feature and annotate these entities with labels (tags). This feature identifies objects, locations, activities, animal species, products, and more.

Label detection differs from Object tracking. Unlike object tracking, label detection provides labels for the entire frame (without bounding boxes).

For example, for a video of a train at a crossing, the Video Intelligence API returns labels such as "train", "transportation", "railroad crossing", and so on. Each label includes a time segment with the time offset (timestamp) for the entity's appearance from the beginning of the video. Each annotation also contains additional information including an entity id that you can use to find more information about the entity in the Google Knowledge Graph Search API.

Each entity returned can also include associated category entities in the `categoryEntities` field. For example the "Terrier" entity label has a category of "Dog". Category entities have a hierarchy. For example, the "Dog" category is a child of the "Mammal" category in the hierarchy. For a list of the common category entities that the Video Intelligence uses, see entry-level-categories.json.

Source: https://cloud.google.com/video-intelligence/docs/feature-label-detection

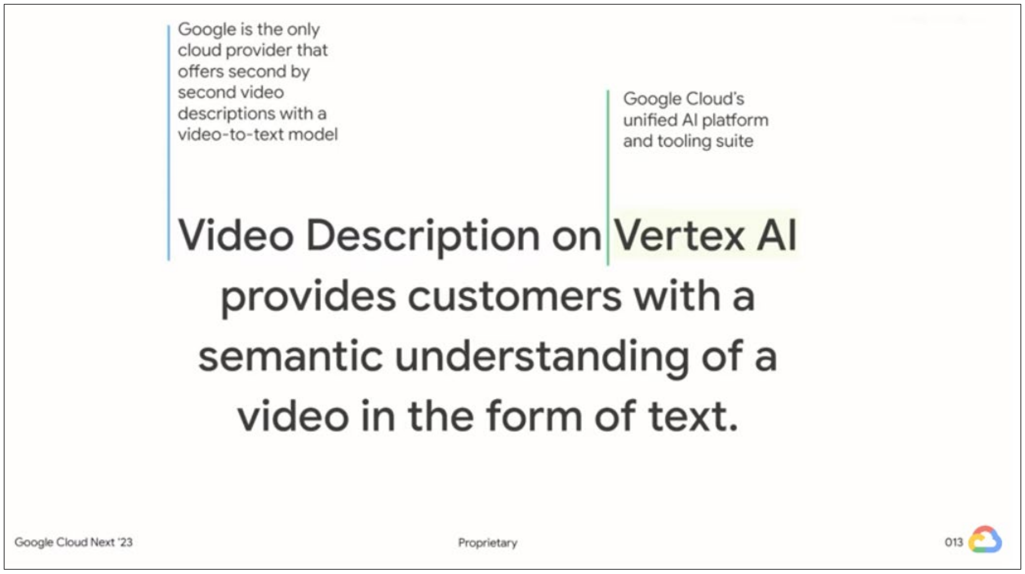
52. In addition, Google explains that Video Intelligence in Vertex AI includes a Video Description (video-to-text) model that allows users generate a semantic understanding of video in the form of text.

> So Video Description in Vertex AI is a feature with which customers can get **semantic understanding of a video in the form of text**.[22]

[21] *See* https://www.youtube.com/watch?v=y-k8oelbmGc at 21:00 to 22:14 (emphasis added).

[22] *See* Google developer presentation titled, "Unlock New Video And Image Use Cases With Imagen On Vertex AI" (available at https://www.youtube.com/watch?v=F4YKREYPkEI) at 7:17 to 7:24 (emphasis added).

Unlock new video and image use cases with Imagen on Vertex AI



Google is the only cloud provider that offers second by second video descriptions with a video-to-text model

Google Cloud's unified AI platform and tooling suite

Video Description on Vertex AI provides customers with a semantic understanding of a video in the form of text.

Source: https://www.youtube.com/watch?v=F4YKREYPkEI at 7:19.

53. On information and belief, Google and its customers use Video Intelligence via Vertex AI to analyze videos and generate contextual summaries for every 15-second video chunk.

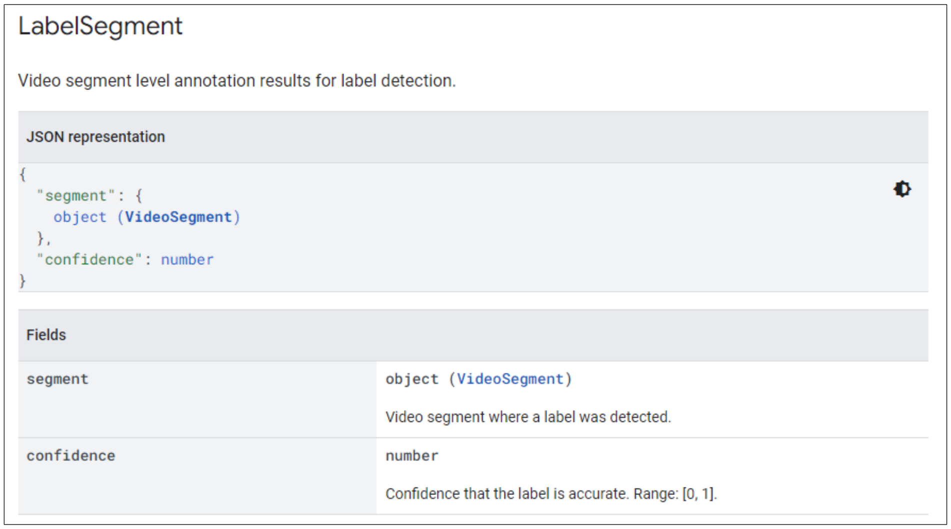
> Let's see a quick demo again. So similarly, same studio, now we have a tab called Video Description. You go ahead and select your video. Again, for a 2-minute video, it's about 20 to 30 seconds latency. And there you go. **For every 15-second chunk, it gives you a description. And in the UI experience, you can actually click on the timestamp and go to that portion of the video so you understand the relevancy and you maintain context.** You can save this as a JSON file. And by the way, all the three features I talked about today all have API support right out the box.[23]

54. The Accused Product performs the step of "assigning a probability of accuracy to the identified object based on the contextual topic." For instance, Google states the following with regard to Video Intelligence:

> And for each frame, you will get image features such as this elephant with 97% accuracy, or animal, or river in that case. And, if you keep going on, you will get other labels for crocodiles and lions.[24]

[23] *Id*. at 11:32 to 12:16 (emphasis added).
[24] *See* https://www.youtube.com/watch?v=y-k8oelbmGc at 21:00 to 22:14.

55. Google's developer documentation further explains that Video Intelligence assigns a confidence (accuracy) rating to the assigned entities, labels and/or tags, and describes various methods or API calls by which that metadata may be retrieved. For instance, Google explains that the confidence rating for a video segment level annotation may be retrieved using the "annotate" method together with the "LabelSegment" call.[25]



LabelSegment

Video segment level annotation results for label detection.

JSON representation

```
{
  "segment": {
    object (VideoSegment)
  },
  "confidence": number
}
```

Fields	
segment	object (VideoSegment) Video segment where a label was detected.
confidence	number Confidence that the label is accurate. Range: [0, 1].

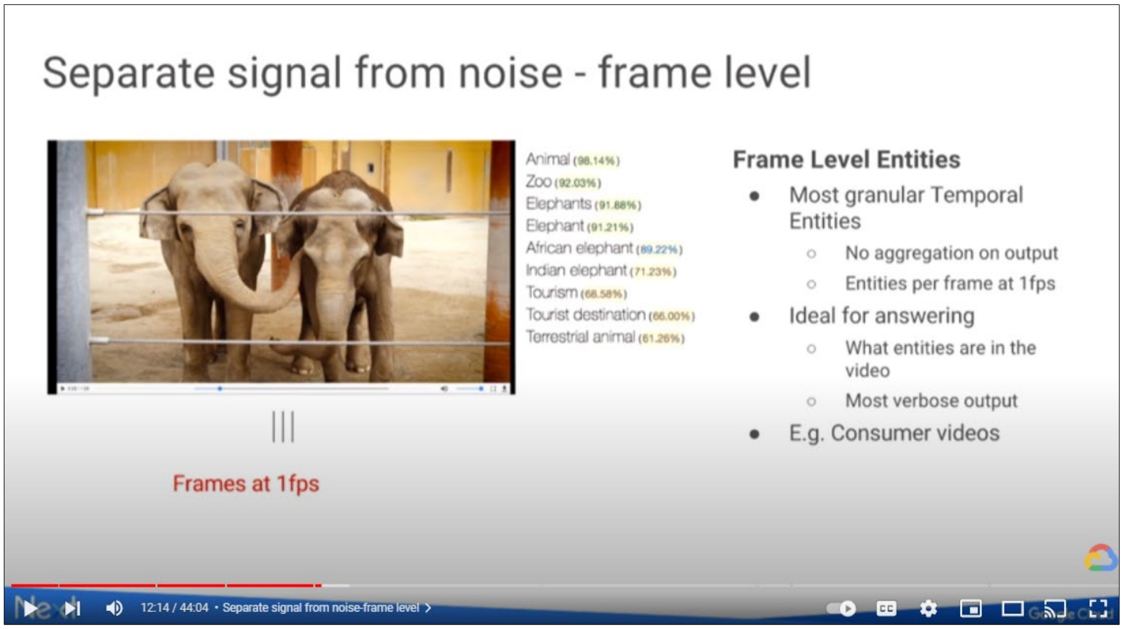
Source: https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse

56. The Accused Product performs the step of "converting the image files associated with the video to video data, wherein the video data comprises the object, the probability, and the contextual topic." For example, as described above, Video Intelligence detects entities within a frame, shot, and/or video segment, calculates the confidence value that the detected entity is correct, and assigns a label and/or contextual description for the analyzed frame, shot, or video segment.

> And we are first only interested in the video stream. And a video stream is often a sequence of images . . . a lot of them. For a five-minute video at 24 frames per second, you will have like 7,000, 8,000 images in them. And they will probably in sequence look a lot similar. So classifying each of those is not too efficient, let's say. So we shot sample the images, and say at one FPS. **So one frame per second, we apply an image classifier. An image classifier that is very similar to -- that**

[25] *See* https://cloud.google.com/video-intelligence/docs/reference/rest/v1/videos/annotate; *see also*, https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse#labelsegment.

> **you can get through the Google Cloud Vision API. And we do it for every image in the entire video. And for each frame, you will get image features such as this elephant with 97% accuracy, or animal, or river in that case. And, if you keep going on, you will get other labels for crocodiles and lions.**[26]



Separate signal from noise - frame level

Animal (98.14%)
Zoo (92.03%)
Elephants (91.88%)
Elephant (91.21%)
African elephant (89.22%)
Indian elephant (71.23%)
Tourism (68.58%)
Tourist destination (66.00%)
Terrestrial animal (61.26%)

Frames at 1fps

Frame Level Entities

- Most granular Temporal Entities
  - No aggregation on output
  - Entities per frame at 1fps
- Ideal for answering
  - What entities are in the video
  - Most verbose output
- E.g. Consumer videos

Source: https://www.youtube.com/watch?v=y-k8oelbmGc at 12:14.

57. As Google explains, this metadata may be generated using the "LABEL_DETECTION" video annotation feature.[27] As discussed above, once generated, this metadata about the contextual topic, object, and probability may be retrieved using various API calls such as the "LabelSegment" or "LabelFrame" elements together with the "annotate" method.[28]

58. The Accused Product further performs the step of "cross-referencing the text and the video data with the video to determine contextual topics." As Google explains, Video

[26] *See* https://www.youtube.com/watch?v=y-k8oelbmGc at 21:00 to 22:14 (emphasis added).

[27] *See* https://cloud.google.com/video-intelligence/docs/reference/rest/v1p3beta1/videos/annotate #request-body.

[28] *See* https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse#labelsegment; https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse#labelframe.

Intelligence utilizes an "aggregating video classifier" that combines the outputs of the analyzed video image data and audio, text, subtitle or other data to determine a contextual topic.

> And at the end of the video, what we do is **we have an aggregating video classifier, and this is really what sits at the core of the -- of this particular API**. **And this – given all of these features throughout the entire video it will then say, hey, this is a documentary about animals in Africa. Another similar example would be if you have a video of people in costumes or pumpkins and candies, this classifier will be able to tell you, hey, this is a video about Halloween.** And this is only possible because we have trained this model on millions of YouTube videos, and human-rated and also like-verified.[29]

59. On information and belief, Video Intelligence via Vertex AI for AutoML also cross-references text and video data with the video to determine contextual topics. For example, Google explains that Vertex AI models process raw video data and associate video data with natural language to allow for semantic video searching.

> Traditionally, video search only allows searching against title, description, and possibly tags. But these descriptions usually typically provide the video uploaders with some poor quality that may not be able to capture the full details of the video. With model pipeline, we can achieve [a] much better experience. So two images are shown. It depicts the content creator Alice uploading videos to a video platform. **And after the raw video is processed by the model pipeline, a user, Bob, can type in the natural language into the search bar for semantic video search.** This could allow Bob to accurately search interesting moments. Inside the long video, a massive video library is automatically—it'll locate the replay position inside these typically long videos.[30]

60. Google explains further that cross-referencing video data with textual data in Video Intelligence via Vertex AI can be used to power better content recommendations, search capabilities, and other use cases.

> There are some other great use cases video description helps unlock. So one of them is, of course, metadata. **Now that you have such detailed information about the video, of course, you have richer metadata, which you can use to power better recommendation engines or searches.** Another one is automated captions, similar to the one we saw earlier with images. So if you have short form

[29] *See* https://www.youtube.com/watch?v=y-k8oelbmGc at 22:14 to 23:11 (emphasis added).

[30] *See* https://www.youtube.com/watch?v=F4YKREYPkEI at 26:44 to 27:30 (emphasis added).

> videos, like 15 seconds, 30 seconds, you can caption those. Accessibility—you say the tourism app example earlier.
>
> **And the next one is understanding and searching the important moments**. So this is a really cool one. So for an entire video, if you had this detailed rendering in the form of text, now you could do something with it. **So now that I have the text format, I could search for important terms in it**. So for example, let's imagine we have a six-hour baseball game video, right? And we have its rendering. Now, I could search for terms like home run, strike, audience cheering. And then once I have those moments, I can stitch them together into, say, a highlight reel.[31]

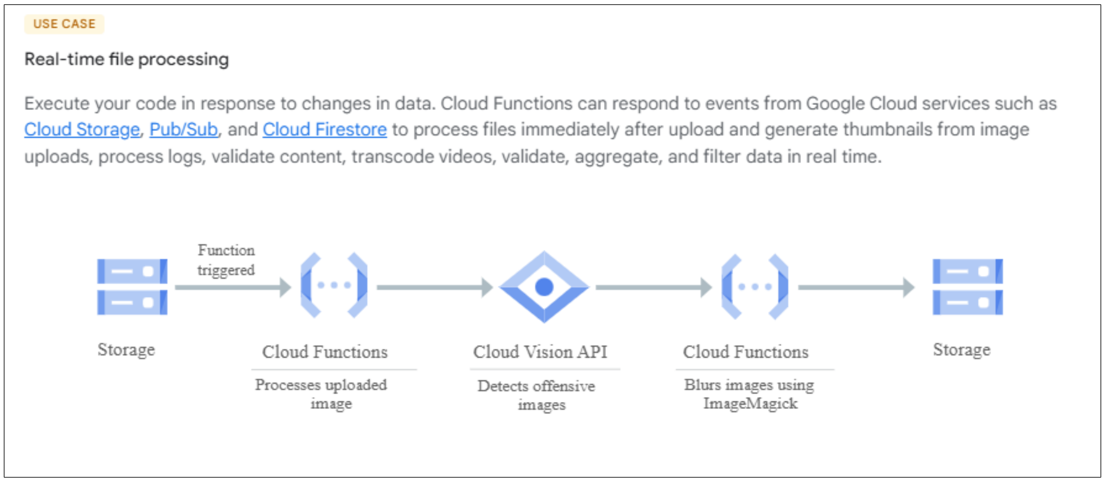
61. The Accused Product performs the step of "generating a contextual text, an image, or an animation based on the determined contextual topics" and "generating a content-rich video based on the generated text, image, or animation." For instance, Google explains that media publishers can use the results output by Video Intelligence to generate a highlight reel of video content related to a particular topic or to rapidly create video descriptions for searching.

> A media publisher could have hundreds of petabytes of video data sitting in storage buckets and one common thing they might want to do is create a highlight reel focused on a specific type of content or search their large library for a specific entity. So let's see how we would use the Video Intelligence API to search a large library of videos, given all this metadata that we get back from it. So we've got a lot of videos here and lets say this media publisher has hours of sports video but they only want to find the content relevant to baseball. So let's go ahead and search our library here for baseball videos. And we can see that not only does it show us which videos have baseball, it tells us exactly when in those videos baseball appears.[32]

62. Google also explains that Video Intelligence may be used in conjunction with Google Cloud Functions ("GCF") to automatically generate thumbnail images to display for a given video on a media content page.

[31] *Id*. at 8:35 to 9:40 (emphasis added).

[32] Google Cloud Tech demonstration titled, "Cloud Video Intelligence API Demo," (available at https://www.youtube.com/watch?v=mDAoLO4G4CQ) at 1:16 to 1:57.



USE CASE

Real-time file processing

Execute your code in response to changes in data. Cloud Functions can respond to events from Google Cloud services such as Cloud Storage, Pub/Sub, and Cloud Firestore to process files immediately after upload and generate thumbnails from image uploads, process logs, validate content, transcode videos, validate, aggregate, and filter data in real time.

Source: https://cloud.google.com/functions/#video-and-image-analysis

63. Similarly, Video Intelligence (via Vertex AI with AutoML) may be used to create text descriptions of videos, highlight reels, and captions, among other applications.

> There are some other great use cases video description helps unlock. So one of them is, of course, metadata. Now that you have such detailed information about the video, of course, you have richer metadata, which you can use to power better recommendation engines or searches. **Another one is automated captions**, similar to the one we saw earlier with images. **So if you have short form videos, like 15 seconds, 30 seconds, you can caption those**. Accessibility—you say the tourism app example earlier.
>
> And the next one is understanding and searching the important moments. So this is a really cool one. So for an entire video, if you had this detailed rendering in the form of text, now you could do something with it. So now that I have the text format, I could search for important terms in it. So for example, let's imagine we have a six-hour baseball game video, right? And we have its rendering. **Now, I could search for terms like home run, strike, audience cheering. And then once I have those moments, I can stitch them together into, say, a highlight reel**.[33]

64. Upon information and belief, Google directly infringes other claims of the ’954 Patent as well, including for the reasons discussed in the preceding paragraphs.

65. At no time has Google been licensed, either expressly or impliedly, under the ’954 Patent.

[33] https://www.youtube.com/watch?v=F4YKREYPkEI at 8:35 to 9:40 (emphasis added).

66. On information and belief, Google has had knowledge of the ’954 Patent at least through its participation in the SPROCKIT event in or around April or May 2015 at which Vu Digital presented its then “patent pending” V2D technology. On information and belief, Google additionally has had knowledge of the ’954 Patent family since at least as early as November 2, 2022, as evidenced by Google’s identification of the ’954 Patent as potential prior art to the inventions described in Google’s U.S. Patent Application No. 17/423,623.[34] At least as of the filing of this Complaint, Google has had knowledge of the ’954 Patent and Google’s infringement thereof.

67. Google’s infringement of the ’954 Patent, which is knowing and willful at least as of the filing of this Complaint, has caused and continues to cause damage to CSI, and CSI is entitled to recover damages sustained as a result of Google’s wrongful acts in an amount subject to proof at trial.

