Face and Voice Recognition

Using Microsoft Cognitive Services

Daniel Egan – Microsoft

Adnan Masood, PhD. Microsoft MVP - UST Global


LFB
Northern
Public subv
I'm Saqib Shaikh.
0:12 / 2:44

“Hello? Is it me you're looking for?”

Face and Voice Recognition

Using Microsoft Cognitive Services

Daniel Egan - Microsoft

Why Microsoft Cognitive Services ?

Easy

Roll your own with REST APIs

Simple to add: just a few lines of code required

Flexible

Make the same API code call on iOS, Android, and Windows

Integrate into the language and platform of your choice

Tested

Built by experts in their field from Microsoft Research, Bing, and Azure Machine Learning

Quality documentation, sample code, and community support

GET A KEY

BUILD

python™

node.js™

GitHub

msdn

uservoice

stackoverflow

Understand the data around your application

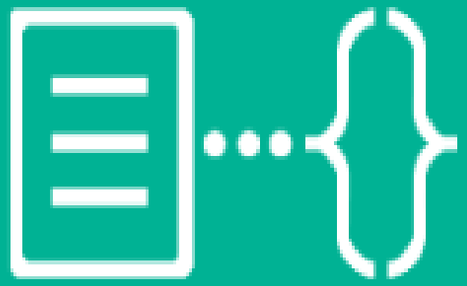
Microsoft Cognitive Services API services will help you understand and interact with audio, text, image, and video

Microsoft Cognitive Services

Cognitive Services

LUIS

(Language Understanding Intelligent Service)



Microsoft Cognitive Services

Cognitive Services

Vision APIs	Face APIs	Speech APIs	LUIS (Language Understanding Intelligent Service)
Analyze Image	Face Detection	Speech Recognition	Detect Intent
OCR	Face Grouping	Text to Speech	Determine Entities
Generate Thumbnail	Face Identification	Speech Intent Recognition	Improve Models

Cognitive Services

microsoft.com/cognitive

Vision	Speech	Language	Knowledge	Search
Computer Vision	Custom Recognition	Bing Spell Check	Academic Knowledge	Bing Web Search
Emotion	Speaker Recognition	Linguistic Analysis	Entity Linking	Bing Image Search
Face	Speech	Language Understanding	Knowledge Exploration	Bing Video Search
Video	Translator	Text Analytics	Recommendations	Bing News Search
		WebLM		Bing Autosuggest

Cognitive Services

microsoft.com/cognitive

Vision	Speech	Language	Knowledge	Search
Computer Vision	Custom Recognition	Bing Spell Check	Academic Knowledge	Bing Web Search
Emotion	Speaker Recognition	Linguistic Analysis	Entity Linking	Bing Image Search
Face	Speech	Language Understanding	Knowledge Exploration	Bing Video Search
Video	Translator	Text Analytics	Recommendations	Bing News Search
		WebLM		Bing Autosuggest

Powerful models

Cognitive Services models are trained using the same deep learning and machine learning techniques that power many products across Microsoft

Easy to use

Microsoft Cognitive Services allows you to focus on your application by easily including these services across platforms through simple REST APIs

Easily include Cognitive Services

```
ProjectOxford.Face.Contract.Face[] detectionResults = new ProjectOxford.Face.Contract.Face[0];
ProjectOxford.Face.Contract.IdentifyResult[] identifyResults = new ProjectOxford.Face.Contract.IdentifyResult[0];

using (var imageFileStream = Context.ContentResolver.OpenInputStream(imageUri))
{
            //Call detection and identification REST API
            detectionResults = await client.DetectAsync(imageStream: imageFileStream, analyzesAge: true, analyzesGender: true);

            identifyResults = await client.IdentifyAsync(personGroupId, detectionResults.Select(face => face.FaceId).ToArray());
}
```

Vision APIs

Analyze an Image

OCR

Get Thumbnail

Vision

Computer Vision API
Distill actionable information from images

Face API
Detect, identify, analyze, organize, and tag faces in photos

Emotion API
Personalize experiences with emotion recognition

Video API
Analyze, edit, and process videos within your app

Content Moderator
Machine-assisted moderation of text and images, augmented with human review tools

Custom Vision Service
Customizable web service that learns to recognize specific content in imagery

Video Indexer
Process and extract smart insights from videos

Analyze Image Service

Understand content and features within an image

Analyze Image – Example

Type of Image:

Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-Line Drawing
Black & White Image	False

Content of Image:

Categories	[{ “name”: “people_swimming”, “score”: 0.099609375 }]
Adult Content	False
Adult Score	0.18533889949321747
Faces	[{ “age”: 27, “gender”: “Male”, “faceRectangle”: {“left”: 472, “top”: 258, “width”: 199, “height”: 199}}]

Image Colors:

Dominant Color Background	White
Dominant Color Foreground	Grey
Dominant Colors	White
Accent Color	

Analyze Image – Example

Content of Image:

Categories

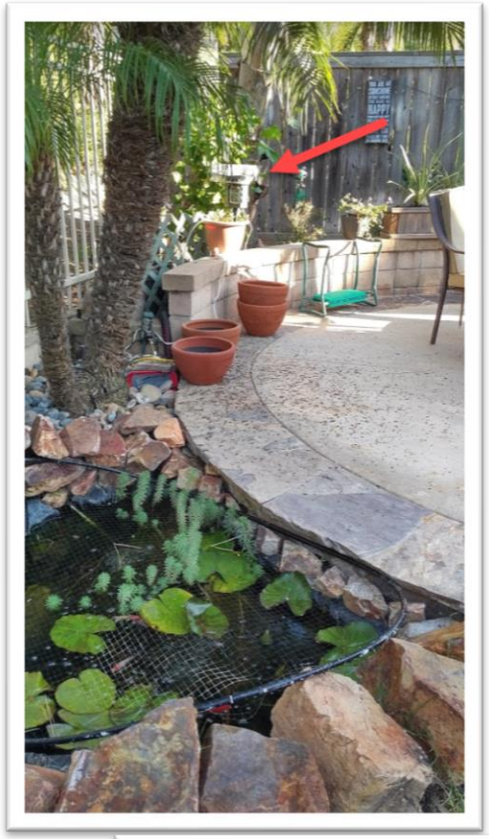

```
v0: [{ “name”: “animal”, “score”: 0.9765625 }]
V1: [{ "name": "grass", "confidence": 0.9999992847442627 },
 { "name": "outdoor", "confidence": 0.9999072551727295 },
 { "name": "cow", "confidence": 0.99954754114151 },
 { "name": "field", "confidence": 0.9976195693016052 },
 { "name": "brown", "confidence": 0.988935649394989 },
 { "name": "animal", "confidence": 0.97904372215271 },
 { "name": "standing", "confidence": 0.9632768630981445 },
 { "name": "mammal", "confidence": 0.9366017580032349,
 "hint": "animal" },
 { "name": "wire", "confidence": 0.8946959376335144 },
 { "name": "green", "confidence": 0.8844101428985596 },
 { "name": "pasture", "confidence": 0.8332059383392334 },
 { "name": "bovine", "confidence": 0.5618471503257751,
 "hint": "animal" },
 { "name": "grassy", "confidence": 0.48627158999443054 },
 { "name": "lush", "confidence": 0.1874018907546997 },
 { "name": "staring", "confidence": 0.165890634059906 }]
```

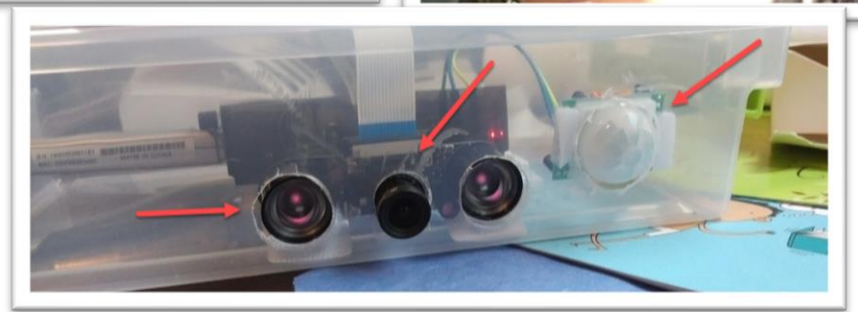

Describe


```
0.975 "a brown cow standing on top of a lush green field“
0.974  “a cow standing on top of a lush green field”
0.965  “a large brown cow standing on top of a lush green field”
```

Wackoon

Wackoon





Content of Image:

