Course name: Developing Applications with Google Cloud
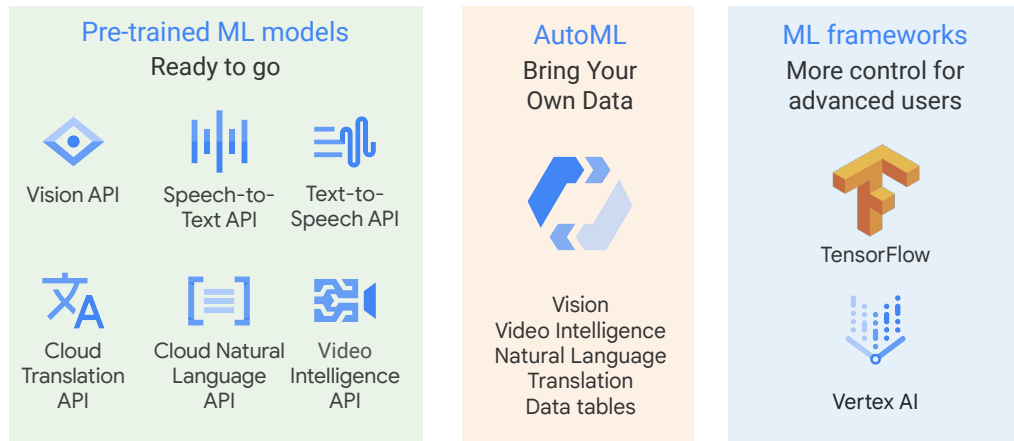**Module name:** Adding Intelligence to Your Application

Featured products: Vision API, Speech-to-Text and Text-to-Speech APIs, Video Intelligence API, Translation API, Natural Language API

Machine learning is about teaching machines to recognize patterns like humans do. While even a two-year-old can easily distinguish between an apple and an orange, it is very difficult to teach a computer to do the same thing. Google has developed pre-trained machine learning models and made them available to you as easy-to-use Google Cloud APIs. Now, with just a few lines of code, you can add artificial intelligence or AI to your own application. In this module, we'll explore Google's pre-trained machine learning APIs for vision, speech, video intelligence, and natural language processing.

Use pre-trained machine learning (ML) models to add intelligence to your applications

Google Cloud offers several pre-trained machine-learning (ML) models that you can use to add intelligence to your application.

- The Vision API enables you to perform complex image detection. For more information, see https://cloud.google.com/vision/.
- The Speech-to-Text and Text-to-Speech APIs enable developers to convert audio to text. It handles 110 languages and variants to support your global user base. You can transcribe the text of users dictating to an application's microphone, enable command-and-control through voice, transcribe audio files, and more. For more information, see https://cloud.google.com/speech/.
- The Cloud Translation API enables you to translate an arbitrary string into any supported language. The Cloud Translation API is highly responsive. Websites and applications can use the Cloud Translation API for fast, dynamic translation of text from a source language to a target language (e.g., Japanese to English). For more information, see https://cloud.google.com/translate/.
- The Cloud Natural Language API enables you to extract information about entities such as people, places, and events that are mentioned in text documents, news articles, or blog posts. You can use the API to understand sentiment about your product on social media or parse intent from customer conversations. For more information, see https://cloud.google.com/natural-language/.
- The Video Intelligence API enables you to search every moment of every video file to extract and understand the video's entities at the shot, frame, or video level. The API annotates videos stored in Google Cloud Storage and helps you identify key noun entities of your video and when they occur within

- the video. For more information, see
  https://cloud.google.com/video-intelligence/.

AutoML is a suite of ML products that enables users with limited ML expertise to train high-quality models specific to their business needs. AutoML leverages more than 10 years of proprietary Google Research technology to help users' ML models achieve faster performance and more accurate predictions. For more information, see https://cloud.google.com/automl/.

You can also use your own data to build and train your own ML models by using TensorFlow and Vertex AI. For more information, see https://www.tensorflow.org/ and https://cloud.google.com/vertex-ai.

Invoke REST APIs to use machine learning APIs; no machine learning knowledge is required

Invoke Vision API

The Vision API can work off an image in Cloud Storage or embedded directly into a POST message. I'll use C

请您爱护和保护卫生创建优美水环境

. That photograph is from http://www.publicdomainpictures.net/view-image.php?image=15842

Image (GCS/embedded)

JSON request

```
# Running Vision API
import base64
IMAGE="gs://cloud-training-demos/vision/sign2.jpg"
vservice = build('vision', 'v1', developerKey=APIKEY)
request = vservice.images().annotate(body={
        'requests': [{
                'image': {
                        'source': {
                                'gcs_image_uri': IMAGE
                        }
                },
                'features': [{
                        'type': 'TEXT_DETECTION',
                        'maxResults': 3,
                }]
        }],
})
responses = request.execute(num_retries=3)
print responses

{u'responses': [{u'textAnnotations': [{u'locale': u'zh', u'description': u'\u8bf7\u6
```

JSON response

Google Cloud

It's really easy to invoke the REST APIs to implement machine learning in your application, no ML knowledge is required. Let's take a look at this example. In this example, we are using the Vision API to process an image that's stored in Cloud Storage. We invoke the REST API and send it a JSON request, and we receive a JSON response with attributes that describe the image, as simple as that.