<u>SECOND CLAIM FOR RELIEF</u>

(Infringement of U.S. Patent. No. 9,940,972)

68. CSI realleges and incorporates by reference the allegations of the foregoing paragraphs set forth above.

69. On April 10, 2018, the USPTO duly and legally issued, after a full and fair examination, U.S. Patent No. 9,940,972 (“the ’972 patent”) titled “Video to data” to inventors Naeem Lakhani and Bartlett Wade Smith, IV. A true and correct copy of the ’972 patent is attached as **Exhibit B** to this Complaint.

70. The ’972 Patent was assigned to CSI, which currently holds all substantial rights, title, and interest in and to the ’972 Patent.

71. Pursuant to 35 U.S.C. § 282, the ’972 Patent is presumed valid.

72. The ’972 Patent is presumed to be patent eligible under 35 U.S.C. § 101.

[34] *See* Google’s November 2, 2022 Information Disclosure Statement with regard to U.S. Patent Application No. 17,423,623 (“**Exhibit D**”).

73. The ’972 Patent is directed to an improvement in the functionality of machine learned video recognition and classification systems, particularly with regard to specific techniques for improving the accuracy of predictions made using image, audio, textual, and other video data. More specifically, the ’972 Patent is directed to techniques that utilize topical information such as the arrangement of certain objects in a series of still images in order to improve computer visual recognition in a video classification system.

74. The ’972 Patent addresses specific technical challenges that arose in video classification systems when attempting to extract meaningful metadata from video content.

75. The ’972 Patent describes the challenges in the field of computer visual recognition systems and explains the advantages of the claimed inventions. For instance, the ’972 Patent explains that using the claimed techniques, “the topics generated from an image or a frame and the topics extracted from audio can be combined.”[35] In this way, “[t]he text can be cross-referenced, and topics common to both texts would be given additional weight.”[36] The ’972 Patent explains that “key words indicating topic and semantic that appear in both [image and audio] texts can be selected or emphasized,” thereby permitting a more holistic and complete summary when generating, for example, text describing the content of the video.[37]

76. As the examiner acknowledged during prosecution, none of the prior art cited during examination disclosed the claimed techniques which recite cross-referencing audio and video semantic information based on “topical metadata” to generate “topics,” “generating video text” based on the cross-referencing, and “generating text, an image, or an animation” based on the generated video text and placing the text, image, or animation in the video.

77. The technological solutions described above are recited in the ’972 Patent claims, including, for example, independent claims 1 and 17 (and their corresponding dependent claims).

[35] ’972 Patent, 5:46–47.

[36] ’972 Patent, 5:47–49.

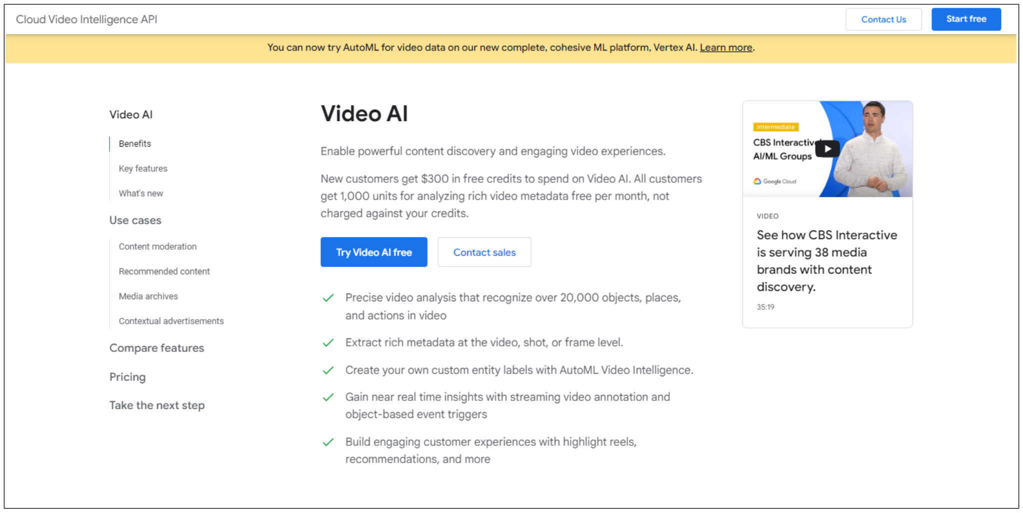
[37] ’972 Patent, 5:49–54.

78. Claim 1 of the ’972 Patent reads as follows:

> 1. A method to generate video data from a video comprising:
>
> generating audio files and image files from the video;
>
> distributing the image files across a plurality of processors and processing the image files in parallel, wherein processing the image files comprises extracting one or more objects and identifying the one or more objects;
>
> processing the audio files;
>
> converting audio files associated with the video to text;
>
> converting the image files associated with the video to video data;
>
> generating a topical meta-data that describes content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio files;
>
> adding the topical meta-data to the video; and
>
> cross-referencing the text and the video data based on the generated topical meta-data to determine topics;
>
> generating video text based on the cross-referencing, wherein the video text describes content of the video;
>
> generating a text, image, or animation based on the video text; and
>
> placing the text, image, or animation in the video.

79. Upon information and belief, Google directly infringes at least claim 1, at least in the exemplary manner described below.

80. Google directly infringes the ’972 Patent by making, using, offering to sell, and/or selling in the United States its Cloud Video Intelligence platform, which relies on ML models and NLP techniques to identify and classify videos using contextual information. Google directs and controls each relevant aspect of the accused technology discussed herein, and benefits from the use of each feature that infringes the ’972 Patent. On information and belief Google uses the ’972 patented inventions and the Accused Product to classify videos uploaded to its online video sharing platform, YouTube (www.youtube.com), and to its Google Photos iOS and Android apps.



You can now try AutoML for video data on our new complete, cohesive ML platform, Vertex AI. Learn more.

Video AI

Enable powerful content discovery and engaging video experiences.

New customers get $300 in free credits to spend on Video AI. All customers get 1,000 units for analyzing rich video metadata free per month, not charged against your credits.

- ✓ Precise video analysis that recognize over 20,000 objects, places, and actions in video
- ✓ Extract rich metadata at the video, shot, or frame level.
- ✓ Create your own custom entity labels with AutoML Video Intelligence.
- ✓ Gain near real time insights with streaming video annotation and object-based event triggers
- ✓ Build engaging customer experiences with highlight reels, recommendations, and more

VIDEO

See how CBS Interactive is serving 38 media brands with content discovery.

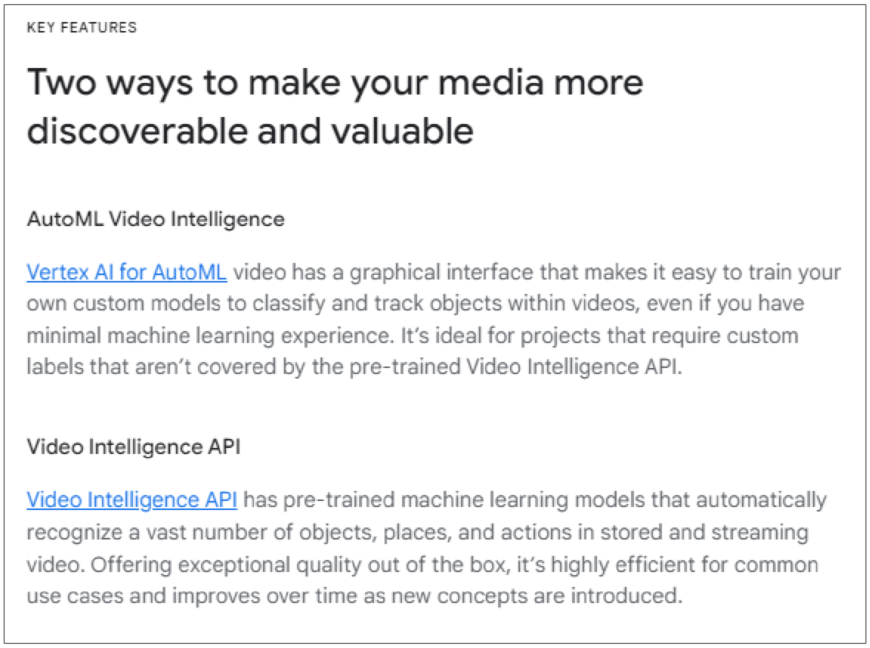
35:19

Source: https://cloud.google.com/video-intelligence

81. According to Google, Video Intelligence may be used with custom (via "Vertex AI for AutoML")[38] or pretrained ("Video Intelligence API") machine learning ("ML") models in order to classify and annotate videos so that video data may be more easily parsed and to enable users to implement, for example, contextual-based advertising and/or content recommendation in their video content libraries.[39]

[38] As discussed above, AutoML has been replaced and/or consolidated with Google's "Vertex AI" product, which continues to provide this capability to customers. *See*, *e.g.*, https://cloud.google.com/video-intelligence/automl/pricing (last accessed March 4, 2024). As Google explains "[a]ll of the functionality of legacy AutoML Video Intelligence and new features are available on the Vertex AI platform." *Id*.

[39] *See* https://cloud.google.com/video-intelligence.



KEY FEATURES

Two ways to make your media more discoverable and valuable

AutoML Video Intelligence

Vertex AI for AutoML video has a graphical interface that makes it easy to train your own custom models to classify and track objects within videos, even if you have minimal machine learning experience. It's ideal for projects that require custom labels that aren't covered by the pre-trained Video Intelligence API.

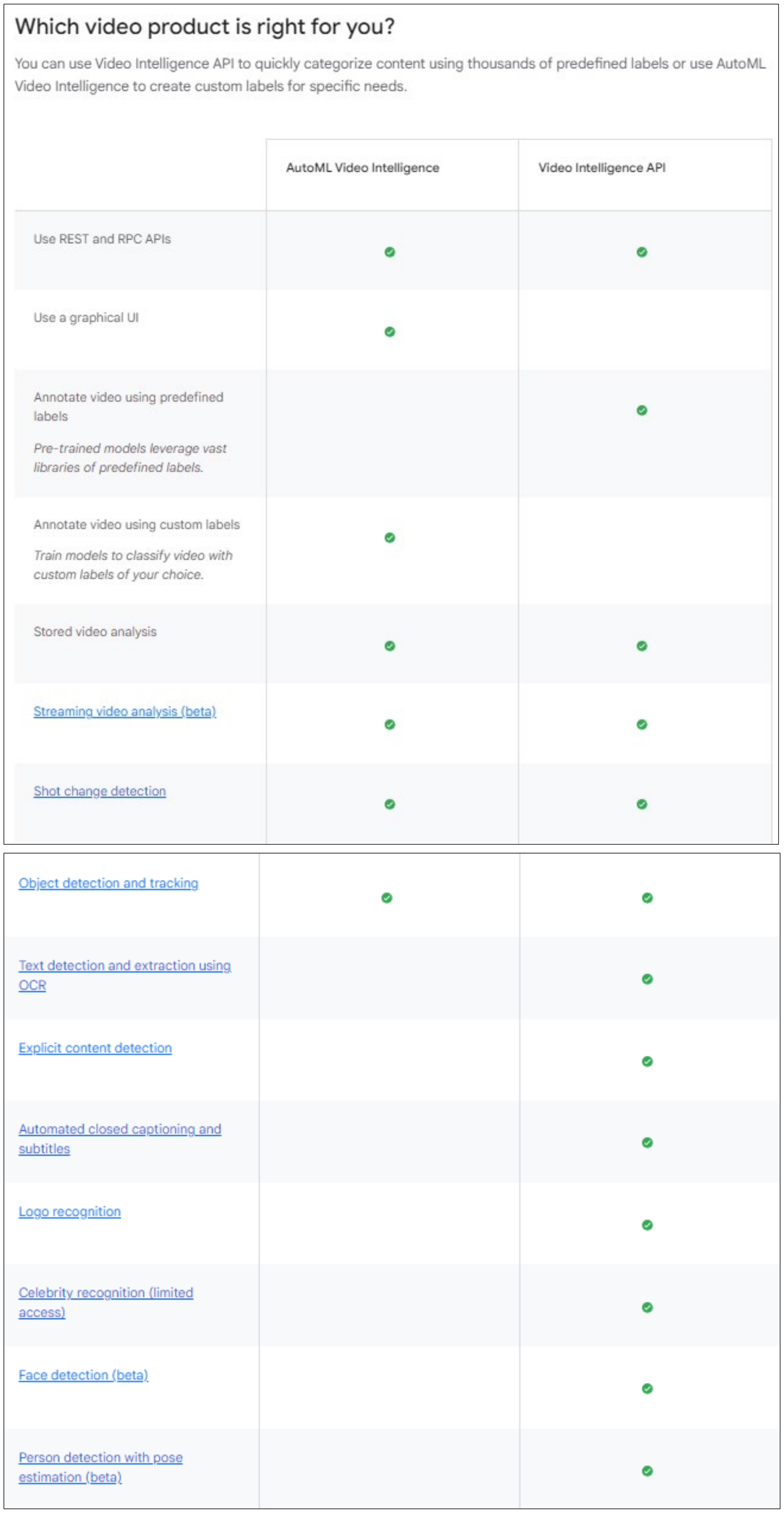
Video Intelligence API

Video Intelligence API has pre-trained machine learning models that automatically recognize a vast number of objects, places, and actions in stored and streaming video. Offering exceptional quality out of the box, it's highly efficient for common use cases and improves over time as new concepts are introduced.

Source: https://cloud.google.com/video-intelligence

82. Video Intelligence may be used to add custom or predefined labels for both stored video content libraries as well as for live-streaming video applications.[40]

[40] *Id.*

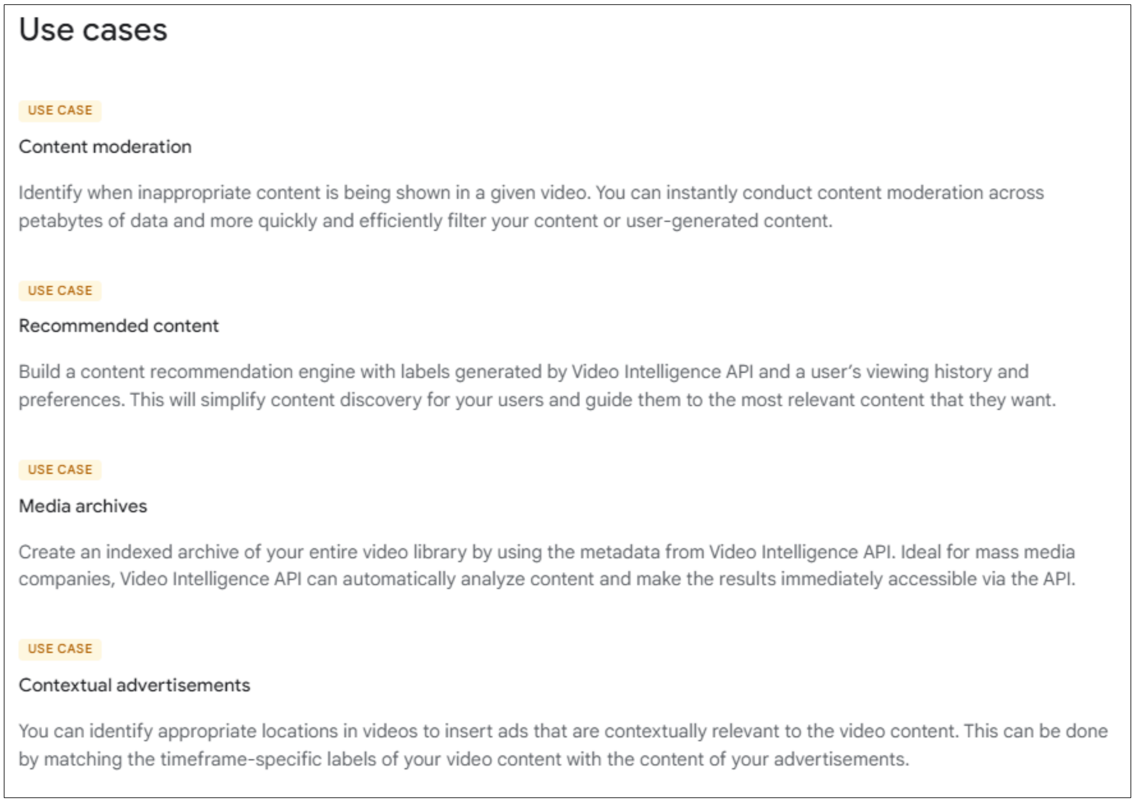


Which video product is right for you?

You can use Video Intelligence API to quickly categorize content using thousands of predefined labels or use AutoML Video Intelligence to create custom labels for specific needs.

	AutoML Video Intelligence	Video Intelligence API
Use REST and RPC APIs	✓	✓
Use a graphical UI	✓	
Annotate video using predefined labels *Pre-trained models leverage vast libraries of predefined labels.*		✓
Annotate video using custom labels *Train models to classify video with custom labels of your choice.*	✓	
Stored video analysis	✓	✓
Streaming video analysis (beta)	✓	✓
Shot change detection	✓	✓
Object detection and tracking	✓	✓
Text detection and extraction using OCR		✓
Explicit content detection		✓
Automated closed captioning and subtitles		✓
Logo recognition		✓
Celebrity recognition (limited access)		✓
Face detection (beta)		✓
Person detection with pose estimation (beta)		✓

Source: https://cloud.google.com/video-intelligence



Use cases

USE CASE

Content moderation

Identify when inappropriate content is being shown in a given video. You can instantly conduct content moderation across petabytes of data and more quickly and efficiently filter your content or user-generated content.

USE CASE

Recommended content

Build a content recommendation engine with labels generated by Video Intelligence API and a user's viewing history and preferences. This will simplify content discovery for your users and guide them to the most relevant content that they want.

USE CASE

Media archives

Create an indexed archive of your entire video library by using the metadata from Video Intelligence API. Ideal for mass media companies, Video Intelligence API can automatically analyze content and make the results immediately accessible via the API.

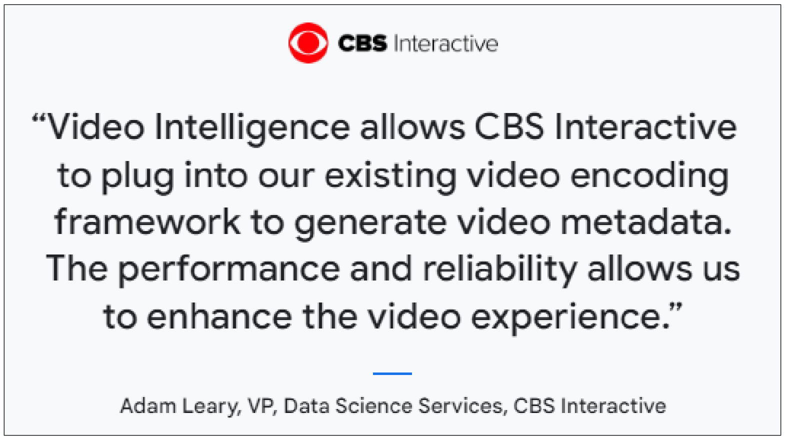
USE CASE

Contextual advertisements

You can identify appropriate locations in videos to insert ads that are contextually relevant to the video content. This can be done by matching the timeframe-specific labels of your video content with the content of your advertisements.

Source: https://cloud.google.com/video-intelligence

83. On information and belief, Google's customers use Video Intelligence to generate video metadata using the context-based techniques of the '972 Patent in order to "enhance the video experience."