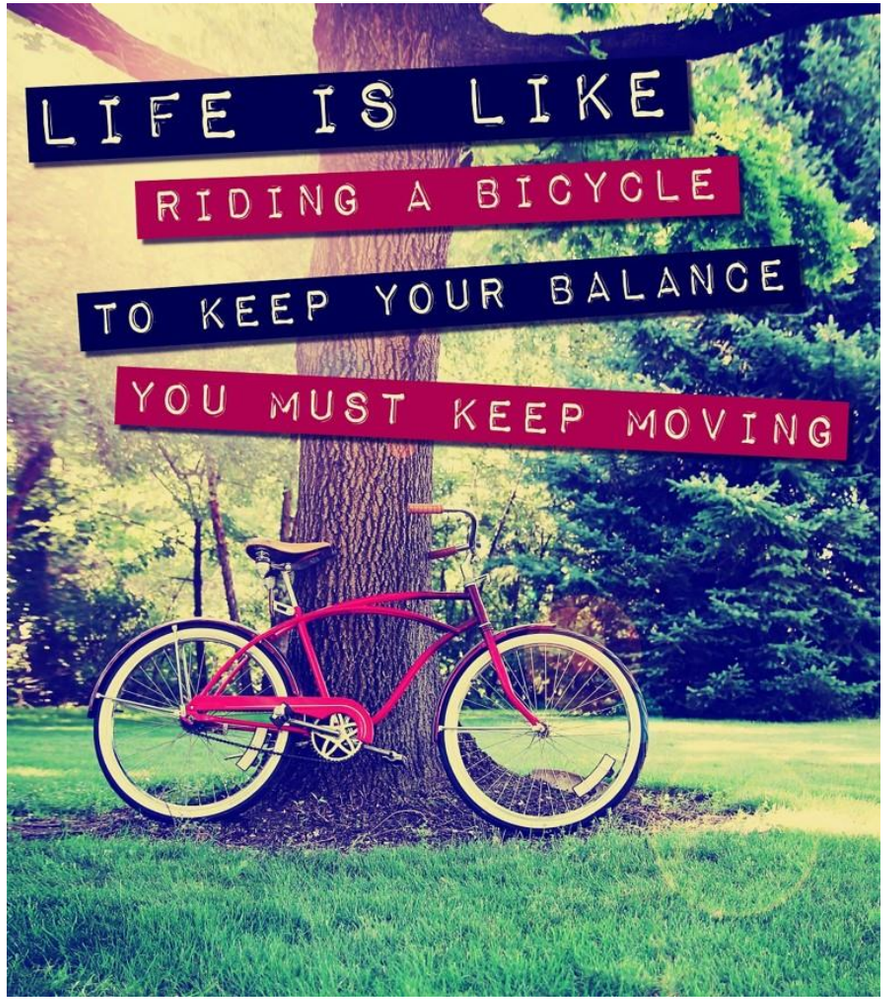
Categories


```
v0: [{ “name”: “animal”, “score”: 0.9765625 }]
V1: [{ "name": "grass", "confidence": 0.9999992847442627 },
 { "name": "outdoor", "confidence": 0.9999072551727295 },
 { "name": "cat", "confidence": 0.99954754114151 },
 { "name": "raccoon", "confidence": 0.9976195693016052 },
 { "name": "grey", "confidence": 0.988935649394989 },
 { "name": "animal", "confidence": 0.97904372215271 },
 { "name": "standing", "confidence": 0.9632768630981445 },
 { "name": "mammal", "confidence": 0.9366017580032349,
"hint": "animal" },
 { "name": "aquarium", "confidence": 0.8946959376335144 },
 { "name": "green", "confidence": 0.8844101428985596 },
 { "name": "grass", "confidence": 0.8332059383392334 },
 { "name": "water", "confidence": 0.5618471503257751 },
 { "name": "grassy", "confidence": 0.48627158999443054 },
 { "name": "lush", "confidence": 0.1874018907546997 },
 { "name": "staring", "confidence": 0.165890634059906 }]
```