Let's take a look at a few examples now.

Analyze images

Label detection
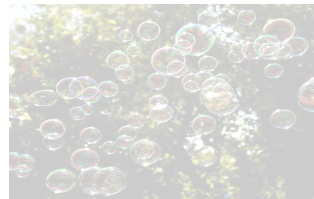

Optical character recognition (OCR)

Landmark detection
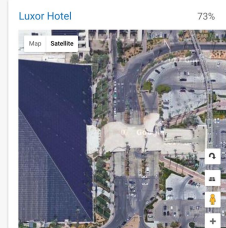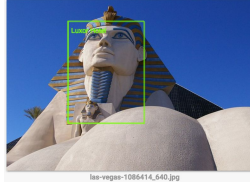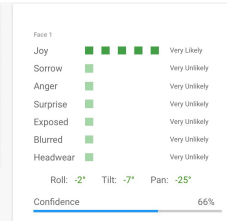

Logo detection

Face detection

Explicit content detection

Google Cloud

The Vision API can categorize objects under labels and perform optical character recognition (OCR). The Vision API can detect landmarks, logos, faces, and explicit content.

Get insight from images

Google Cloud

For example, Vision API can analyze faces and return information about emotions and headwear. In the wedding picture, the API accurately returns the emotional expressions on the faces in the picture.

In the picture of the Sphinx, Vision API correctly detects that the image is from the Sphinx in Las Vegas and not the Sphinx in Egypt.

Let your users talk to you

Google Cloud
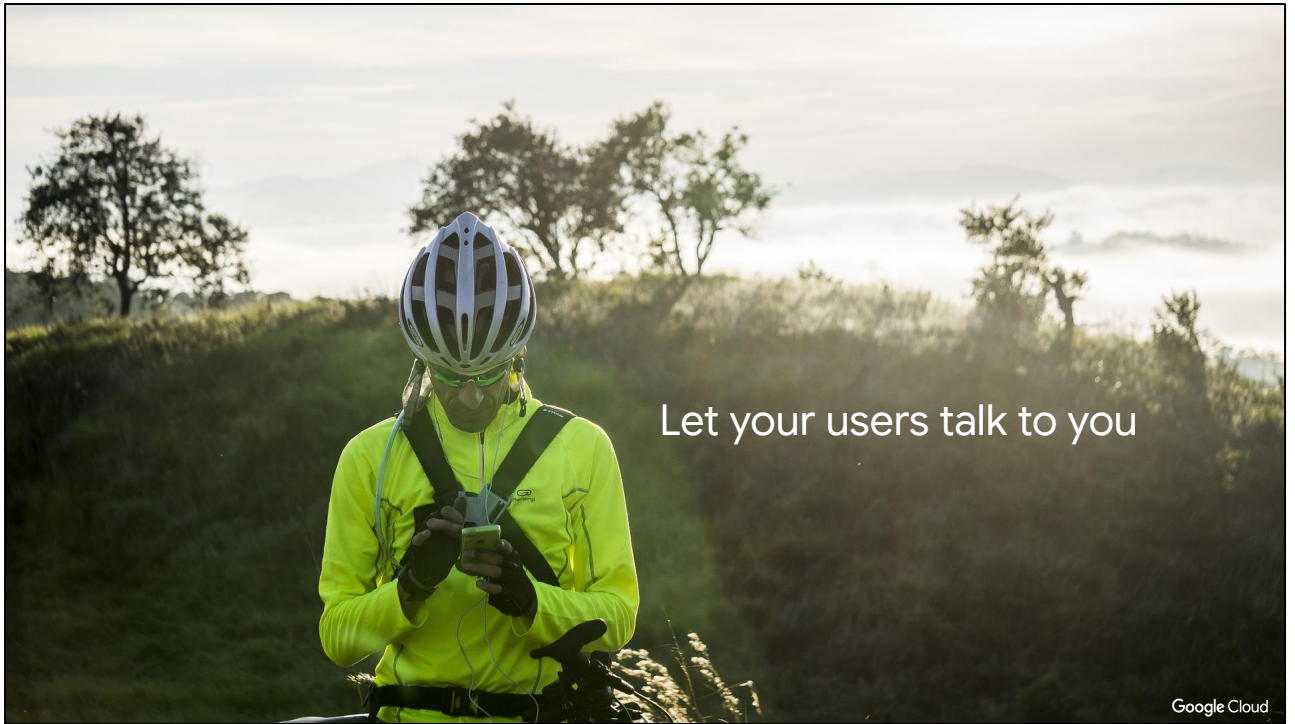
Speech API enables developers to convert audio to text. It handles 110 languages and variants to support your global user base. You can transcribe the text of users dictating to an application's microphone, enable command-and-control through voice, transcribe audio files, and more.