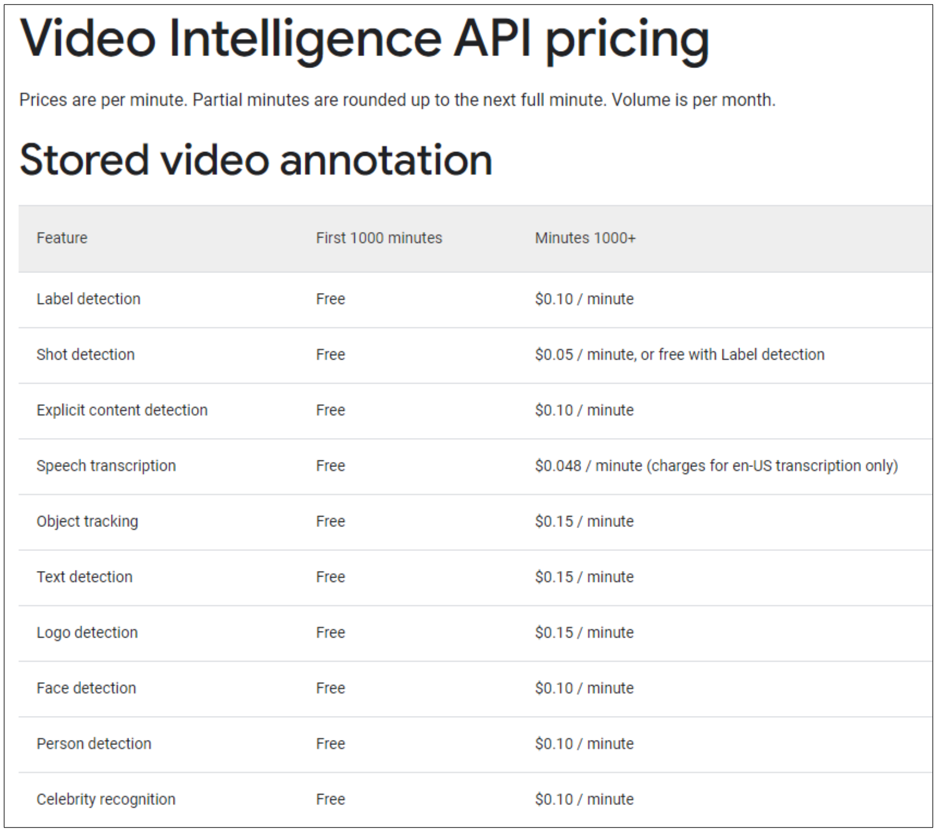
CBS Interactive

"Video Intelligence allows CBS Interactive to plug into our existing video encoding framework to generate video metadata. The performance and reliability allows us to enhance the video experience."

Adam Leary, VP, Data Science Services, CBS Interactive

Source : https://cloud.google.com/video-intelligence

84. Google provides the Video Intelligence product to users through Google Cloud Platform ("GCP") and charges users for the amount of resources (*e.g.*, cloud server virtual CPUs) consumed for video processing on a per minute basis.



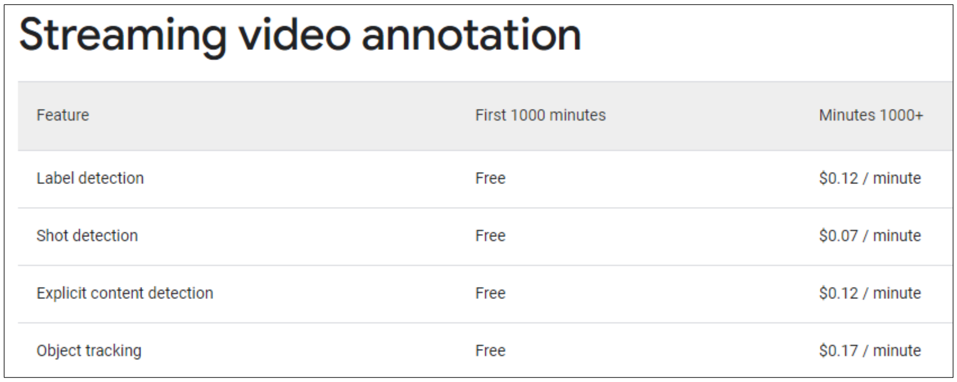
Video Intelligence API pricing

Prices are per minute. Partial minutes are rounded up to the next full minute. Volume is per month.

Stored video annotation

Feature	First 1000 minutes	Minutes 1000+
Label detection	Free	$0.10 / minute
Shot detection	Free	$0.05 / minute, or free with Label detection
Explicit content detection	Free	$0.10 / minute
Speech transcription	Free	$0.048 / minute (charges for en-US transcription only)
Object tracking	Free	$0.15 / minute
Text detection	Free	$0.15 / minute
Logo detection	Free	$0.15 / minute
Face detection	Free	$0.10 / minute
Person detection	Free	$0.10 / minute
Celebrity recognition	Free	$0.10 / minute

Source: https://cloud.google.com/video-intelligence/pricing



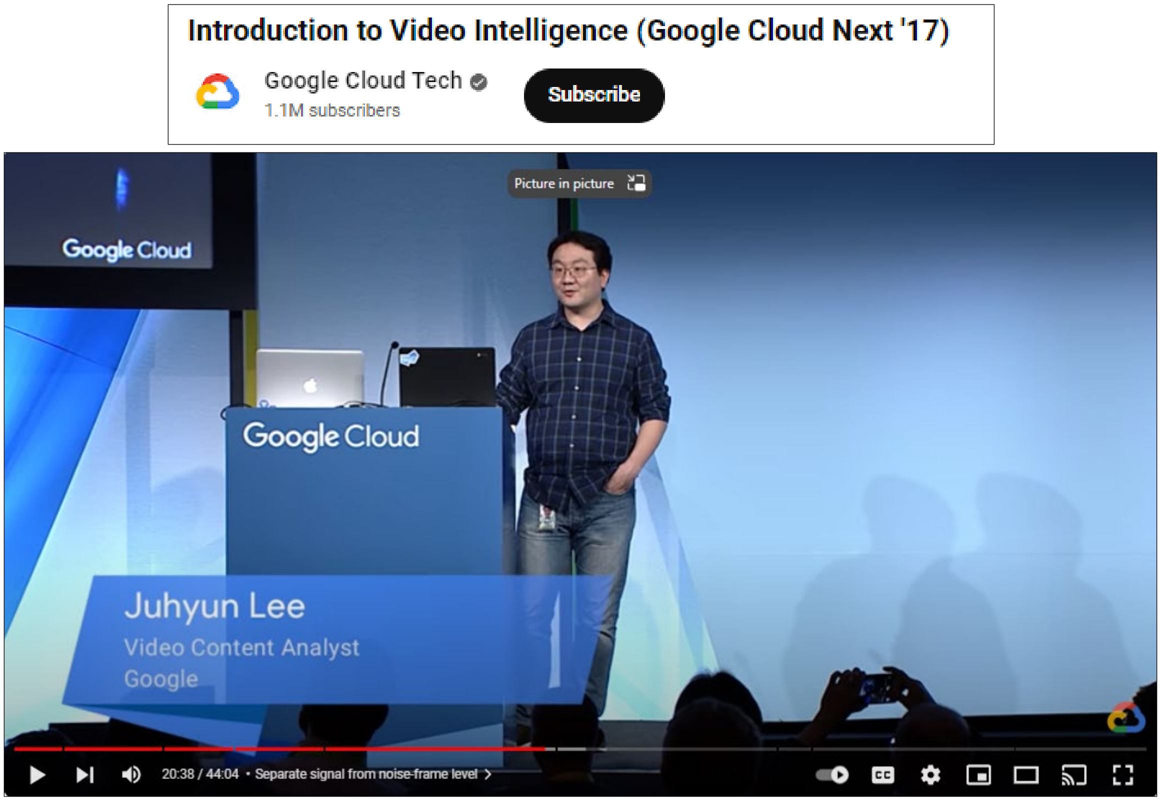
Streaming video annotation

Feature	First 1000 minutes	Minutes 1000+
Label detection	Free	$0.12 / minute
Shot detection	Free	$0.07 / minute
Explicit content detection	Free	$0.12 / minute
Object tracking	Free	$0.17 / minute

Source: https://cloud.google.com/video-intelligence/pricing#streaming_video_annotation

85. With regard to claim 1 of the '972 Patent, Google practices "[a] method to generate video data from a video comprising: generating audio files and image files from the video." For

example, at the Google Next 2017 developer conference, Google explained that Video Intelligence begins by decoding video into video data, audio data, subtitle data, and various other metadata.[41]



Introduction to Video Intelligence (Google Cloud Next '17)

Google Cloud Tech

1.1M subscribers

Source: https://www.youtube.com/watch?v=y-k8oelbmGc

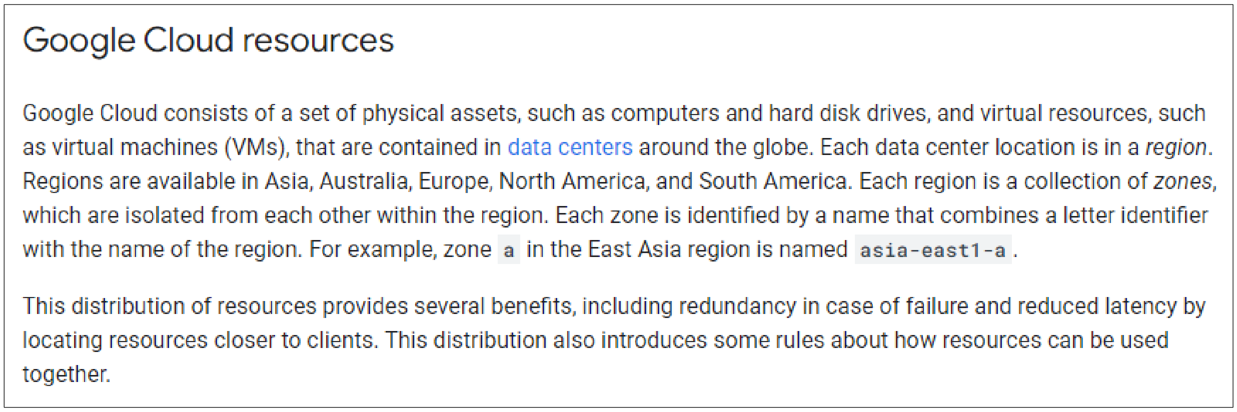
> So given a video in your GCS [Google Cloud Storage] bucket, **the first thing that we do is we decode the video**. And usually a video container contains like **audio streams, video streams, subtitles and all of the metadata**.[42]

86. The Accused Product performs the step of distributing the image files across a plurality of processors and processing the image files in parallel, wherein processing the image files comprises extracting one or more objects and identifying the one or more objects. On information and belief, Video Intelligence executes on a distributed network of one or more Google Cloud servers. As Google explains, Video Intelligence uses the Google Cloud Platform

[41] Introduction to Video Intelligence (Google Cloud Next '17) (available at https://www.youtube.com/watch?v=y-k8oelbmGc).

[42] *Id*. at 21:00 to 22:14 (emphasis added).

("GCP") which comprises a number of computer resources housed in Google's data centers around the world.[43]

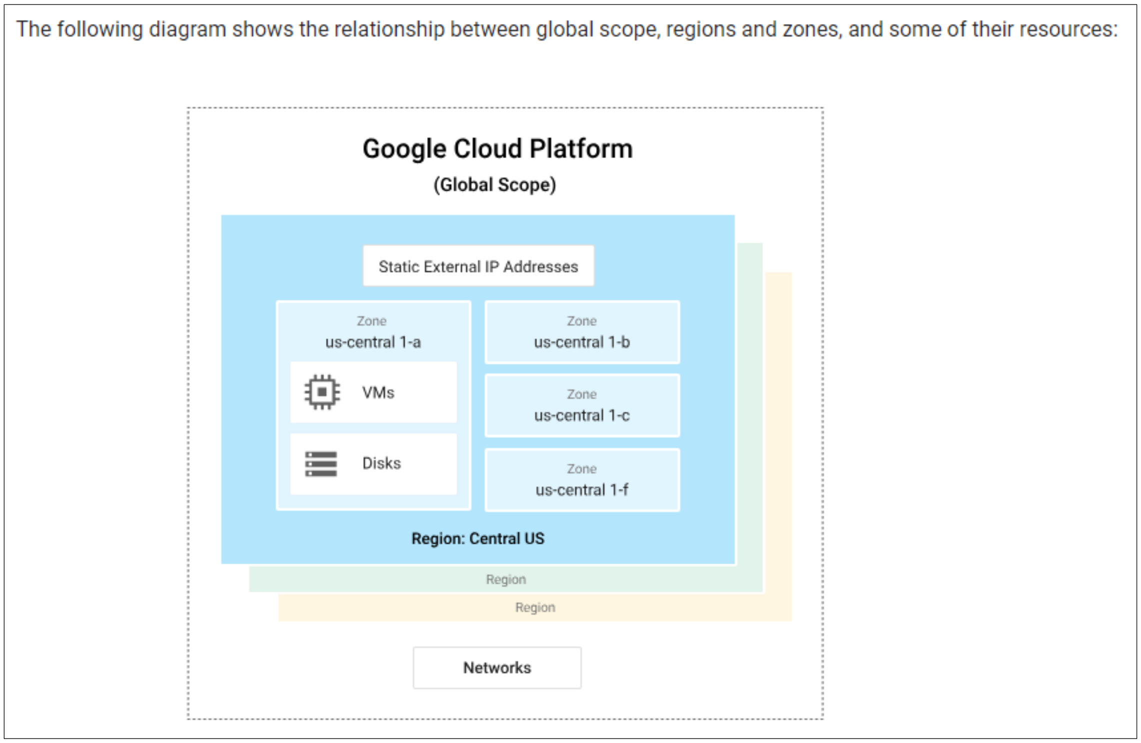


Google Cloud resources

Google Cloud consists of a set of physical assets, such as computers and hard disk drives, and virtual resources, such as virtual machines (VMs), that are contained in data centers around the globe. Each data center location is in a *region*. Regions are available in Asia, Australia, Europe, North America, and South America. Each region is a collection of *zones*, which are isolated from each other within the region. Each zone is identified by a name that combines a letter identifier with the name of the region. For example, zone `a` in the East Asia region is named `asia-east1-a`.

This distribution of resources provides several benefits, including redundancy in case of failure and reduced latency by locating resources closer to clients. This distribution also introduces some rules about how resources can be used together.

Source: https://cloud.google.com/docs/overview

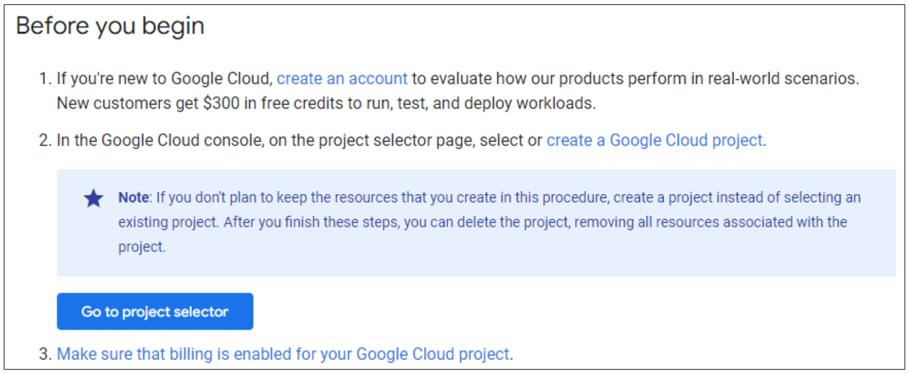


The following diagram shows the relationship between global scope, regions and zones, and some of their resources:

Source: https://cloud.google.com/docs/overview

87. Google explains that a GCP account and project are required in order to use the Video Intelligence API.

[43] *See* https://cloud.google.com/docs/overview.



Before you begin

1. If you're new to Google Cloud, create an account to evaluate how our products perform in real-world scenarios. New customers get $300 in free credits to run, test, and deploy workloads.
2. In the Google Cloud console, on the project selector page, select or create a Google Cloud project.

★ **Note**: If you don't plan to keep the resources that you create in this procedure, create a project instead of selecting an existing project. After you finish these steps, you can delete the project, removing all resources associated with the project.

3. Make sure that billing is enabled for your Google Cloud project.

Source: https://cloud.google.com/video-intelligence/docs/annotate-video-command-line

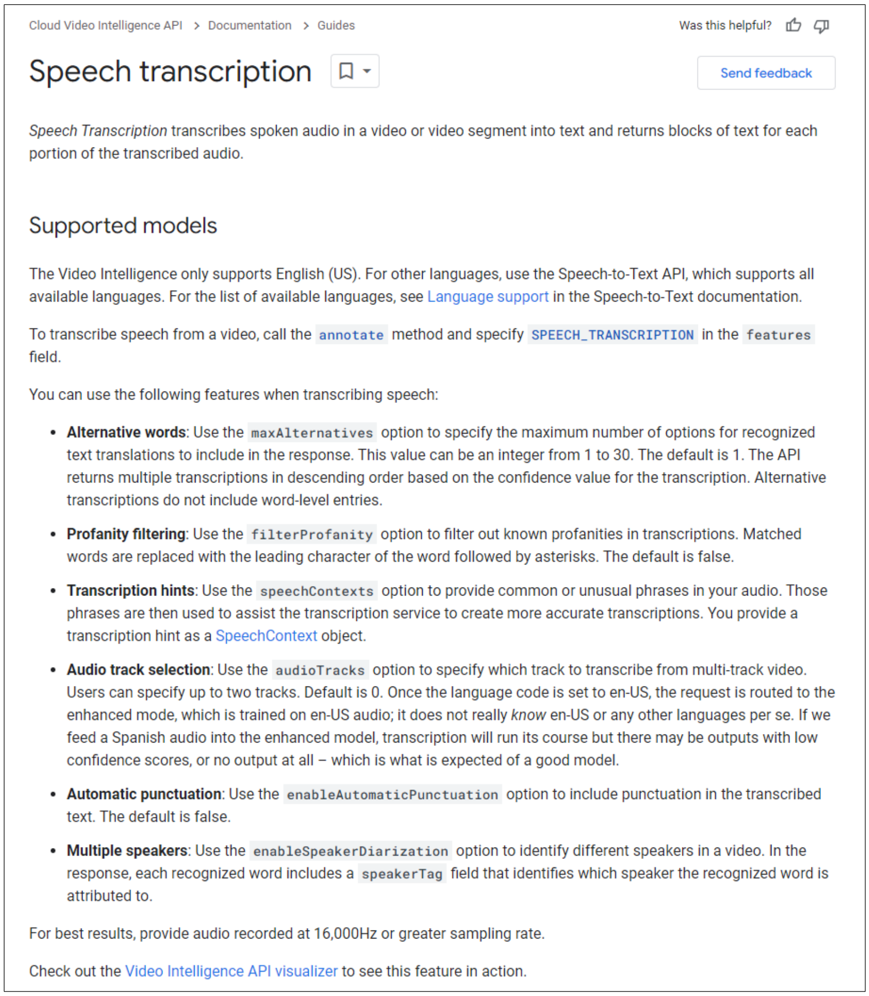
88. The Accused Product additionally performs the step of "processing the audio files" and "converting audio files associated with the video to text." For example, as explained above, Video Intelligence begins by decoding video files into image data, audio data, subtitle data, and/or additional data streams.

> So given a video in your GCS [Google Cloud Storage] bucket, **the first thing that we do is we decode the video**. And usually a video container contains like **audio streams, video streams, subtitles and all of the metadata**.[44]

89. Google explains that Video Intelligence extracts audio data and converts that data into text using the "SPEECH_TRANSCRIPTION" feature.[45] According to Google, Video Intelligence extracts audio data from the video and transcribes that data into "text translations" using the "SPEECH_TRANSCRIPTION" feature. The "maxAlternatives" option allows users to customize the flexibility of the natural language models used by Video Intelligence in order to influence the confidence scores generated by the ML model during the recognition process. The API then returns multiple transcriptions based on the confidence value for the transcription.

[44] https://www.youtube.com/watch?v=y-k8oelbmGc at 21:00 to 22:14 (emphasis added).

[45] *See* https://cloud.google.com/video-intelligence/docs/feature-speech-transcription.



Speech transcription

Speech Transcription transcribes spoken audio in a video or video segment into text and returns blocks of text for each portion of the transcribed audio.