OCR Service

Detect and recognize words within a photo

OCR – Example



TEXT:
LIFE IS LIKE
RIDING A BICYCLE
TO KEEP YOUR BALANCE
YOU MUST KEEP MOVING

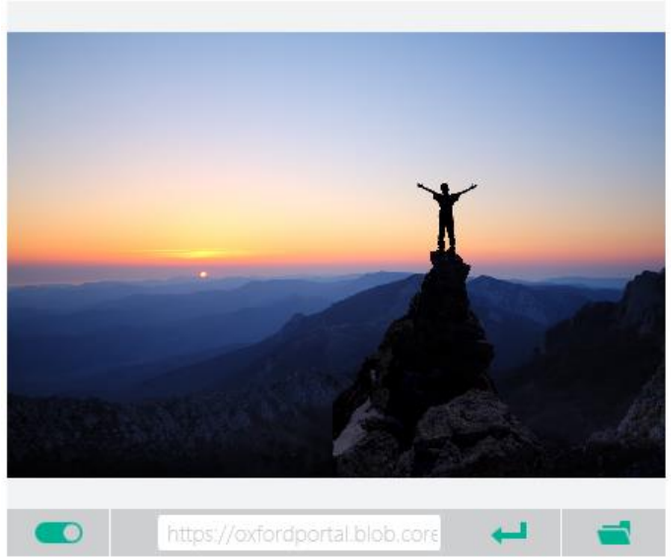
JSON:

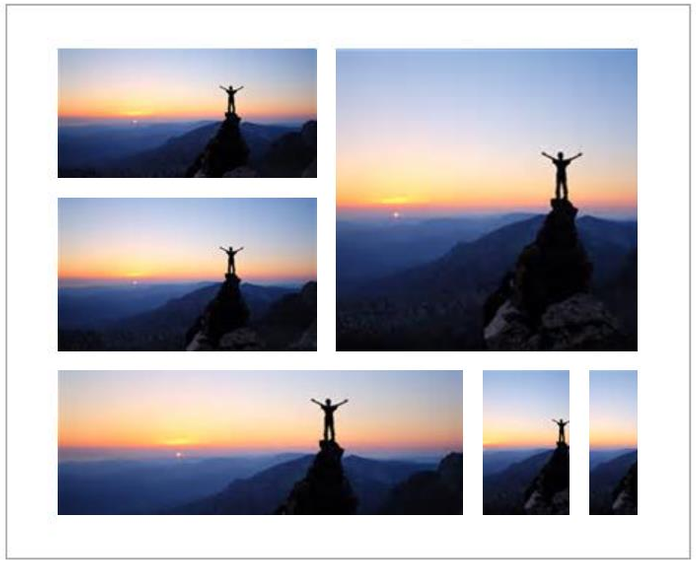

```
{
 "language": "en",
 "orientation": "Up",
 "regions": [
  {
   "boundingBox": "41,77,918,440",
   "lines": [
    {
     "boundingBox": "41,77,723,89",
     "words": [
      {
       "boundingBox": "41,102,225,64",
       "text": "LIFE"
      },
      {
       "boundingBox": "356,89,94,62",
       "text": "IS"
      },
      {
       "boundingBox": "539,77,225,64",
       "text": "LIKE"
      }
. . .
```