Meeting Nanny: Use images to take action
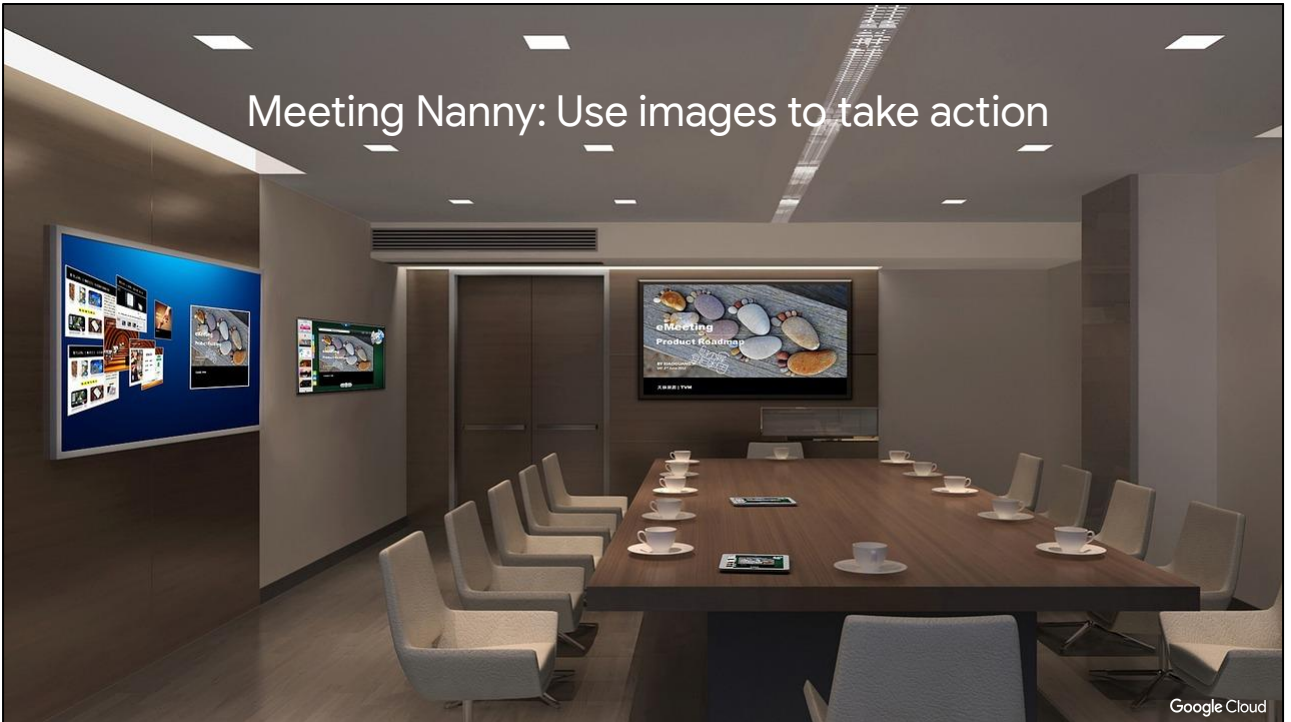
Google's conference room systems perform occupancy detection by using motion detection with the VC camera and by call ID matching. Every 30 seconds, the VC unit sends a Cloud Pub/Sub notification indicating whether motion was detected or not. It also sends a Cloud Pub/Sub notification when a call starts or ends.

If motion is detected between 6 and 8 minutes after the meeting start time, the room counts as occupied. Otherwise, it's empty.

Quiz

You are developing an application that tags all surveillance video before storing the files. Which API should you use?

A. Vision API

B. Speech-to-Text API

C. Video Intelligence API

D. Translation API

You are developing an application that will tag all surveillance video before storing the files. Which API should you use?

a. Vision API
b. Speech-to-Text API
c. Video Intelligence API
d. Translation API

Quiz

You are developing an application that tags all surveillance video before storing the files. Which API should you use?

A. Vision API

B. Speech-to-Text API

C. Video Intelligence API

D. Translation API

If you are developing an application that will tag all surveillance video before storing the files, you should use the Video Intelligence API. (C)