Supported models

The Video Intelligence only supports English (US). For other languages, use the Speech-to-Text API, which supports all available languages. For the list of available languages, see Language support in the Speech-to-Text documentation.

To transcribe speech from a video, call the `annotate` method and specify `SPEECH_TRANSCRIPTION` in the `features` field.

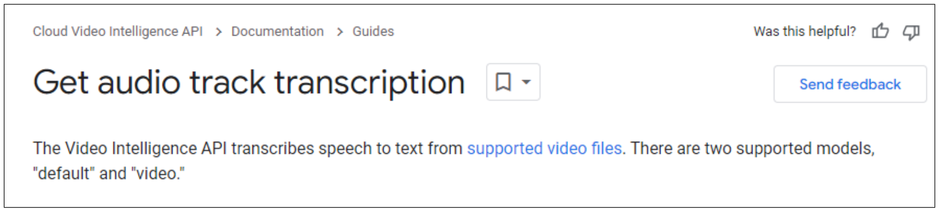
You can use the following features when transcribing speech:

- **Alternative words**: Use the `maxAlternatives` option to specify the maximum number of options for recognized text translations to include in the response. This value can be an integer from 1 to 30. The default is 1. The API returns multiple transcriptions in descending order based on the confidence value for the transcription. Alternative transcriptions do not include word-level entries.
- **Profanity filtering**: Use the `filterProfanity` option to filter out known profanities in transcriptions. Matched words are replaced with the leading character of the word followed by asterisks. The default is false.
- **Transcription hints**: Use the `speechContexts` option to provide common or unusual phrases in your audio. Those phrases are then used to assist the transcription service to create more accurate transcriptions. You provide a transcription hint as a SpeechContext object.
- **Audio track selection**: Use the `audioTracks` option to specify which track to transcribe from multi-track video. Users can specify up to two tracks. Default is 0. Once the language code is set to en-US, the request is routed to the enhanced mode, which is trained on en-US audio; it does not really *know* en-US or any other languages per se. If we feed a Spanish audio into the enhanced model, transcription will run its course but there may be outputs with low confidence scores, or no output at all – which is what is expected of a good model.
- **Automatic punctuation**: Use the `enableAutomaticPunctuation` option to include punctuation in the transcribed text. The default is false.
- **Multiple speakers**: Use the `enableSpeakerDiarization` option to identify different speakers in a video. In the response, each recognized word includes a `speakerTag` field that identifies which speaker the recognized word is attributed to.

For best results, provide audio recorded at 16,000Hz or greater sampling rate.

Check out the Video Intelligence API visualizer to see this feature in action.

Source: https://cloud.google.com/video-intelligence/docs/feature-speech-transcription



Get audio track transcription

The Video Intelligence API transcribes speech to text from supported video files. There are two supported models, "default" and "video."

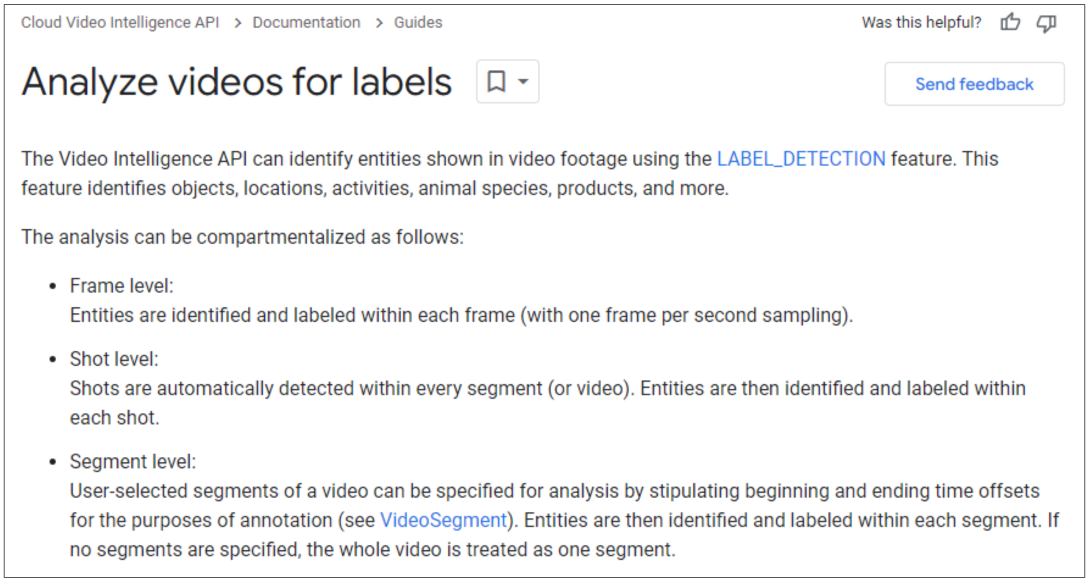
Source: https://cloud.google.com/video-intelligence/docs/transcription

90. As Google explains, using audio data is important and necessary to ensure the accuracy of the Video Intelligence system.

> Assume that you have a video of an animated music video. It would be sad if we give you a label saying that it's an animated cartoon. **You definitely want your**

> **audio signal in there. So we are looking at various audio signals to improve the quality of the model.**[46]

91. The Accused Product performs the step of converting the image files associated with the video to video data. For example, using Video Intelligence, labels may be created for the extracted image files at the "frame," "shot," and/or "video" levels.



Analyze videos for labels

The Video Intelligence API can identify entities shown in video footage using the LABEL_DETECTION feature. This feature identifies objects, locations, activities, animal species, products, and more.

The analysis can be compartmentalized as follows:

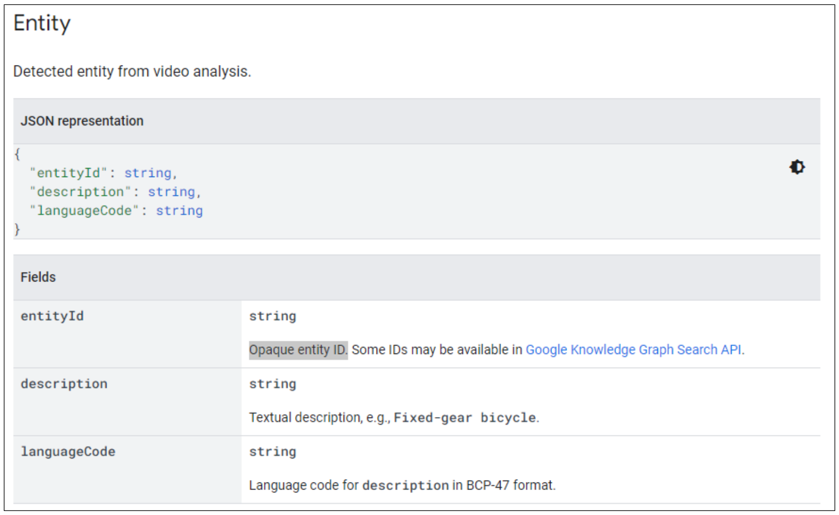
- Frame level:
  Entities are identified and labeled within each frame (with one frame per second sampling).
- Shot level:
  Shots are automatically detected within every segment (or video). Entities are then identified and labeled within each shot.
- Segment level:
  User-selected segments of a video can be specified for analysis by stipulating beginning and ending time offsets for the purposes of annotation (see VideoSegment). Entities are then identified and labeled within each segment. If no segments are specified, the whole video is treated as one segment.

Source: https://cloud.google.com/video-intelligence/docs/analyze-labels

92. Google explains that Video Intelligence identifies and creates annotations for "entities" within a video, video segment, and/or frame using the "LABEL_DETECTION" feature.[47] Entity and label information is associated with the analyzed image data along with other information, such as time stamps as to where in the video a particular entity and/or label appears, so that that data can be retrieved later using various API calls such as "Entity," "LabelSegment," and "LabelFrame."

[46] https://www.youtube.com/watch?v=y-k8oelbmGc at 23:44 to 24:07 (emphasis added).

[47] https://cloud.google.com/video-intelligence/docs/analyze-labels.



Entity

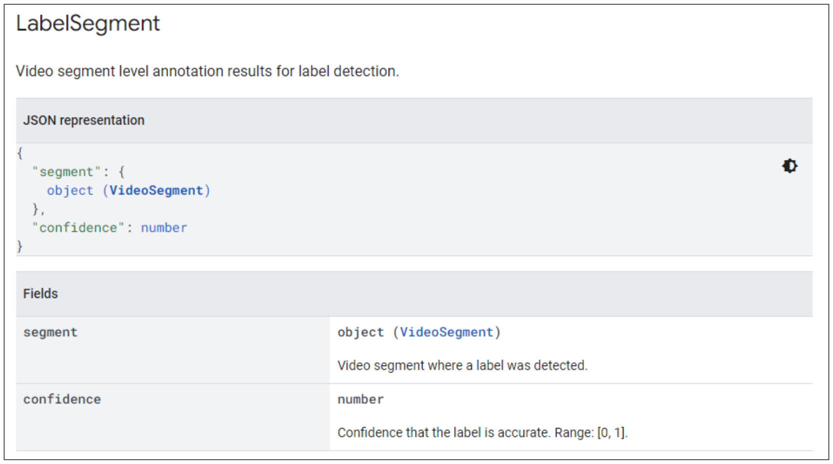
Detected entity from video analysis.

JSON representation

```
{
  "entityId": string,
  "description": string,
  "languageCode": string
}
```

Fields	
`entityId`	`string` Opaque entity ID. Some IDs may be available in Google Knowledge Graph Search API.
`description`	`string` Textual description, e.g., `Fixed-gear bicycle`.
`languageCode`	`string` Language code for `description` in BCP-47 format.

Source: https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse



LabelSegment

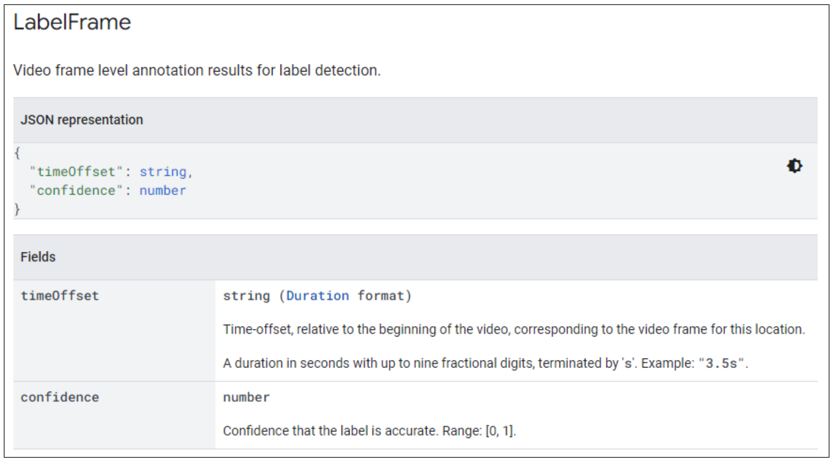
Video segment level annotation results for label detection.

JSON representation

```
{
  "segment": {
    object (VideoSegment)
  },
  "confidence": number
}
```

Fields	
`segment`	`object (VideoSegment)` Video segment where a label was detected.
`confidence`	`number` Confidence that the label is accurate. Range: [0, 1].

Source: https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse



LabelFrame

Video frame level annotation results for label detection.

JSON representation

```
{
  "timeOffset": string,
  "confidence": number
}
```

Fields	
`timeOffset`	`string (Duration format)` Time-offset, relative to the beginning of the video, corresponding to the video frame for this location. A duration in seconds with up to nine fractional digits, terminated by 's'. Example: `"3.5s"`.
`confidence`	`number` Confidence that the label is accurate. Range: [0, 1].

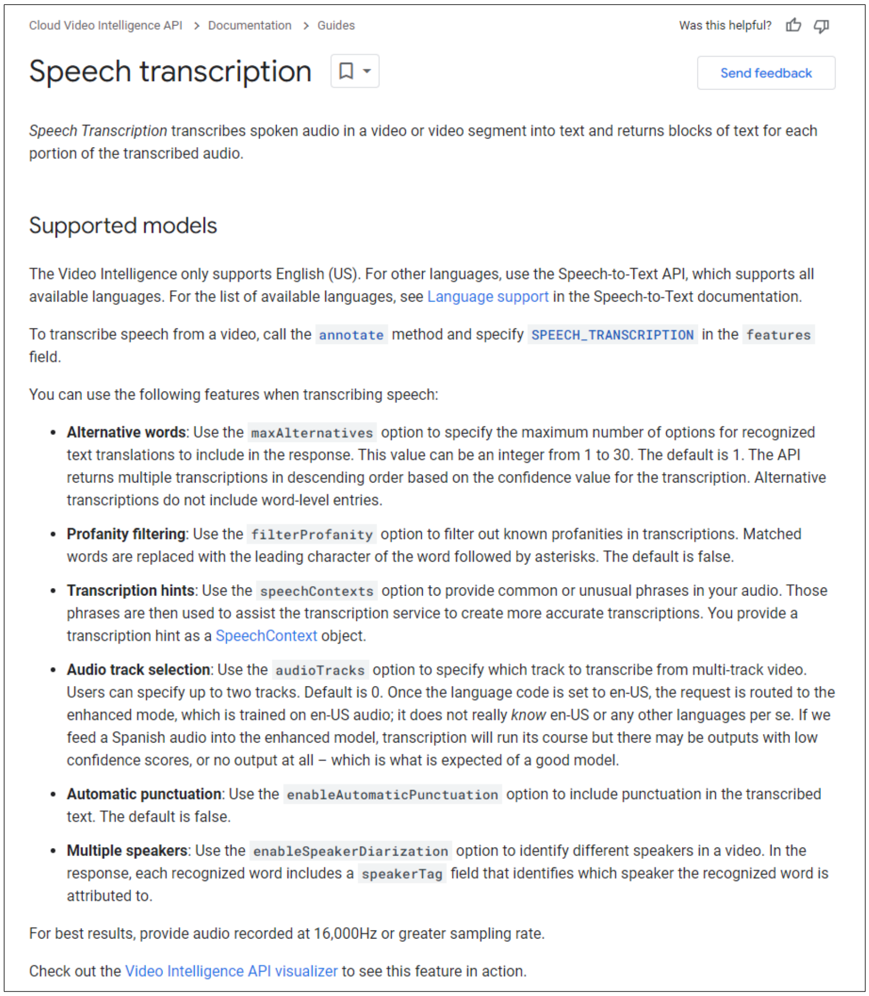
Source: https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse

93. The Accused Product performs the step of generating a topical meta-data that describes content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio files. For example, Google explains that Video Intelligence first identifies the various objects in the images that make up a given video and uses that information to assign labels.

> And we are first only interested in the video stream. And a video stream is often a sequence of images . . . a lot of them. For a five-minute video at 24 frames per second, you will have like 7,000, 8,000 images in them. And they will probably in sequence look a lot similar. So classifying each of those is not too efficient, let's say. So we shot sample the images, and say at one FPS. So one frame per second, we apply an image classifier. An image classifier that is very similar to – that you can get through the Google Cloud Vision API. And we do it for every image in the entire video. And for each frame, you will get image features such as this elephant with 97% accuracy, or animal, or river in that case. And, if you keep going on, you will get other labels for crocodiles and lions.[48]

94. Video Intelligence also generates topical metadata by deriving semantic information from extracted audio data. For example, Video Intelligence extracts audio data from a video and transcribes that data into "text translations" using the natural language models that enable the "SPEECH_TRANSCRIPTION" feature. As discussed above, the "maxAlternatives" option allows users to customize the flexibility of the natural language models used by Video Intelligence in order to influence the confidence scores generated by the ML model during the recognition process.

[48] https://www.youtube.com/watch?v=y-k8oelbmGc at 21:00 to 22:14.



Speech transcription

Speech Transcription transcribes spoken audio in a video or video segment into text and returns blocks of text for each portion of the transcribed audio.

Supported models

The Video Intelligence only supports English (US). For other languages, use the Speech-to-Text API, which supports all available languages. For the list of available languages, see Language support in the Speech-to-Text documentation.

To transcribe speech from a video, call the `annotate` method and specify `SPEECH_TRANSCRIPTION` in the `features` field.

You can use the following features when transcribing speech:

- **Alternative words**: Use the `maxAlternatives` option to specify the maximum number of options for recognized text translations to include in the response. This value can be an integer from 1 to 30. The default is 1. The API returns multiple transcriptions in descending order based on the confidence value for the transcription. Alternative transcriptions do not include word-level entries.
- **Profanity filtering**: Use the `filterProfanity` option to filter out known profanities in transcriptions. Matched words are replaced with the leading character of the word followed by asterisks. The default is false.
- **Transcription hints**: Use the `speechContexts` option to provide common or unusual phrases in your audio. Those phrases are then used to assist the transcription service to create more accurate transcriptions. You provide a transcription hint as a SpeechContext object.
- **Audio track selection**: Use the `audioTracks` option to specify which track to transcribe from multi-track video. Users can specify up to two tracks. Default is 0. Once the language code is set to en-US, the request is routed to the enhanced mode, which is trained on en-US audio; it does not really *know* en-US or any other languages per se. If we feed a Spanish audio into the enhanced model, transcription will run its course but there may be outputs with low confidence scores, or no output at all – which is what is expected of a good model.
- **Automatic punctuation**: Use the `enableAutomaticPunctuation` option to include punctuation in the transcribed text. The default is false.
- **Multiple speakers**: Use the `enableSpeakerDiarization` option to identify different speakers in a video. In the response, each recognized word includes a `speakerTag` field that identifies which speaker the recognized word is attributed to.

For best results, provide audio recorded at 16,000Hz or greater sampling rate.

Check out the Video Intelligence API visualizer to see this feature in action.

Source: https://cloud.google.com/video-intelligence/docs/feature-speech-transcription

95. As Google explains, both the semantic information from the identification of the one or more objects and the semantic information from the audio files are used to generate topical metadata. For example, Google explains that Video Intelligence utilizes an "aggregating video classifier" that combines the outputs of the analyzed video image data and audio, text, subtitle or other data to determine a contextual topic.

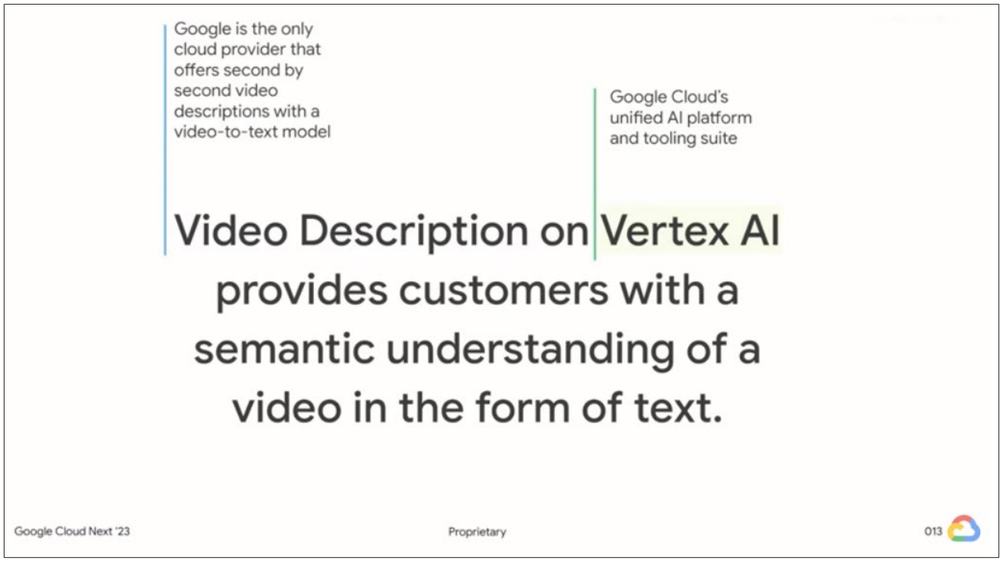
> And at the end of the video, what we do is **we have an aggregating video classifier, and this is really what sits at the core of the -- of this particular API. And this – given all of these features throughout the entire video it will then say, hey, this is a documentary about animals in Africa. Another similar example would be if you have a video of people in costumes or pumpkins and candies, this classifier will be able to tell you, hey, this is a video about Halloween.** And this

> is only possible because we have trained this model on millions of YouTube videos, and human-rated and also like-verified.[49]

96. Google also explains that using audio data is important and necessary to ensure the accuracy of the Video Intelligence system.