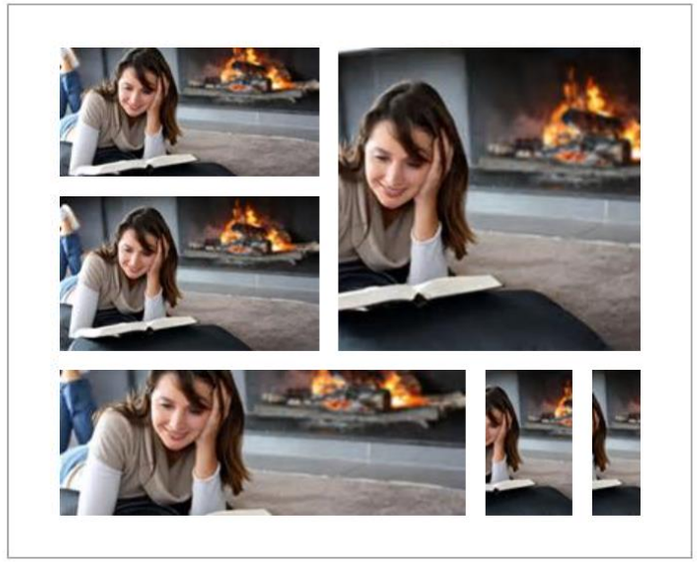
Smart Thumbnail Service

Scale and crop an image,
while retaining key content

Smart Thumbnail – Example







Face APIs

Detection

Verification

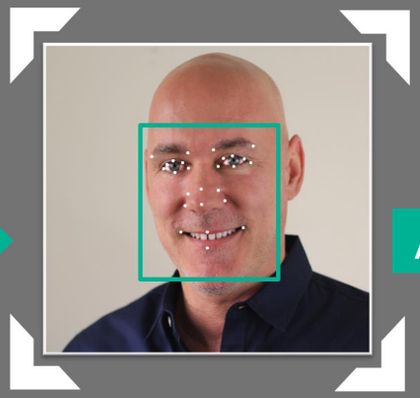
Grouping

Identification

Face API – Detection

INPUT
IMAGE

DETECTION



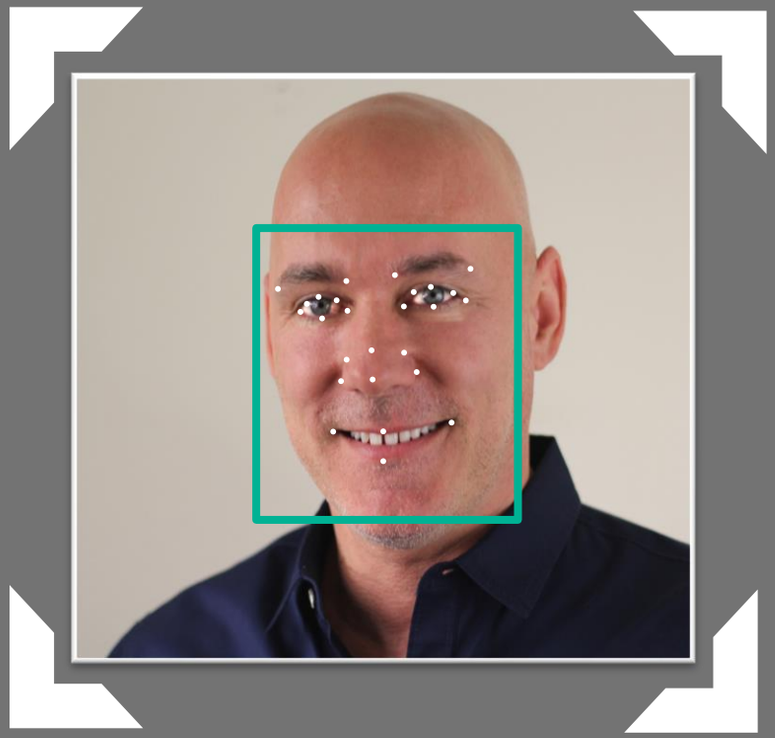
FACIAL
RECTANGLE + LANDMARKS

ATTRIBUTES

Detection Result:
JSON:


```
[
  {
    "faceRectangle": {
      "width": 109,
      "height": 109,
      "left": 62,
      "top": 62
    },
    "attributes": {
      "age": 41,
      "gender": "male",
      "headPose": {
        "roll": "2.9",
        "yaw": "-1.3",
        "pitch": "0.0"
      }
    "faceLandmarks": {
      "pupilLeft": {
        "x": "93.6",
        "y": "88.2"
      },
      "pupilRight": {
        "x": "138.4",
        "y": "91.7"
      },
     ...
```

Face API – Detection



Detection Result:
JSON:


```
[
  {
    "faceRectangle": {
      "width": 109,
      "height": 109,
      "left": 62,
      "top": 62
    },
    "attributes": {
      "age": 31,
      "gender": "male",
      "headPose": {
        "roll": "2.9",
        "yaw": "-1.3",
        "pitch": "0.0"
      }
    "faceLandmarks": {
      "pupilLeft": {
        "x": "93.6",
        "y": "88.2"
      },
      "pupilRight": {
        "x": "138.4",
        "y": "91.7"
      },
     ...
```

Face API – Verification

Given two faces, determine whether they are the same person

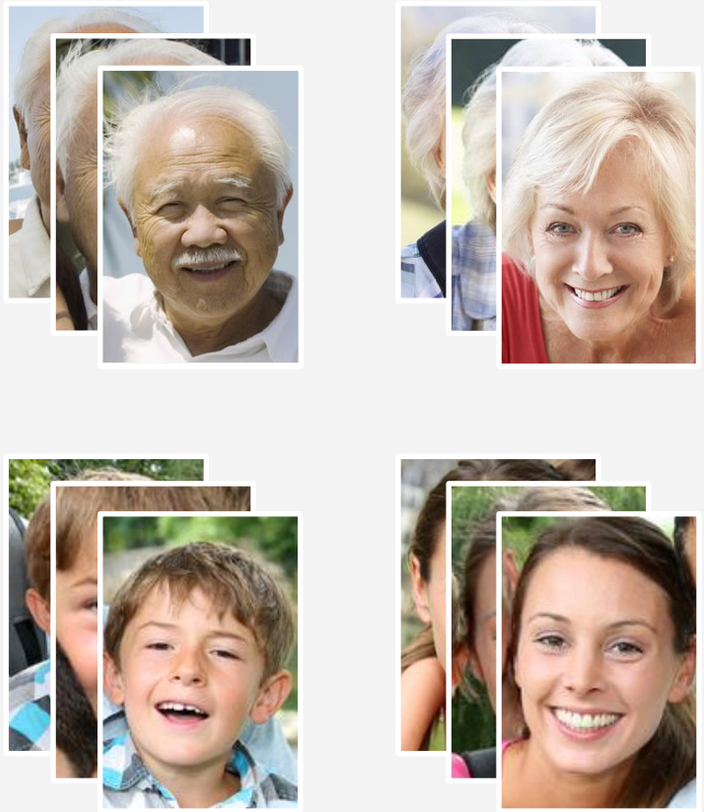
Verification Result:
JSON:

```
[
    {
        "isIdentical":false,
        "confidence":0.01
    }
]
```

Demo

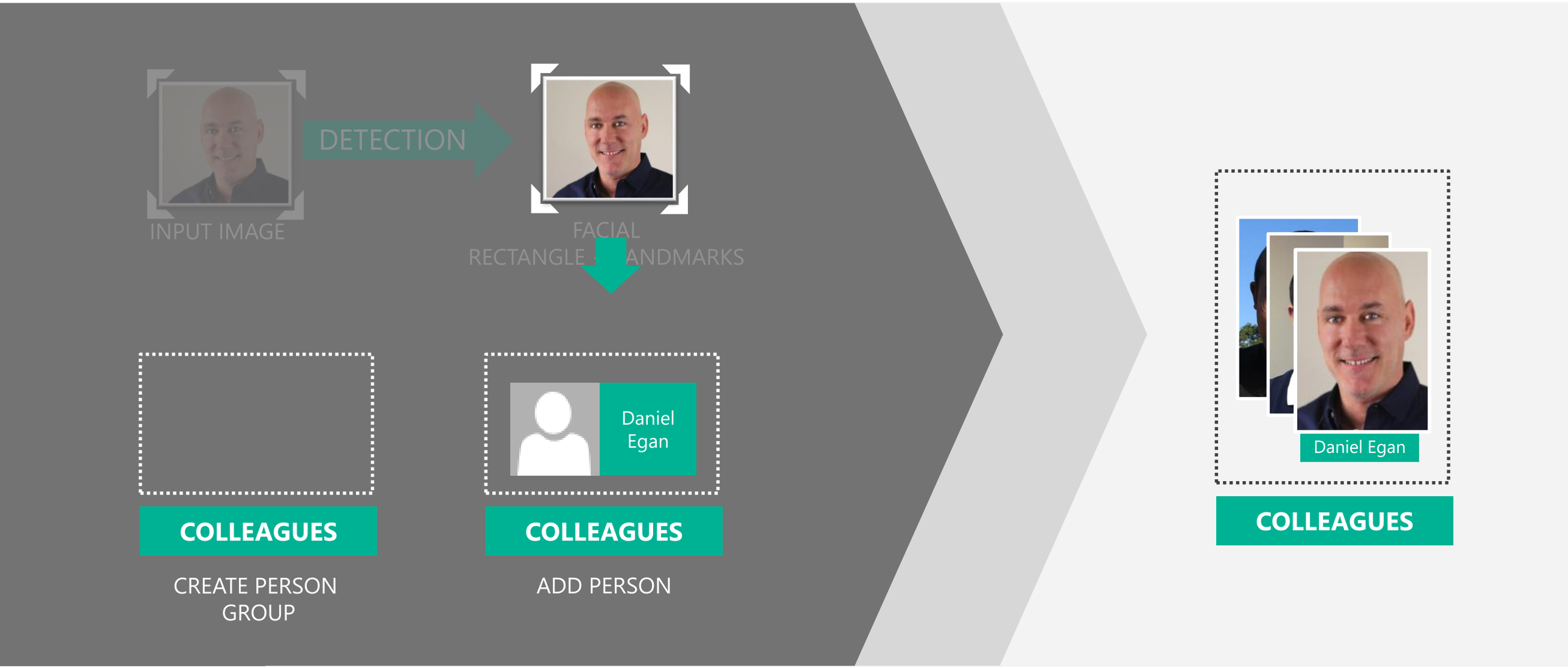
Lets see the online API

Face API – Grouping



CLUSTERED BY
DETECTED PEOPLE

Face API – Create Person Object



Face API – Identify

NEW INPUT IMAGE

IDENTIFY

Natalie Huber

GROUP PERSON OBJECTS

Daniel Egan

RECOGNITION

COLLEAGUES

He is
Daniel Egan.