> Assume that you have a video of an animated music video. It would be sad if we give you a label saying that it's an animated cartoon. **You definitely want your audio signal in there. So we are looking at various audio signals to improve the quality of the model.**[50]

97. In addition, Video Intelligence via Vertex AI includes a video description (video-to-text) feature that allows users to generate a semantic understanding of video in the form of text.



Google is the only cloud provider that offers second by second video descriptions with a video-to-text model

Google Cloud's unified AI platform and tooling suite

Video Description on Vertex AI provides customers with a semantic understanding of a video in the form of text.

Google Cloud Next '23

Proprietary

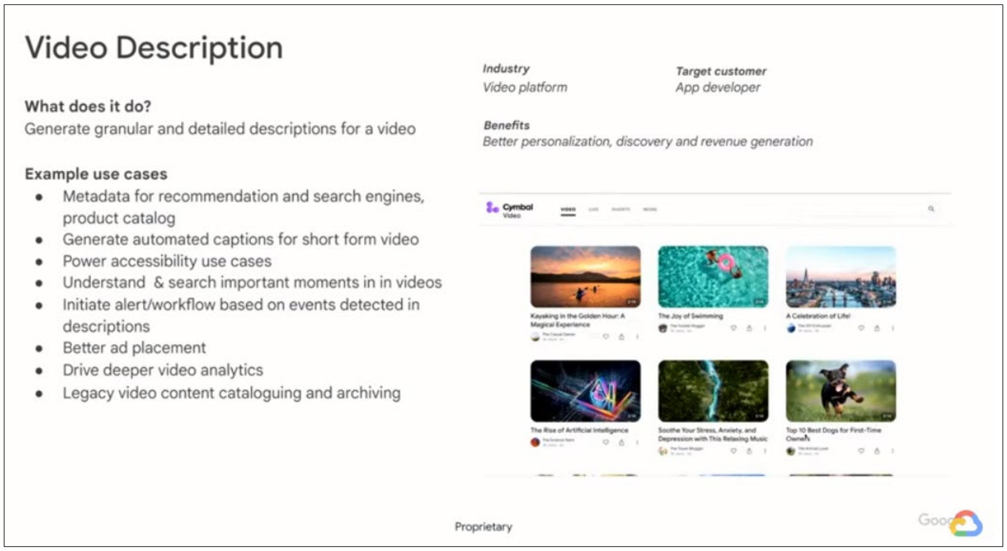
013

Source: https://www.youtube.com/watch?v=F4YKREYPkEI at 7:19.

98. As Google explains, Video Intelligence via Vertex AI may be used to generate topical metadata, such as video descriptions or summaries, for analyzed videos.

[49] *See* https://www.youtube.com/watch?v=y-k8oelbmGc at 22:14 to 23:11 (emphasis added).

[50] https://www.youtube.com/watch?v=y-k8oelbmGc at 23:44 to 24:07 (emphasis added).



Video Description

What does it do?
Generate granular and detailed descriptions for a video

Example use cases

- Metadata for recommendation and search engines, product catalog
- Generate automated captions for short form video
- Power accessibility use cases
- Understand & search important moments in in videos
- Initiate alert/workflow based on events detected in descriptions
- Better ad placement
- Drive deeper video analytics
- Legacy video content cataloguing and archiving

Industry
Video platform

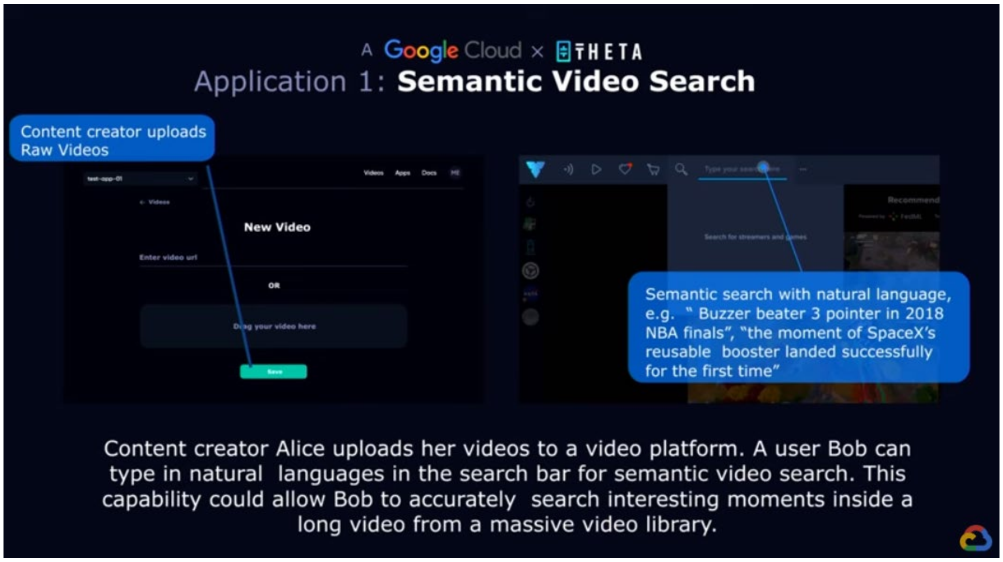
Target customer
App developer

Benefits
Better personalization, discovery and revenue generation

Cymbal Video
Kayaking in the Golden Hour: A Magical Experience
The Joy of Swimming
A Celebration of Life!
The Rise of Artificial Intelligence
Soothe Your Stress, Anxiety, and Depression with This Relaxing Music
Top 10 Best Dogs for First-Time Owners

Proprietary

Source: https://www.youtube.com/watch?v=F4YKREYPkEI at 8:33.



A Google Cloud × THETA

Application 1: Semantic Video Search

Content creator uploads Raw Videos

Semantic search with natural language, e.g. " Buzzer beater 3 pointer in 2018 NBA finals", "the moment of SpaceX's reusable booster landed successfully for the first time"

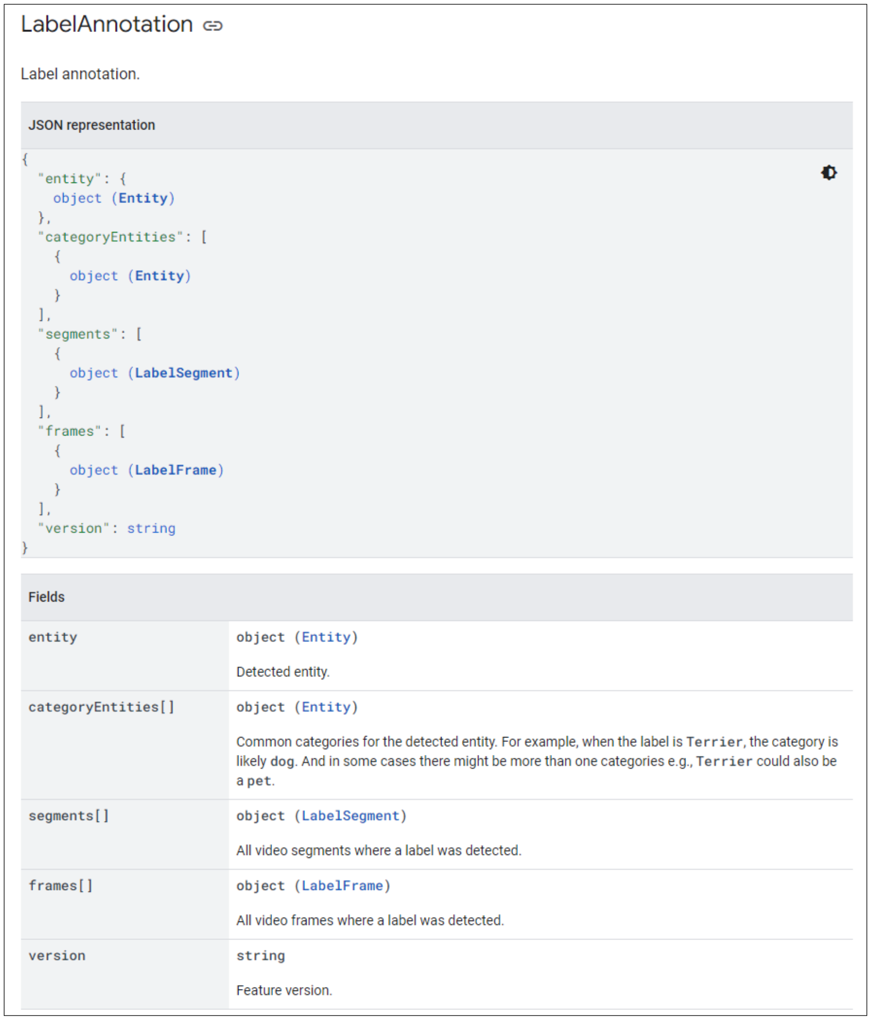
Content creator Alice uploads her videos to a video platform. A user Bob can type in natural languages in the search bar for semantic video search. This capability could allow Bob to accurately search interesting moments inside a long video from a massive video library.

Source: https://www.youtube.com/watch?v=F4YKREYPkEI at 27:13.

Traditionally, video search only allows searching against title, description, and possibly tags. But these descriptions usually typically provide the video uploaders with some poor quality that may not be able to capture the full details of the video. With model pipeline, we can achieve [a] much better experience. So two images are shown. It depicts the content creator Alice uploading videos to a video platform. **And after the raw video is processed by the model pipeline, a user, Bob, can type in the natural language into the search bar for semantic video search.** This could allow Bob to accurately search interesting moments. Inside the

> long video, a massive video library is automatically—it'll locate the replay position inside these typically long videos.[51]

99. The Accused Product performs the step of adding the topical meta-data to the video. On information and belief, "entities," "labels," "tags" and other metadata generated from the analyzed video are associated with the video file. Google explains that this annotation metadata may be subsequently accessed using, for example, calls to the Video Intelligence API. The "LabelAnnotation" parameter, for instance, allows a user to access "label" annotation metadata associated with an analyzed video.



LabelAnnotation

Label annotation.

JSON representation

```
{
  "entity": {
    object (Entity)
  },
  "categoryEntities": [
    {
      object (Entity)
    }
  ],
  "segments": [
    {
      object (LabelSegment)
    }
  ],
  "frames": [
    {
      object (LabelFrame)
    }
  ],
  "version": string
}
```

Fields	
entity	object (Entity) Detected entity.
categoryEntities[]	object (Entity) Common categories for the detected entity. For example, when the label is Terrier, the category is likely dog. And in some cases there might be more than one categories e.g., Terrier could also be a pet.
segments[]	object (LabelSegment) All video segments where a label was detected.
frames[]	object (LabelFrame) All video frames where a label was detected.
version	string Feature version.

Source: https://cloud.google.com/video-intelligence/docs/reference/rest/v1/AnnotateVideoResponse#labelannotation

[51] https://www.youtube.com/watch?v=F4YKREYPkEI at 26:44 to 27:30 (emphasis added).

100. The Accused Product performs the step of cross-referencing the text and the video data based on the generated topical meta-data to determine topics. Video Intelligence includes an aggregating classifier that combines the outputs of analyzed video image and audio, text, subtitle data and/or other annotation metadata to determine a contextual topic for a given video.

> And at the end of the video, what we do is **we have an aggregating video classifier, and this is really what sits at the core of the—of this particular API**. **And this—given all of these features throughout the entire video it will then say, hey, this is a documentary about animals in Africa. Another similar example would be if you have a video of people in costumes or pumpkins and candies, this classifier will be able to tell you, hey, this is a video about Halloween.** And this is only possible because we have trained this model on millions of YouTube videos, and human-rated and also like-verified.[52]

101. In addition, as discussed above, Video Intelligence via Vertex AI allows customers to generate a semantic understanding of video in the form of text based on the combined product of recognized objects and text from the constituent video and audio components of a video file.[53]

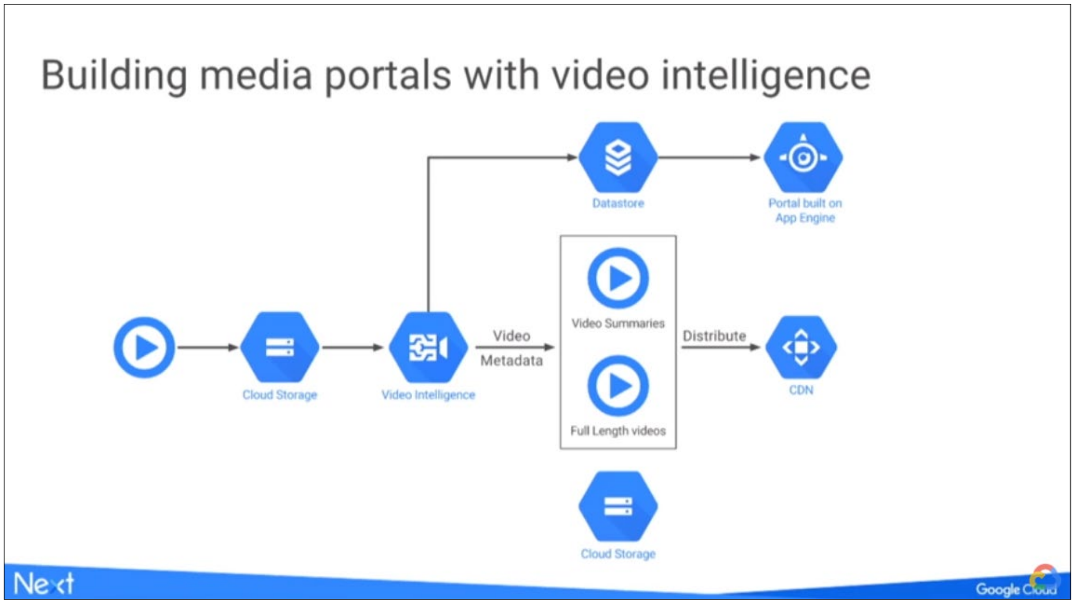
102. The Accused Product further performs the step of "generating video text based on the cross-referencing, wherein the video text describes content of the video." For example, as discussed above, Google explains that media publishers can use the results output by Video Intelligence to generate a highlight reel of video content related to a particular topic or to rapidly create video descriptions for searching.[54] Media publishers may also use Video Intelligence to

[52] https://www.youtube.com/watch?v=y-k8oelbmGc at 22:14 to 23:11 (emphasis added).

[53] https://www.youtube.com/watch?v=F4YKREYPkEI at 7:17–7:24 ("Video Description in Vertex AI is a feature with which customers can get semantic understanding of a video in the form of text."); *see also*, *id*. at 26:44 to 27:30 ("With model pipeline, we can achieve [a] much better experience. So two images are shown. It depicts the content creator Alice uploading videos to a video platform. And after the raw video is processed by the model pipeline, a user, Bob, can type in the natural language into the search bar for semantic video search. This could allow Bob to accurately search interesting moments. Inside the long video, a massive video library is automatically—it'll locate the replay position inside these typically long videos.").

[54] https://www.youtube.com/watch?v=mDAoLO4G4CQ at 1:16 to 1:57. ("A media publisher could have hundreds of petabytes of video data sitting in storage buckets and one common thing they might want to do is create a highlight reel focused on a specific type of content or search their large library for a specific entity. So let's see how we would use the Video Intelligence API to search a large library of videos, given all this metadata that we get back from it.").

easily generate thumbnails or video summaries that highlight the most interesting aspects of a particular video file.



Building media portals with video intelligence

Source: https://www.youtube.com/watch?v=y-k8oelbmGc at 31:21 (high-level illustration of Video Intelligence's use for creating thumbnails or "video summaries").

103. Additionally, Google explains that Video Intelligence via Vertex AI may be used to provide video descriptions or summaries for analyzed video files based on the extracted semantic information from the image data, audio data, and/or other data:

> Let's see a quick demo again. So similarly, same studio, now we have a tab called Video Description. You go ahead and select your video. Again, for a 2-minute video, it's about 20 to 30 seconds latency. And there you go. For every 15-second chunk, it gives you a description. And in the UI experience, you can actually click on the timestamp and go to that portion of the video so you understand the relevancy and you maintain context**.** You can save this as a JSON file. And by the way, all the three features I talked about today all have API support right out the box.[55]

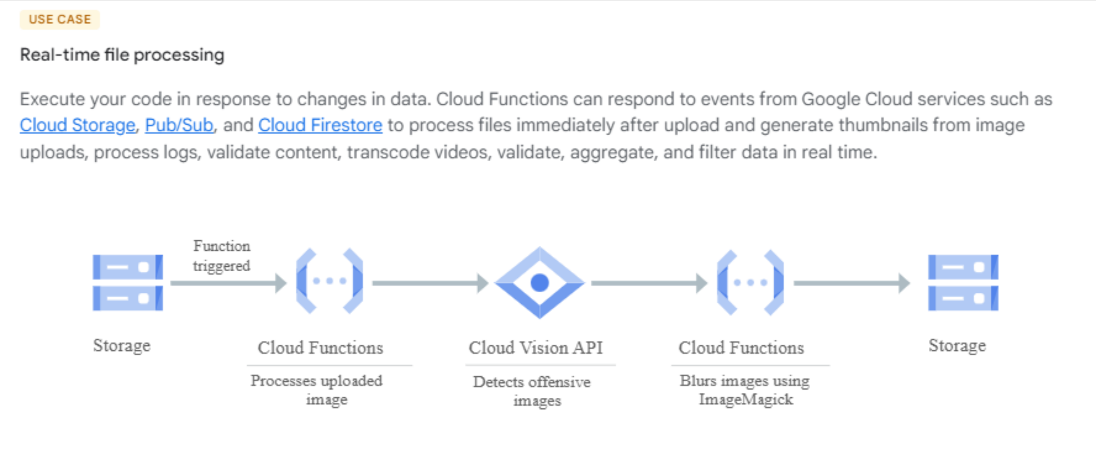
104. Finally, the Accused Product performs the steps of "generating a text, image, or animation based on the video text" and "placing the text, image, or animation in the video." For example, as previously discussed, media publishers may also use Video Intelligence to easily

[55] https://www.youtube.com/watch?v=F4YKREYPkEI at 11:32 to 12:16.

generate thumbnails or "video summaries" that highlight the most interesting aspects of a particular video file.

> Now I'm going to talk a little bit about what's coming next, a preview on video previews, also known as video summaries or thumbnail. . . . So assume that you have a bunch of customer videos and you have to show it in a list; you need to make a listing page. Now you need thumbnails for each of those videos. How do you pick the right video frame for the video that you don't know about? You could pick the first frame in the video, or you could apply the Wadsworth Constant, and skip to . . . 30% of the video and then show that frame. At Google, we use machine learning and what we do is we have a machine learned classifier that tells what's the interesting-ness of this particular video frame as a thumbnail.[56]

105. Google also explains that Video Intelligence may be used to automatically detect inappropriate content and blur the identified frames, shots, and/or segments. As Google explains in its developer documentation, this may be performed using the ImageMagick function provided by Google Cloud Functions ("GCF").