Best practices for Devs

Samples and SDKs exist

For ObjectiveC/Swift/iOS, Java/Android, C#/Windows, and Python (Jupyter notebook)

https://www.microsoft.com/cognitive-services/en-us/SDK-Sample?api=computer%20vision

Limitations

Computer Vision API describes images in English only

Face API detects up to 64 human faces in one image

Facial detection: JPEG, PNG, GIF (first frame), and BMP supported, image file size of 1KB-4MB, detectable face size 36x36-4096x4096 pixels, returned faces ordered by face rect size desc

Fun random details

FindSimilarFace has 2 modes: matchPerson (default, same person) and matchFace (similar faces)

FaceGroup API takes between 2-1000 candidate faces

Documentation: https://www.microsoft.com/cognitive-services/en-us/documentation

Data

Computer Vision

Description, tags, clip art, line drawing, black & white, IsAdultContent/Score, IsRacy/Score, categories, faces, dominant colors, accent color

https://www.microsoft.com/cognitive-services/en-us/computer-vision-api

Emotions

Anger, contempt, disgust, fear, happiness, sadness, surprise, and neutral

https://www.microsoft.com/cognitive-services/en-us/emotion-api

Face

Bounding box, 27 facial landmarks, age, gender, head pose, smile, facial hair, glasses

https://www.microsoft.com/cognitive-services/en-us/face-api

Custom Vision Service

Custom Vision Service

A customizable web service that learns to recognize specific content in imagery

Upload Images

Upload your own labeled images, or use Custom Vision Service to quickly tag any unlabeled images.

Train

Use your labeled images to teach Custom Vision Service the concepts you want it to learn.

Evaluate

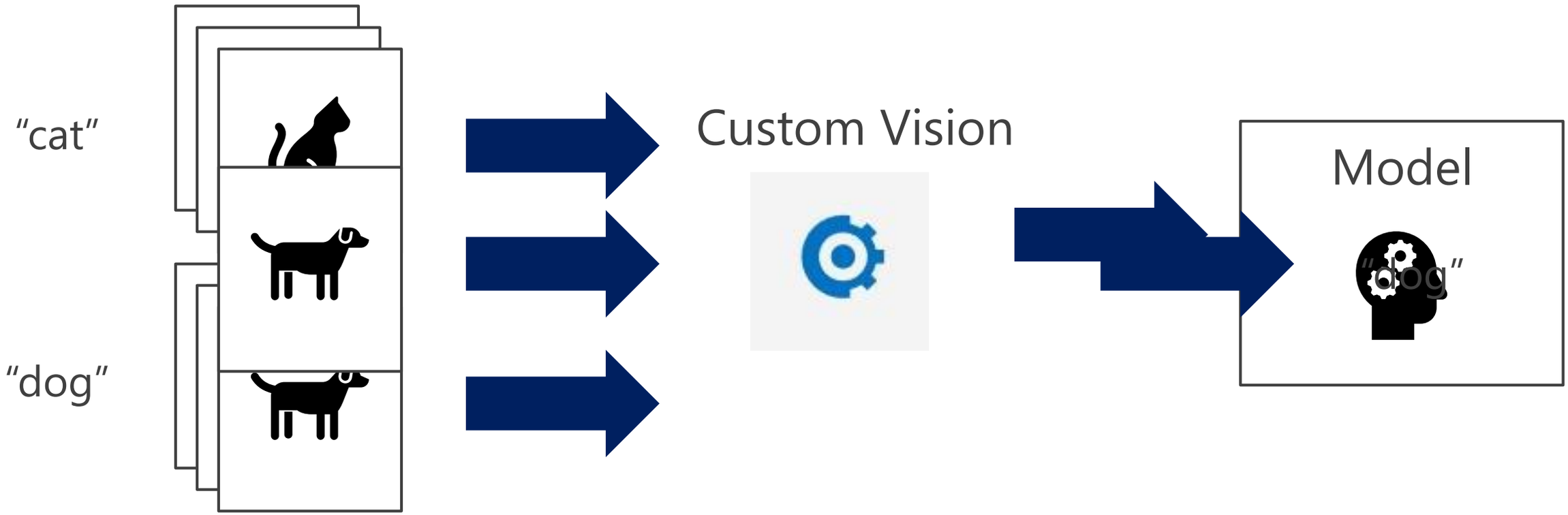
Use simple REST API calls to quickly tag images with your new custom computer vision model.