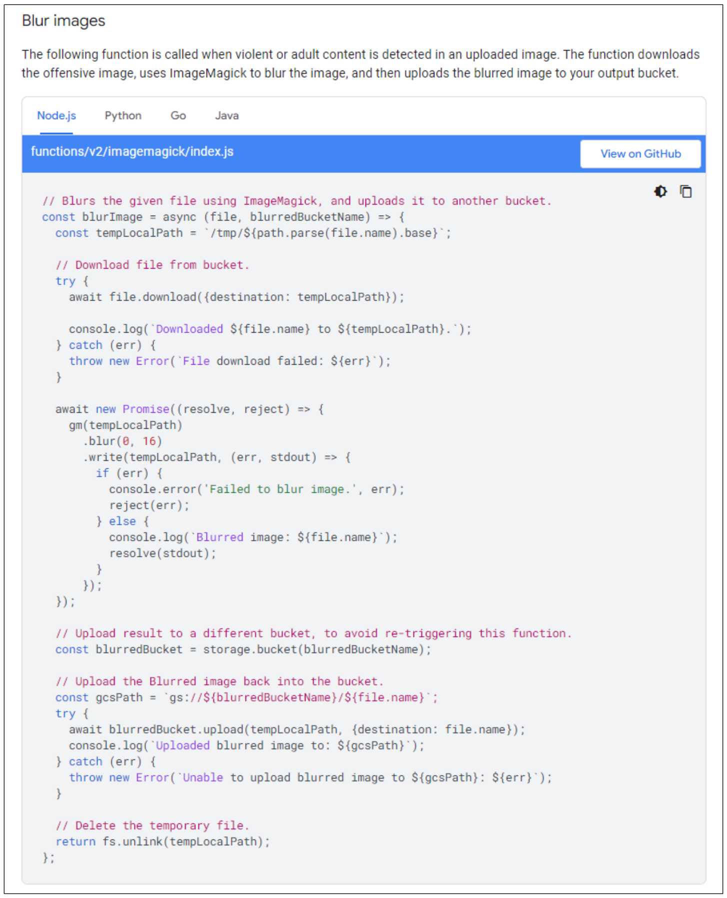
USE CASE

Real-time file processing

Execute your code in response to changes in data. Cloud Functions can respond to events from Google Cloud services such as Cloud Storage, Pub/Sub, and Cloud Firestore to process files immediately after upload and generate thumbnails from image uploads, process logs, validate content, transcode videos, validate, aggregate, and filter data in real time.

Source: https://cloud.google.com/functions/#video-and-image-analysis

[56] https://www.youtube.com/watch?v=y-k8oelbmGc at 24:57 to 26:07.



Blur images

The following function is called when violent or adult content is detected in an uploaded image. The function downloads the offensive image, uses ImageMagick to blur the image, and then uploads the blurred image to your output bucket.

```
// Blurs the given file using ImageMagick, and uploads it to another bucket.
const blurImage = async (file, blurredBucketName) => {
  const tempLocalPath = `/tmp/${path.parse(file.name).base}`;

  // Download file from bucket.
  try {
    await file.download({destination: tempLocalPath});

    console.log(`Downloaded ${file.name} to ${tempLocalPath}.`);
  } catch (err) {
    throw new Error(`File download failed: ${err}`);
  }

  await new Promise((resolve, reject) => {
    gm(tempLocalPath)
      .blur(0, 16)
      .write(tempLocalPath, (err, stdout) => {
        if (err) {
          console.error('Failed to blur image.', err);
          reject(err);
        } else {
          console.log(`Blurred image: ${file.name}`);
          resolve(stdout);
        }
      });
  });

  // Upload result to a different bucket, to avoid re-triggering this function.
  const blurredBucket = storage.bucket(blurredBucketName);

  // Upload the Blurred image back into the bucket.
  const gcsPath = `gs://${blurredBucketName}/${file.name}`;
  try {
    await blurredBucket.upload(tempLocalPath, {destination: file.name});
    console.log(`Uploaded blurred image to: ${gcsPath}`);
  } catch (err) {
    throw new Error(`Unable to upload blurred image to ${gcsPath}: ${err}`);
  }

  // Delete the temporary file.
  return fs.unlink(tempLocalPath);
};
```

Source: https://cloud.google.com/functions/docs/tutorials/imagemagick#blur_images

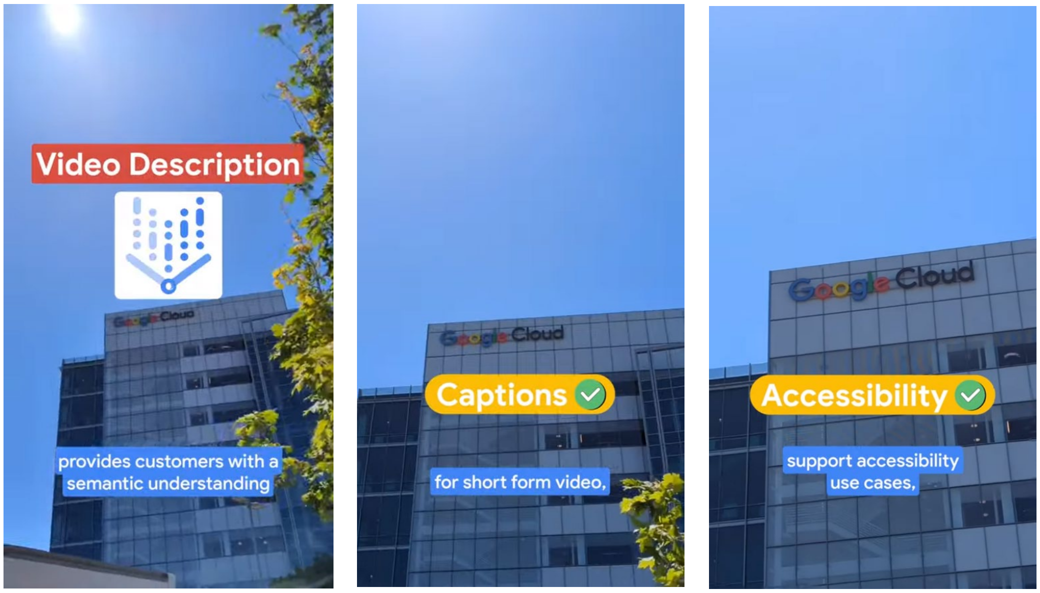
106. In addition, as previously discussed, Video Intelligence via Vertex AI allows users to generate video descriptions, summaries, and captions for a video file.

> There are some other great use cases video description helps unlock. So one of them is, of course, metadata. **Now that you have such detailed information about the video, of course, you have richer metadata, which you can use to power better recommendation engines or searches. Another one is automated captions, similar to the one we saw earlier with images.** So if you have short form videos, like 15 seconds, 30 seconds, you can caption those. Accessibility—you say the tourism app example earlier.
>
> And the next one is understanding and searching the important moments. So this is a really cool one. So for an entire video, if you had this detailed rendering in the form of text, now you could do something with it. **So now that I have the text format, I could search for important terms in it**. So for example, let's imagine we have a six-hour baseball game video, right? And we have its rendering. **Now, I**

> **could search for terms like home run, strike, audience cheering. And then once I have those moments, I can stitch them together into, say, a highlight reel**.[57]

107. As Google explains, Video Intelligence via Vertex AI may place those captions in the video:

> Did you know, Google is the only cloud provider that offers second by second video descriptions with a video-to-text mode? Video Description on Vertex AI provides customers with a semantic understanding of a video in the form of text, generating granular and detailed descriptions for a video. For example, you can generate automated captions for short form video, support accessibility use cases, understand and search important moments in videos, initiate and alert or workflow based on events detected in descriptions, improve ad placement, drive deeper video analytics, and catalog and archive legacy video content. You can even gain deeper insights into your videos and your users who consume them. Like finding out why people stopped watching at a certain time, and identify moments in the video that especially resonated with your users, and why.[58]



Source: https://www.youtube.com/watch?v=zFQrbm0Gk-I at 0:09, 0:18, 0:21.

108. Upon information and belief, Google directly infringes other claims of the ’972 patent as well, including for the reasons discussed in the preceding paragraphs.

[57] https://www.youtube.com/watch?v=F4YKREYPkEI at 8:35 to 9:40 (emphasis added).

[58] https://www.youtube.com/watch?v=zFQrbm0Gk-I at 0:00-0:47.

109. At no time has Google been licensed, either expressly or impliedly, under the ’972 Patent.

110. On information and belief, Google has had knowledge of the ’972 Patent at least through its participation in the 2015 SPROCKIT event at which Vu Digital presented its then patent pending V2D technology. On information and belief, Google additionally has had knowledge of the ’954 Patent family (including the ’972 Patent) since at least as early as November 2, 2022 as evidenced by Google’s identification of the ’954 Patent as potential prior art to the inventions described in Google’s U.S. Patent Application No. 17/423,623.[59] At least as of the filing of this Complaint, Google has had knowledge of the ’972 patent and Google’s infringement thereof.

111. Google’s infringement of the ’972 Patent, which is knowing and willful at least as of the filing of this Complaint, has caused and continues to cause damage to CSI, and CSI is entitled to recover damages sustained as a result of Google’s wrongful acts in an amount subject to proof at trial.

THIRD CLAIM FOR RELIEF

(Infringement of U.S. Patent. No. 11,126,853)

112. CSI realleges and incorporates by reference the allegations of the foregoing paragraphs.

113. On September 21, 2021, the United States Patent and Trademark Office (“USPTO”) duly and legally issued, after a full and fair examination, United States Patent No. 11,126,853 (the “’853 Patent”) entitled “Video to Data” to inventors Bartlett Wade Smith, IV, Allison A. Talley, John Carlos Shields. A true and correct copy of the ’853 Patent is attached as **Exhibit C** to this Complaint.

[59] *See* Google’s November 2, 2022 Information Disclosure Statement with regard to U.S. Patent Application No. 17,423,623 (“**Exhibit D**”).

114. The ’853 Patent was assigned to CSI, which currently holds all substantial rights, title, and interest in and to the ’853 Patent.

115. Pursuant to 35 U.S.C. § 282, the ’853 Patent is presumed valid.

116. The ’853 Patent is presumed to be patent eligible under 35 U.S.C. § 101.

117. The ’853 Patent is directed to an improvement in the functionality of machine learned video recognition and classification systems, particularly with regard to specific techniques for improving the accuracy of object detection and recognition for still images that are extracted from video content. The ’853 Patent is directed to techniques that utilize contextual information such as the arrangement of certain objects in a still image as well as techniques that address image “noise” in order to improve computer visual recognition for video classification systems.

118. The ’853 Patent addresses specific technical challenges that arose in video classification systems when attempting to extract meaningful metadata from video content.

119. The ’853 Patent describes the challenges in the field of computer visual recognition systems and explains the advantages of the claimed inventions:

> [I]dentification of an individual can be a difficult task. For example, facial recognition can become difficult when an individual’s face is obstructed by another object like a football, a baseball helmet, a musical instrument, or other obstructions. An advantage of some embodiments described herein can include the ability to identify an individual without identification of the individual’s face. Embodiments can use contextual information such as association of objects, text, and/or other context within an image or video. As one example, a football placer scores a touchdown but rather than identifying the player using facial recognition, the play can be identified by object recognition of, for example, the player’s team’s logo, text recognition of the player’s jersey number, and by cross referencing this data with that team’s roster (as oppose to another team, which is an example of why the logo recognition can be important). Such embodiments can further learn to identify that player more readily and save his image as data.[60]

120. The ’853 Patent further describes specific challenges encountered when analyzing noisy images and explains how the claimed inventions can be used to address these challenges.

[60] ’853 Patent, 7:55–8:5.

> [I]mages from the video signal can be processed to address, for example, the problem of object noise in a given frame or image. Often images are segmented only to locate and positively identify one or very few main images in the foreground of a given frame. The non-primary or background images are often treated as noise. Nevertheless, these can provide useful information, context, and/or branding for example. . . . For certain specific embodiments, identification of only certain clearly identifiable faces or large unobstructed objects or band logos can be required with all other image noise disregarded or filtered, which can require less computational processing and image database referencing, in turn reducing costs. However, it may become necessary or desirable to detect more detail from a frame or set of frames. In such circumstances, the computational thresholds for identification of an object, face, etc. can be altered according to a then stated need or desire for non-primary, background, obstructed and/or grainy type images.[61]

121. The technological solutions described above are recited in the ’853 Patent claims, including, for example, independent claim 1 (and the corresponding dependent claims).

122. Claim 1 of the ’853 Patent reads as follows:

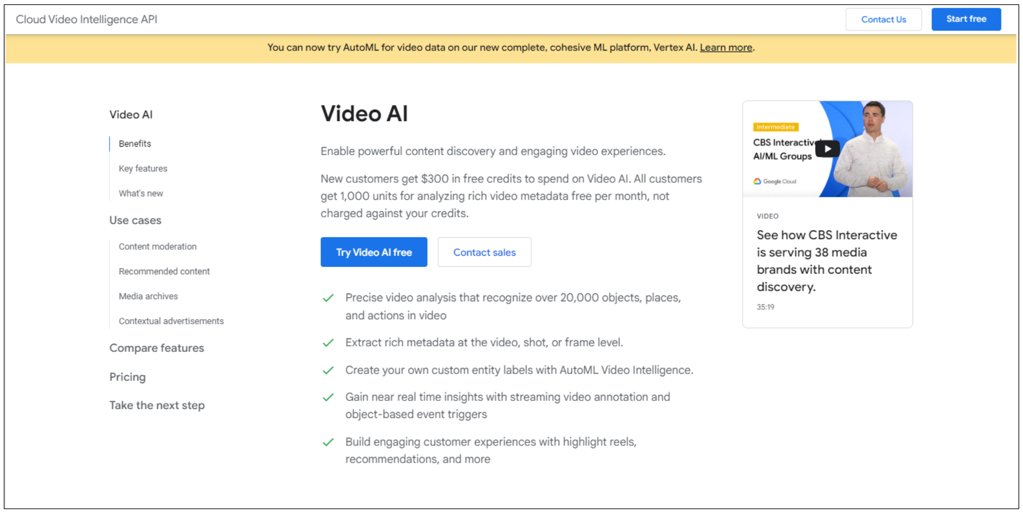
> 1. A system for generating data from a video, comprising:
>
> a coordinator communicatively coupled to a splitter and to a plurality of demultiplexer nodes, wherein the splitter is configured to segment the video, wherein the demultiplexer nodes are configured to extract audio files from the video and to extract still frame images from the video;
>
> an image detector configured to detect an image of an object in the still frame images, wherein the image detector is adjustable to increase detection of non-primary images in the video; and
>
> an object recognizer configured to compare the image of the object to a fractal, wherein the fractal includes a representation of the object based on landmarks associated with the object, wherein the recognizer is further configured to update the fractal with the image.[62]

123. Upon information and belief, Google directly infringes at least claim 1 of the ’853 Patent, at least in the exemplary manner described below.

[61] ’853 Patent, 9:6–26.

[62] ’853 Patent, claim 1.

124. Google directly infringes the ’853 Patent by making, using, offering to sell, and/or selling in the United States its Cloud Video Intelligence product, which relies on ML models and NLP techniques to identify and classify videos using contextual information. Google directs and controls each relevant aspect of the accused technology discussed herein, and benefits from the use of each feature that infringes the ’853 Patent. Additionally, on information and belief, Google uses the ’853 patented inventions and the Accused Product to classify videos uploaded to its online video sharing platform, YouTube (www.youtube.com), and to its Google Photos iOS and Android apps.



Cloud Video Intelligence API

You can now try AutoML for video data on our new complete, cohesive ML platform, Vertex AI. Learn more.

Video AI

Enable powerful content discovery and engaging video experiences.

New customers get $300 in free credits to spend on Video AI. All customers get 1,000 units for analyzing rich video metadata free per month, not charged against your credits.

- ✓ Precise video analysis that recognize over 20,000 objects, places, and actions in video
- ✓ Extract rich metadata at the video, shot, or frame level.
- ✓ Create your own custom entity labels with AutoML Video Intelligence.
- ✓ Gain near real time insights with streaming video annotation and object-based event triggers
- ✓ Build engaging customer experiences with highlight reels, recommendations, and more

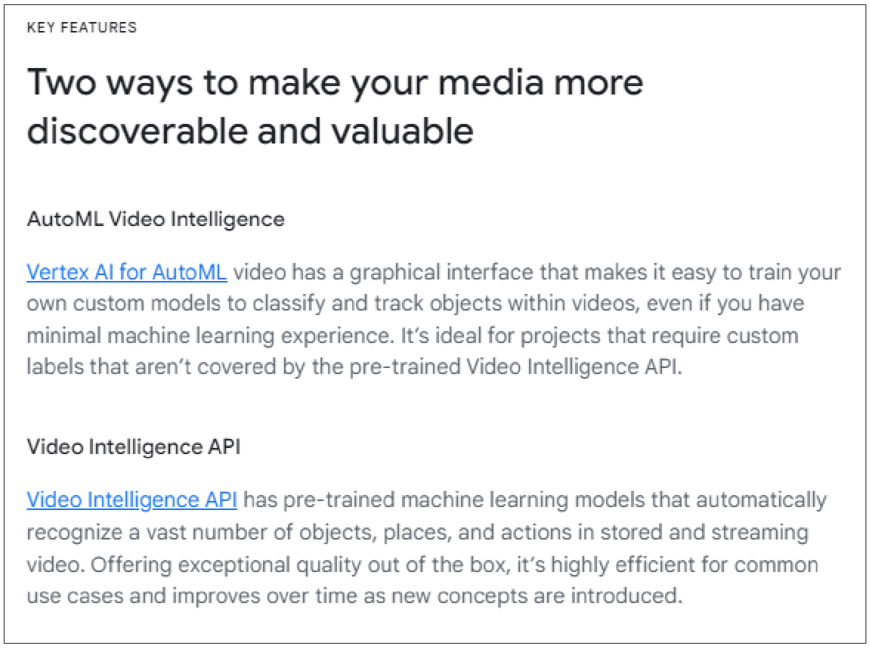
See how CBS Interactive is serving 38 media brands with content discovery.

Source: https://cloud.google.com/video-intelligence.

125. According to Google, Video Intelligence may be used with custom (via “Vertex AI for AutoML”)[63] or pretrained (“Video Intelligence API”) machine learning (“ML”) models in order to classify and annotate videos so that video data may be more easily parsed and to enable users to implement, for example, contextual-based advertising and/or content recommendation in their video content libraries.[64]

[63] As discussed above, AutoML has been replaced and/or consolidated with Google’s “Vertex AI” product, which continues to provide this capability to customers. *See*, *e.g.*, https://cloud.google.com/video-intelligence/automl/pricing (last accessed March 4, 2024). As Google explains “[a]ll of the functionality of legacy AutoML Video Intelligence and new features are available on the Vertex AI platform.” *Id*.

[64] *See* https://cloud.google.com/video-intelligence.



KEY FEATURES

Two ways to make your media more discoverable and valuable

AutoML Video Intelligence

Vertex AI for AutoML video has a graphical interface that makes it easy to train your own custom models to classify and track objects within videos, even if you have minimal machine learning experience. It's ideal for projects that require custom labels that aren't covered by the pre-trained Video Intelligence API.

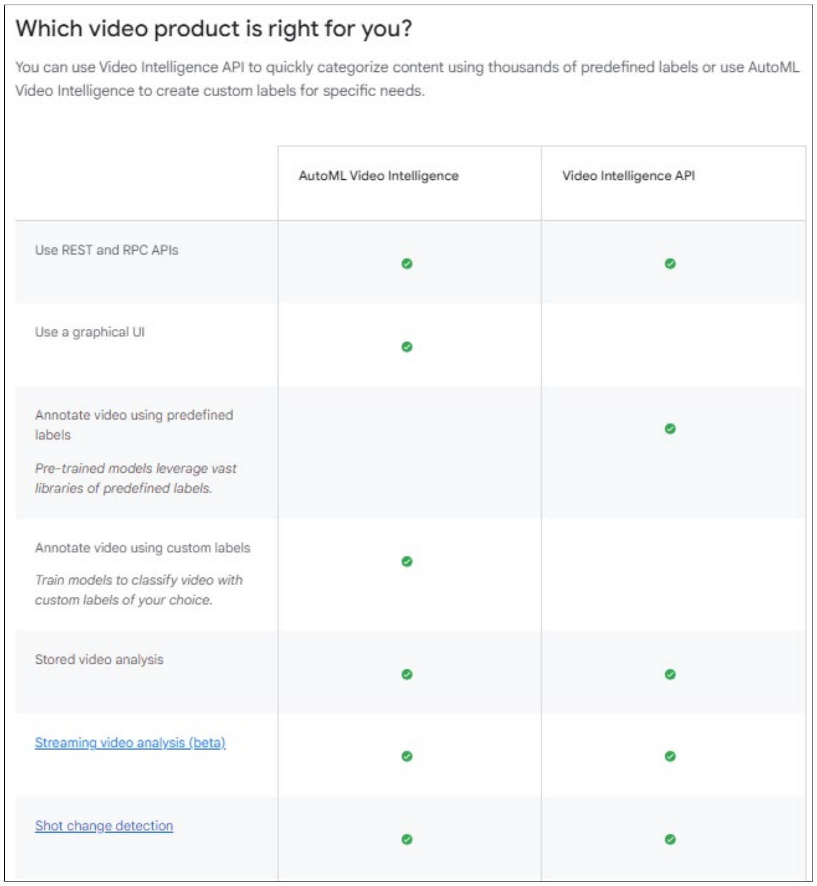
Video Intelligence API

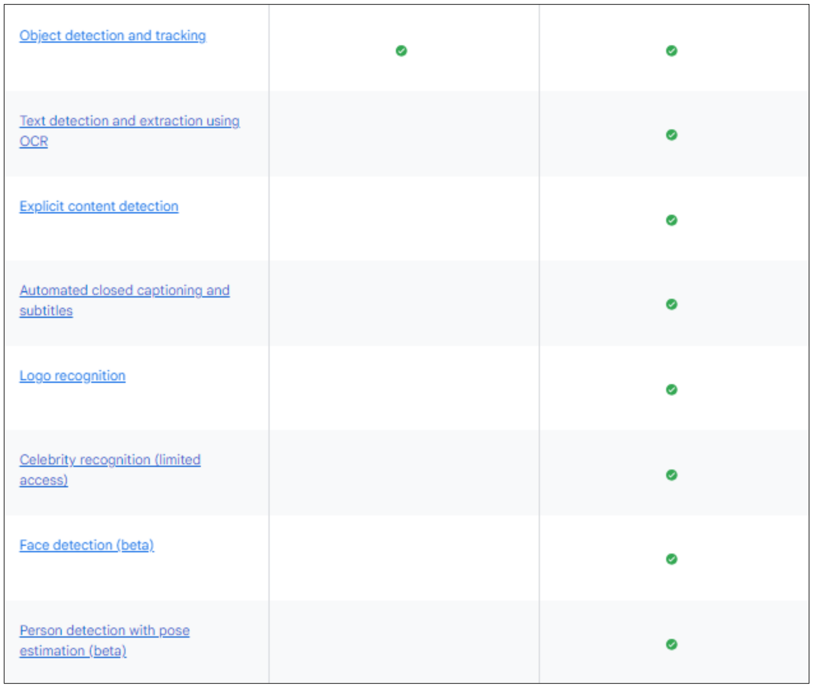
Video Intelligence API has pre-trained machine learning models that automatically recognize a vast number of objects, places, and actions in stored and streaming video. Offering exceptional quality out of the box, it's highly efficient for common use cases and improves over time as new concepts are introduced.

Source: https://cloud.google.com/video-intelligence

126. Video Intelligence may be used to add custom or predefined annotations for both stored video content library as well as for live-streaming video applications.[65]

[65] *Id.*



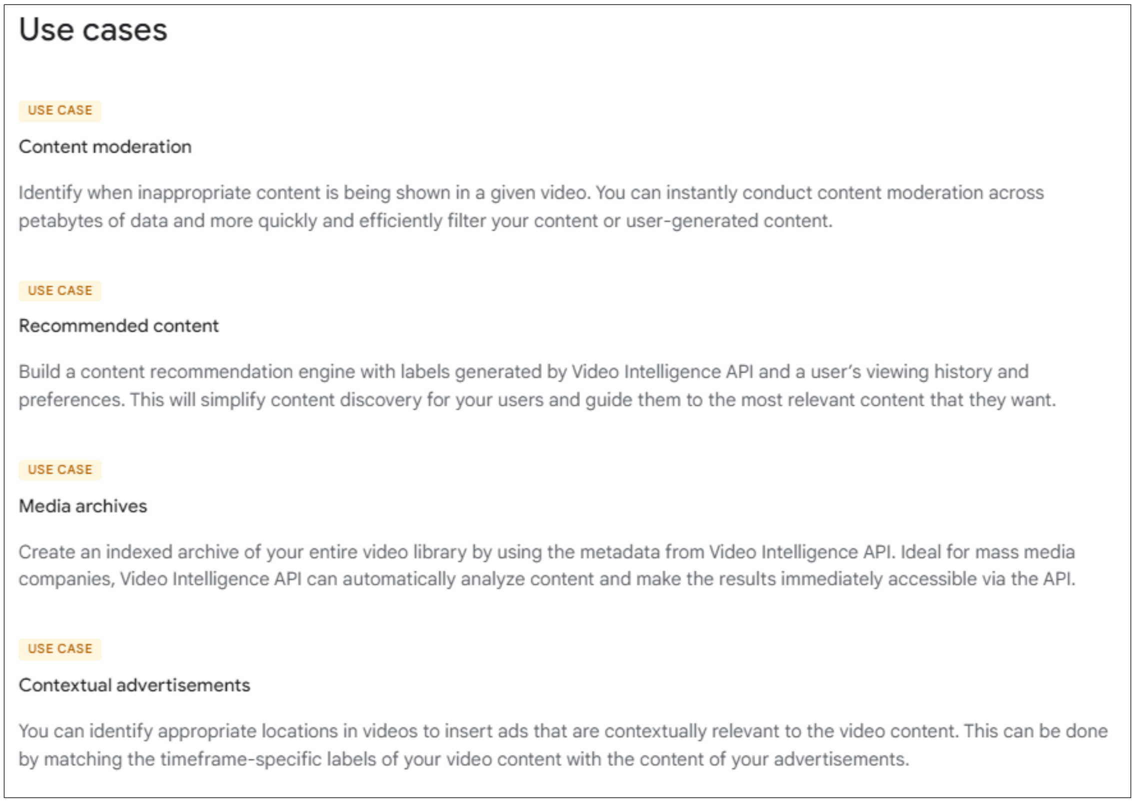


Which video product is right for you?

You can use Video Intelligence API to quickly categorize content using thousands of predefined labels or use AutoML Video Intelligence to create custom labels for specific needs.

	AutoML Video Intelligence	Video Intelligence API
Use REST and RPC APIs	✓	✓
Use a graphical UI	✓	
Annotate video using predefined labels *Pre-trained models leverage vast libraries of predefined labels.*		✓
Annotate video using custom labels *Train models to classify video with custom labels of your choice.*	✓	
Stored video analysis	✓	✓
Streaming video analysis (beta)	✓	✓
Shot change detection	✓	✓
Object detection and tracking	✓	✓
Text detection and extraction using OCR		✓
Explicit content detection		✓
Automated closed captioning and subtitles		✓
Logo recognition		✓
Celebrity recognition (limited access)		✓
Face detection (beta)		✓
Person detection with pose estimation (beta)		✓

Source: https://cloud.google.com/video-intelligence



Use cases

USE CASE

Content moderation

Identify when inappropriate content is being shown in a given video. You can instantly conduct content moderation across petabytes of data and more quickly and efficiently filter your content or user-generated content.

USE CASE

Recommended content

Build a content recommendation engine with labels generated by Video Intelligence API and a user's viewing history and preferences. This will simplify content discovery for your users and guide them to the most relevant content that they want.

USE CASE

Media archives

Create an indexed archive of your entire video library by using the metadata from Video Intelligence API. Ideal for mass media companies, Video Intelligence API can automatically analyze content and make the results immediately accessible via the API.

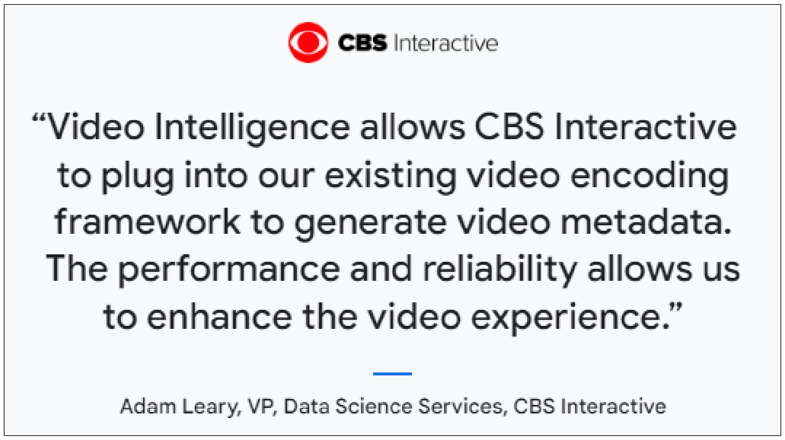
USE CASE

Contextual advertisements

You can identify appropriate locations in videos to insert ads that are contextually relevant to the video content. This can be done by matching the timeframe-specific labels of your video content with the content of your advertisements.

Source: https://cloud.google.com/video-intelligence

127. On information and belief, Google and its customers use Video Intelligence to generate video metadata using the context-based techniques of the '853 Patent in order to "enhance the video experience."



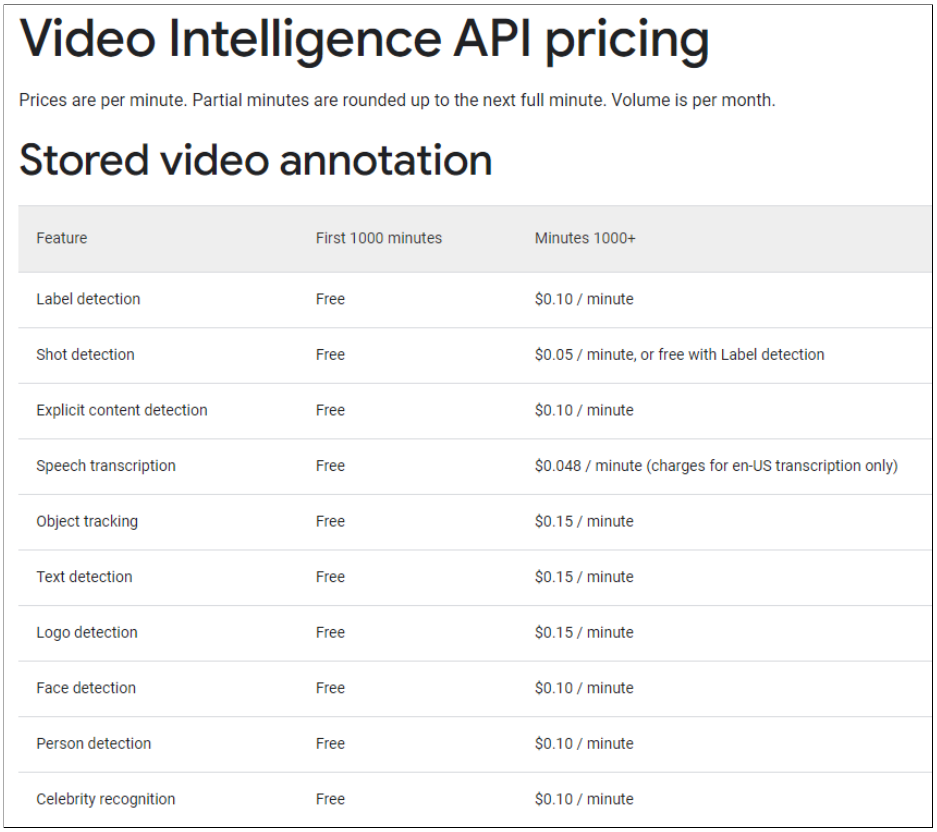
CBS Interactive

"Video Intelligence allows CBS Interactive to plug into our existing video encoding framework to generate video metadata. The performance and reliability allows us to enhance the video experience."

Adam Leary, VP, Data Science Services, CBS Interactive

Source: https://cloud.google.com/video-intelligence

128. Google provides the Video Intelligence product to users through Google Cloud Platform ("GCP") and charges users for the amount of resources (*e.g.*, cloud server virtual CPUs) consumed for video processing on a per minute basis.



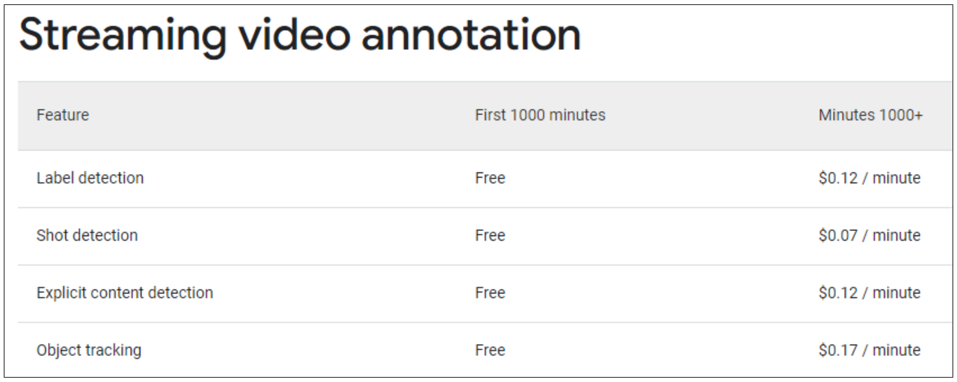
Video Intelligence API pricing

Prices are per minute. Partial minutes are rounded up to the next full minute. Volume is per month.

Stored video annotation

Feature	First 1000 minutes	Minutes 1000+
Label detection	Free	$0.10 / minute
Shot detection	Free	$0.05 / minute, or free with Label detection
Explicit content detection	Free	$0.10 / minute
Speech transcription	Free	$0.048 / minute (charges for en-US transcription only)
Object tracking	Free	$0.15 / minute
Text detection	Free	$0.15 / minute
Logo detection	Free	$0.15 / minute
Face detection	Free	$0.10 / minute
Person detection	Free	$0.10 / minute
Celebrity recognition	Free	$0.10 / minute

Source: https://cloud.google.com/video-intelligence/pricing



Streaming video annotation

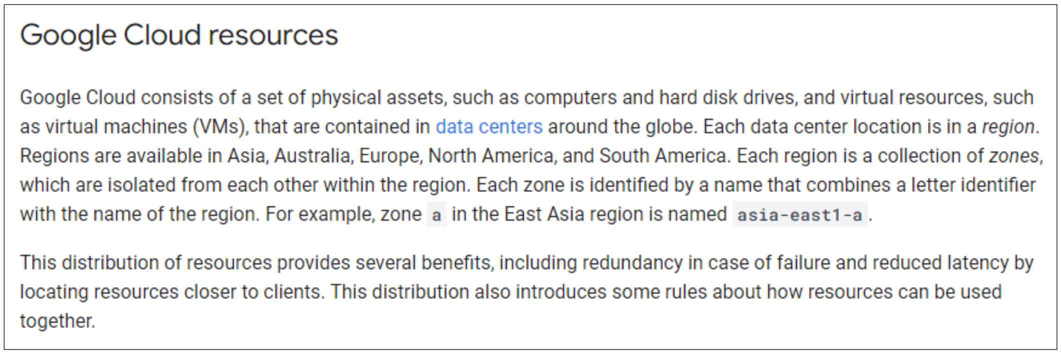
Feature	First 1000 minutes	Minutes 1000+
Label detection	Free	$0.12 / minute
Shot detection	Free	$0.07 / minute
Explicit content detection	Free	$0.12 / minute
Object tracking	Free	$0.17 / minute

Source: https://cloud.google.com/video-intelligence/pricing#streaming_video_annotation

129. With regard to claim 1 of the '853 Patent, Google makes and/or uses a system for generating data from a video comprising "a coordinator communicatively coupled to a splitter and to a plurality of demultiplexer nodes, wherein the splitter is configured to segment the video,

wherein the demultiplexer nodes are configured to extract audio files from the video and to extract still frame images from the video."

130. For example, on information and belief, Video Intelligence requires use of Google's Cloud Platform ("GCP"), a distributed network of "cloud" servers, and Google Cloud Storage in order to perform the relevant video processing functions.[66]

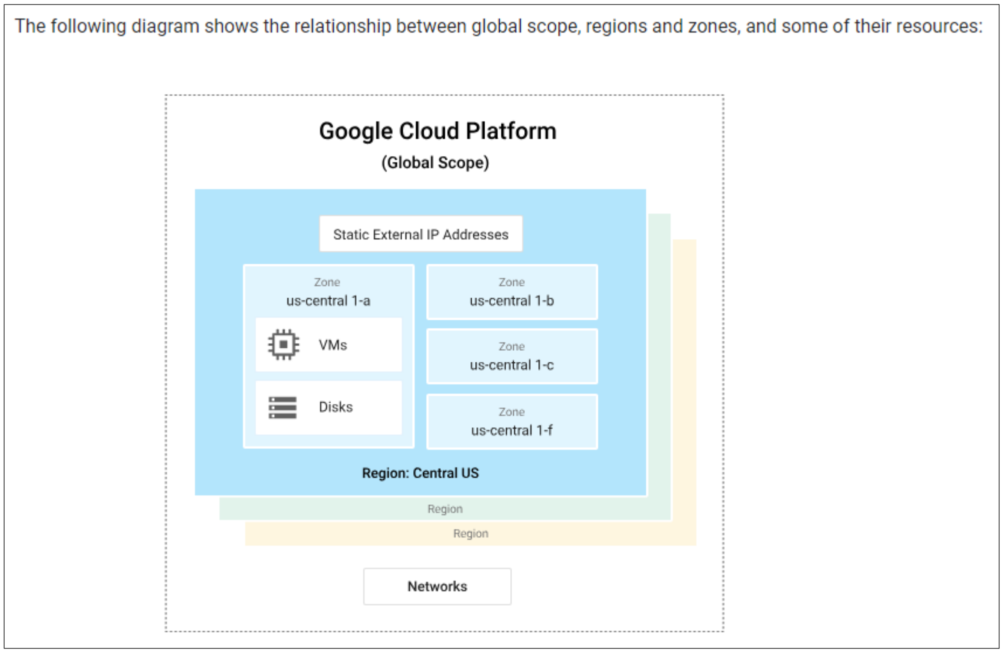


Google Cloud resources

Google Cloud consists of a set of physical assets, such as computers and hard disk drives, and virtual resources, such as virtual machines (VMs), that are contained in data centers around the globe. Each data center location is in a *region*. Regions are available in Asia, Australia, Europe, North America, and South America. Each region is a collection of *zones*, which are isolated from each other within the region. Each zone is identified by a name that combines a letter identifier with the name of the region. For example, zone `a` in the East Asia region is named `asia-east1-a`.

This distribution of resources provides several benefits, including redundancy in case of failure and reduced latency by locating resources closer to clients. This distribution also introduces some rules about how resources can be used together.

Source: https://cloud.google.com/docs/overview

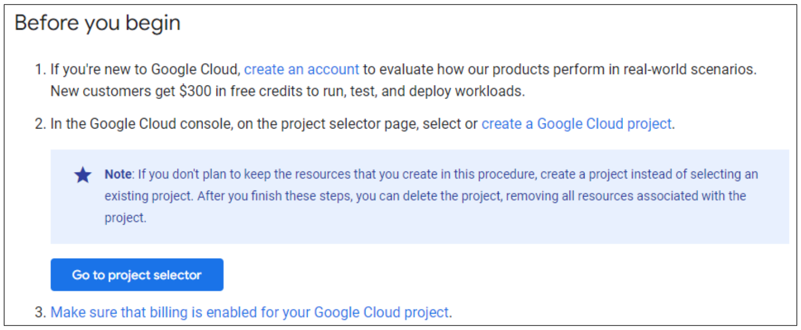


The following diagram shows the relationship between global scope, regions and zones, and some of their resources:

Source: https://cloud.google.com/docs/overview

[66] *See* https://cloud.google.com/video-intelligence?hl=en ("From Video Intelligence's API's monthly free minutes to AutoML Video Intelligence's customizable services, **Google Cloud's video intelligence products** offer pricing options that work with your needs." (emphasis added)).

131. Google also explains that a GCP account and project are required in order to use the Video Intelligence API.



Before you begin

1. If you're new to Google Cloud, create an account to evaluate how our products perform in real-world scenarios. New customers get $300 in free credits to run, test, and deploy workloads.
2. In the Google Cloud console, on the project selector page, select or create a Google Cloud project.

★ **Note**: If you don't plan to keep the resources that you create in this procedure, create a project instead of selecting an existing project. After you finish these steps, you can delete the project, removing all resources associated with the project.

3. Make sure that billing is enabled for your Google Cloud project.

Source: https://cloud.google.com/video-intelligence/docs/annotate-video-command-line

132. In addition, Google explains that the Video Intelligence system begins by segmenting a video and extracting audio and image data from the segmented video.

> So given a video in your GCS [Google Cloud Storage] bucket, **the first thing that we do is we decode the video**. And usually a video container contains like **audio streams, video streams, subtitles and all of the metadata**.[67]

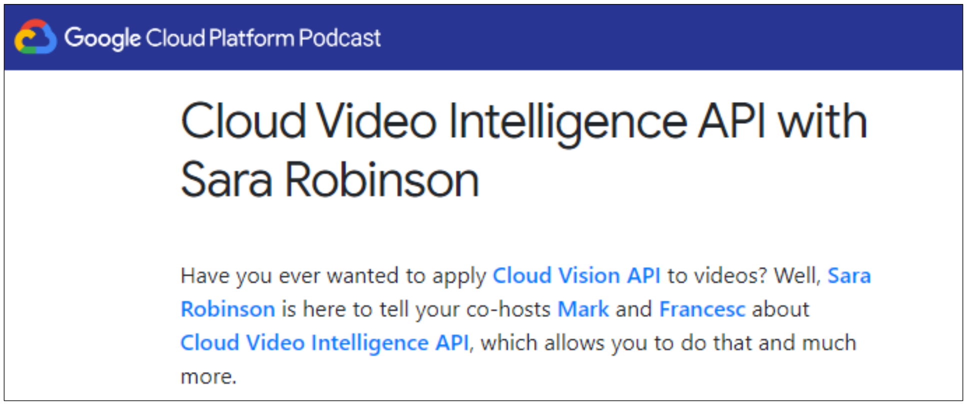
133. Google explains that Video Intelligence extracts still images from the segments, shots, or frames of extracted video data stream.

> And we are first only interested in the video stream. And a video stream is often a sequence of images . . . a lot of them. For a five-minute video at 24 frames per second, you will have like 7,000, 8,000 images in them. And they will probably in sequence look a lot similar. So classifying each of those is not too efficient, let's say. **So we shot sample the images, and say at one FPS. So one frame per second, we apply an image classifier. An image classifier that is very similar to—that you can get through the Google Cloud Vision API.** And we do it for every image in the entire video. And for each frame, you will get image features such as this elephant with 97% accuracy, or animal, or river in that case. And, if you keep going on, you will get other labels for crocodiles and lions.[68]

[67] https://www.youtube.com/watch?v=y-k8oelbmGc at 21:00 to 22:14 (emphasis added).

[68] https://www.youtube.com/watch?v=y-k8oelbmGc at 21:00 to 22:14 (emphasis added).

134. The Accused Product further comprises "an image detector configured to detect an image of an object in the still frame images, wherein the image detector is adjustable to increase detection of non-primary images in the video." For example, as discussed above, the Accused Product comprises an "image classifier."[69] Google further explains that the Video Intelligence image classifier is essentially the same system that Google employs for its Cloud Vision product.[70]



Cloud Video Intelligence API with Sara Robinson

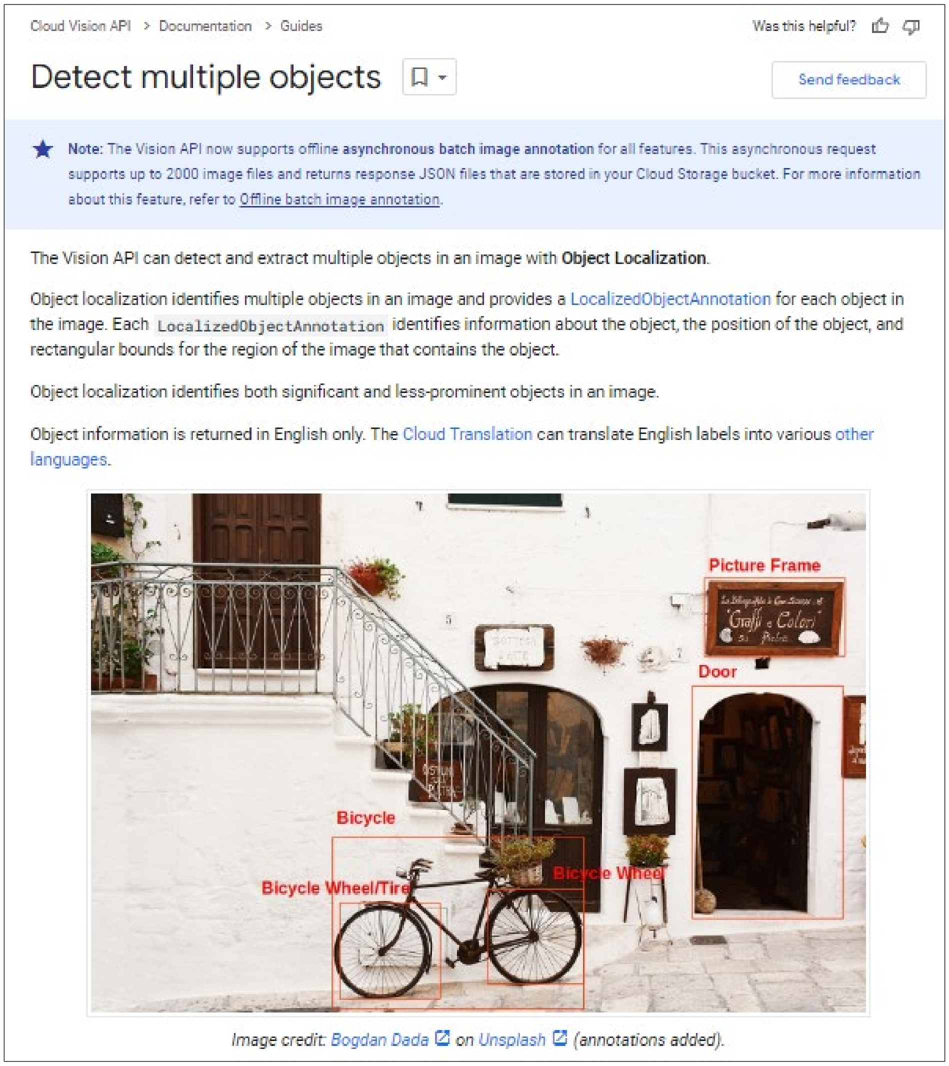
Have you ever wanted to apply Cloud Vision API to videos? Well, Sara Robinson is here to tell your co-hosts Mark and Francesc about Cloud Video Intelligence API, which allows you to do that and much more.

Source: https://www.gcppodcast.com/post/episode-74-video-intelligence-api-with-sara-robinson/.

[69] *Id.*

[70] *Id.*

135. On information and belief, both the Cloud Vision API and Video Intelligence employ an image classifier capable of detecting multiple objects, including primary and non-primary objects.[71]



Cloud Vision API > Documentation > Guides

Was this helpful?

Detect multiple objects

Send feedback

★ **Note:** The Vision API now supports offline **asynchronous batch image annotation** for all features. This asynchronous request supports up to 2000 image files and returns response JSON files that are stored in your Cloud Storage bucket. For more information about this feature, refer to Offline batch image annotation.

The Vision API can detect and extract multiple objects in an image with **Object Localization**.

Object localization identifies multiple objects in an image and provides a LocalizedObjectAnnotation for each object in the image. Each `LocalizedObjectAnnotation` identifies information about the object, the position of the object, and rectangular bounds for the region of the image that contains the object.

Object localization identifies both significant and less-prominent objects in an image.

Object information is returned in English only. The Cloud Translation can translate English labels into various other languages.

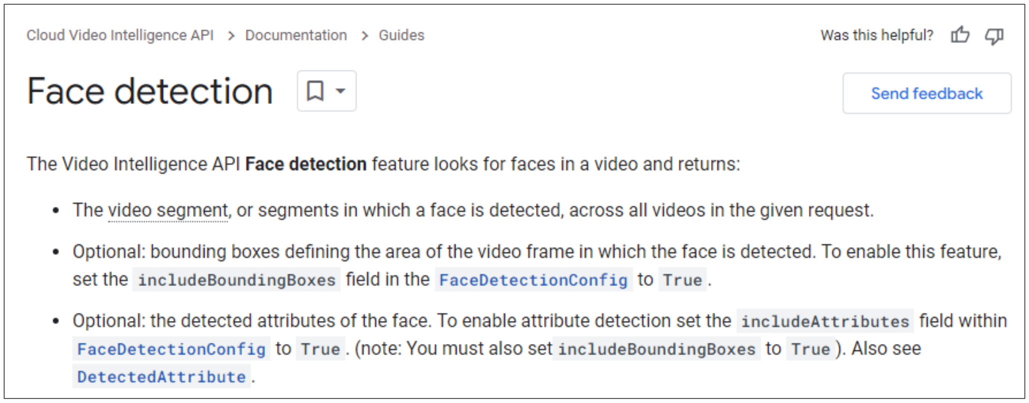
Image credit: Bogdan Dada on Unsplash (annotations added).

Source: https://cloud.google.com/vision/docs/object-localizer (last accessed March 4, 2024).

136. The Accused Product further comprises "an object recognizer configured to compare the image of the object to a fractal, wherein the fractal includes a representation of the object based on landmarks associated with the object, wherein the recognizer is further configured

[71] *See id.*

to update the fractal with the image." For example, Video Intelligence may be used to perform facial recognition for images extracted from a video using detected attributes of a face.

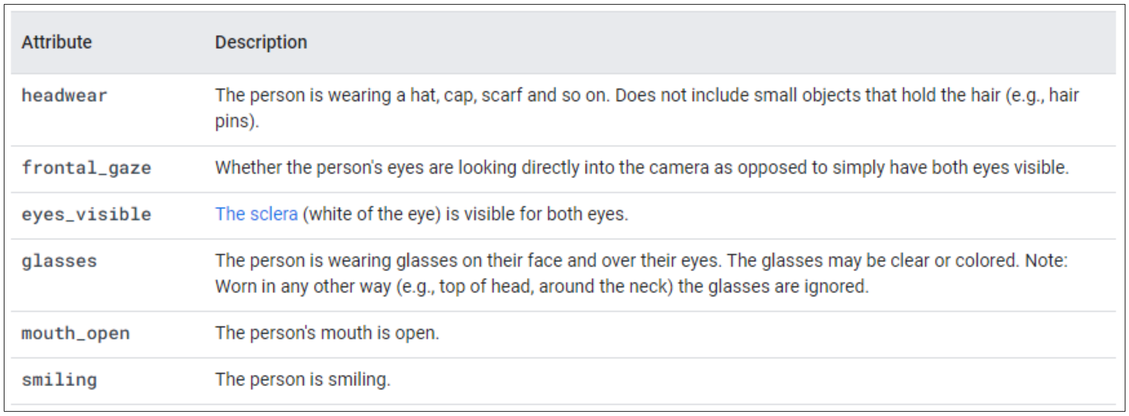


Face detection

The Video Intelligence API **Face detection** feature looks for faces in a video and returns:

- The video segment, or segments in which a face is detected, across all videos in the given request.
- Optional: bounding boxes defining the area of the video frame in which the face is detected. To enable this feature, set the `includeBoundingBoxes` field in the `FaceDetectionConfig` to `True`.
- Optional: the detected attributes of the face. To enable attribute detection set the `includeAttributes` field within `FaceDetectionConfig` to `True`. (note: You must also set `includeBoundingBoxes` to `True`). Also see `DetectedAttribute`.

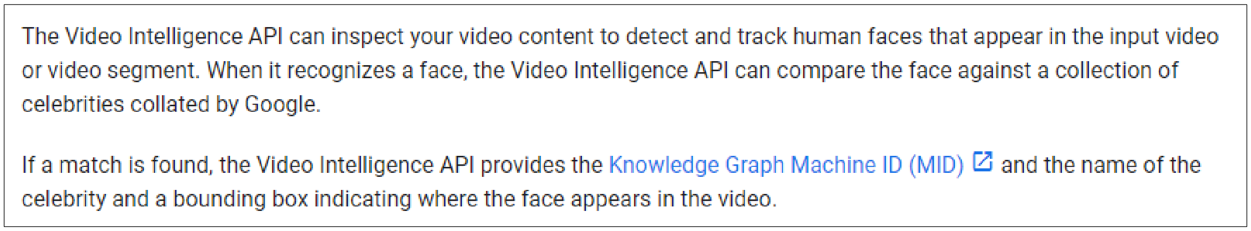
Source: https://cloud.google.com/video-intelligence/docs/feature-face-detection (last accessed March 4, 2024).



Attribute	Description
`headwear`	The person is wearing a hat, cap, scarf and so on. Does not include small objects that hold the hair (e.g., hair pins).
`frontal_gaze`	Whether the person's eyes are looking directly into the camera as opposed to simply have both eyes visible.
`eyes_visible`	The sclera (white of the eye) is visible for both eyes.
`glasses`	The person is wearing glasses on their face and over their eyes. The glasses may be clear or colored. Note: Worn in any other way (e.g., top of head, around the neck) the glasses are ignored.
`mouth_open`	The person's mouth is open.
`smiling`	The person is smiling.

Source: https://cloud.google.com/video-intelligence/docs/feature-face-detection (last accessed March 4, 2024).

137. As another example, Google explains that it maintains a curated collection of image data for celebrities that can be used for facial recognition with Video Intelligence.



The Video Intelligence API can inspect your video content to detect and track human faces that appear in the input video or video segment. When it recognizes a face, the Video Intelligence API can compare the face against a collection of celebrities collated by Google.

If a match is found, the Video Intelligence API provides the Knowledge Graph Machine ID (MID) ↗ and the name of the celebrity and a bounding box indicating where the face appears in the video.

Source: https://cloud.google.com/video-intelligence/docs/celebrity-recognition (last accessed March 4, 2024).

138. In addition, Google explains that Video Intelligence is trained and updated using data extracted from millions of videos in order to improve object recognition accuracy.

> Another similar example would be if you have a video of people in costumes or pumpkins and candies. This classifier will be able to tell you , hey, this is a video about Halloween. And this is only possible because we have trained this model on millions of YouTube videos, and human rated and also like-verified. Ram talked about also the shot level annotation a minute ago. And those are essentially treating each shot as a mini video, and then running with the same logic.[72]

Google further explains that it is "constantly" updating the models for Video Intelligence, YouTube, and Google Photos.

> And one thing I wanted to emphasize is with this release, we are not keeping the models as is. **We are constantly improving the model, especially when Google internally releases a new model for YouTube, Photos, then we also update the models for Google Cloud Video API**.[73]

139. Upon information and belief, Google directly infringes other claims of the '853 patent as well, including for the reasons discussed in the preceding paragraphs.

140. At all relevant times, CSI has complied with the marking requirement under 35 U.S.C. § 287 with respect to the '853 Patent.

141. At no time has Google been licensed, either expressly or impliedly, under the '853 Patent.

142. At least as of the filing of this Complaint, Google has had knowledge of the '853 Patent and Google's infringement thereof.

143. Google's infringement of the '853 Patent, which is knowing and willful at least as of the filing of this Complaint, has caused and continues to cause damage to CSI, and CSI is entitled to recover damages sustained as a result of Google's wrongful acts in an amount subject to proof at trial.

[72] https://www.youtube.com/watch?v=y-k8oelbmGc at 22:38 to 23:21 (last accessed March 4, 2024).

[73] *Id*. at 24:07 to 24:32 (cleaned up; emphasis added).

PRAYER FOR RELIEF

CSI respectfully requests that judgment be entered in its favor against Google as follows:

1. Entering judgment declaring that Google has directly infringed one or more claims of the Patents-in-Suit, in violation of 35 U.S.C. § 271;
2. Entering judgment permanently enjoining Google, its officers, agents, subsidiaries, parents, employees, and those in privity or in active concert with it from further activities that constitute infringement of the Patents-in-Suit;
3. Entering judgment ordering that CSI be awarded damages in an amount no less than a reasonable royalty for each asserted patent arising out of Google's infringement of the Patents-in-Suit, together with any other monetary amounts recoverable by CSI;
4. Entering judgment declaring that this an exceptional case under 35 U.S.C. § 285 and awarding CSI its attorneys' fees; and
5. Awarding CSI court costs and such other and further relief as the Court deems just and proper under the circumstances.

DEMAND FOR JURY TRIAL

Pursuant to Rule 38 of the Federal Rules of Civil Procedure, CSI demands a trial by jury on all issues so triable.

DATED: May 9, 2024

Respectfully submitted,

/s/ *Robert S. Hill*
Robert S. Hill, TX Bar No. 24050764
robert.hill@hklaw.com
David C. Schulte, TX Bar No. 24037456
david.schulte@hklaw.com
Sara S. Staha, TX Bar No. 24088368
Sara.staha@hklaw.com
Morgan J. Delabar, TX Bar No. 24116625
morgan.delabar@hklaw.com
HOLLAND & KNIGHT LLP
1722 Routh St., Suite 1500
Dallas, Texas 75201
Telephone: (214) 964-9500
Facsimile: (214) 964-9501

Robert K. Jain, TX Bar No. 24139315
(*admission pending*)
robert.jain@hklaw.com
HOLLAND & KNIGHT LLP
98 San Jacinto Boulevard, Suite 1900
Austin, Texas 78701
Telephone: (512) 472-1081
Facsimile: (512) 472-7473

Jacob W. Schneider, (*pro hac vice forthcoming*)
MA Bar No. 675315
jacob.schneider@hklaw.com
Allison M. Lucier, (*pro hac vice forthcoming*)
MA Bar No. 569193
allison.lucier@hklaw.com
HOLLAND & KNIGHT LLP
10th St. James Avenue, 11th Floor
Boston, MA 02116
Telephone: (617) 523-2700
Facsimile: (617) 523-6850

Anthony J. Fuga, (*pro hac vice forthcoming*)
IL Bar No. 6301658
anthony.fuga@hklaw.com
Tiffany Lee, (*pro hac vice forthcoming*)
IL Bar No. 6342343
tiffany.lee@hklaw.com
HOLLAND & KNIGHT LLP
150 North Riverside Plaza, Suite 2700

Chicago, IL 60606
Telephone: (312) 263-3600
Facsimile: (312) 578-6666

Deron R. Dacus, TX Bar No. 00790553
ddacus@dacusfirm.com
THE DACUS FIRM, P.C.
821 E. SE Loop 323, Suite 430
Tyler, TX 75701
Telephone: (903) 705-7233
Facsimile: (903) 581-2543

ATTORNEYS FOR PLAINTIFF CELLULAR SOUTH, INC.