Active learning

Images evaluated through your custom vision model become part of a feedback loop you can use to keep improving your classifier.

What is it?

Custom Vision Service is an easy-to-use tool for prototyping, improving, and deploying a custom image classifier to a cloud service, without any background in computer vision or deep learning required.



Export models to mobile!

- Announcement: https://aka.ms/cvsexport
- Sample: https://github.com/Azure-Samples/cognitive-services-ios-customvision-sample
- Xamarin port: https://github.com/Xamarin/ios-samples/tree/master/ios11/Co

Demo

Dog vs. cat classifier: http://customvision.ai

Intelligent Kiosk: https://aka.ms/kioskapp

Export to CoreML on iOS11: https://aka.ms/cvsexport

Best Practices for using Custom Vision

- Use at least 30 images for each tag
- Images should be the focus of the picture
- Use sufficiently diverse images and backgrounds (ex: cats with red background and dogs with blue background)
- Train with images that are similar in {quality, resolution, lighting, etc.} to the images that will be used in prod
- Supports Microsoft accounts (MSA) and AAD

Gotchas to watch for

- V1 doesn’t currently do object detection with bounding boxes within an image
- Intended to be robust to subtle differences, so V1 is not well suited to tasks like defect detection/quality assurance
- Current project limitations while in preview: 1000 images, 50 tags, 20 iterations saved
- Current account limitations while in preview: 20 projects, 1000 predictions per day

Example Customer Scenarios

Customer Support

- Enable a customer to identify a product for support by taking a photo. No finding the manual or pulling the appliance out to identify it!

Service Engineers

- Identify parts for ordering

Manufacturing

- Fault detection on assembly lines to avoid machine downtime and drop in production rates (provided differences are obvious)

Data Scientists

- Automatic tagging instead of manual, to create features or labels

Resources: Custom Vision Service

Get started at http://customvision.ai

Build 2017 Talk:
https://channel9.msdn.com/Events/Build/2017/T6022

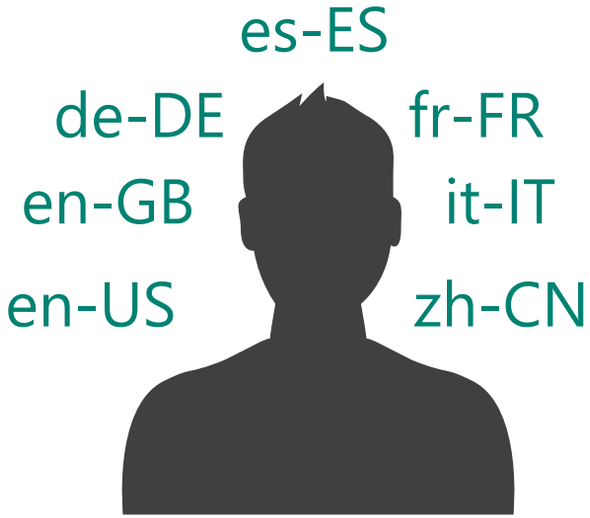
Programmatic API access using C# (Python and Node SDKs coming soon):
https://github.com/Microsoft/Cognitive-CustomVision-Windows

Speech APIs powered by Bing

Voice Recognition (Speech to Text)

Voice Output (Text to Speech)

Voice Recognition



- Converts spoken audio to text
- Same backend which powers Cortana
- Support for 7 languages at launch

Voice Recognition

	REST API	Client Library
SUPPORTED PLATFORMS	Any	Windows, Android, iOS
DATA SUPPORT	Yes	Yes
MIC SUPPORT	No	Yes
SILENCE DETECTION ON MIC	No	Yes
LENGTH OF UTTERANCE	Short	Short and long
NUMBER OF RESPONSES	n-best response back	multiple partial results, n-best (short) and multiple phrases (long)

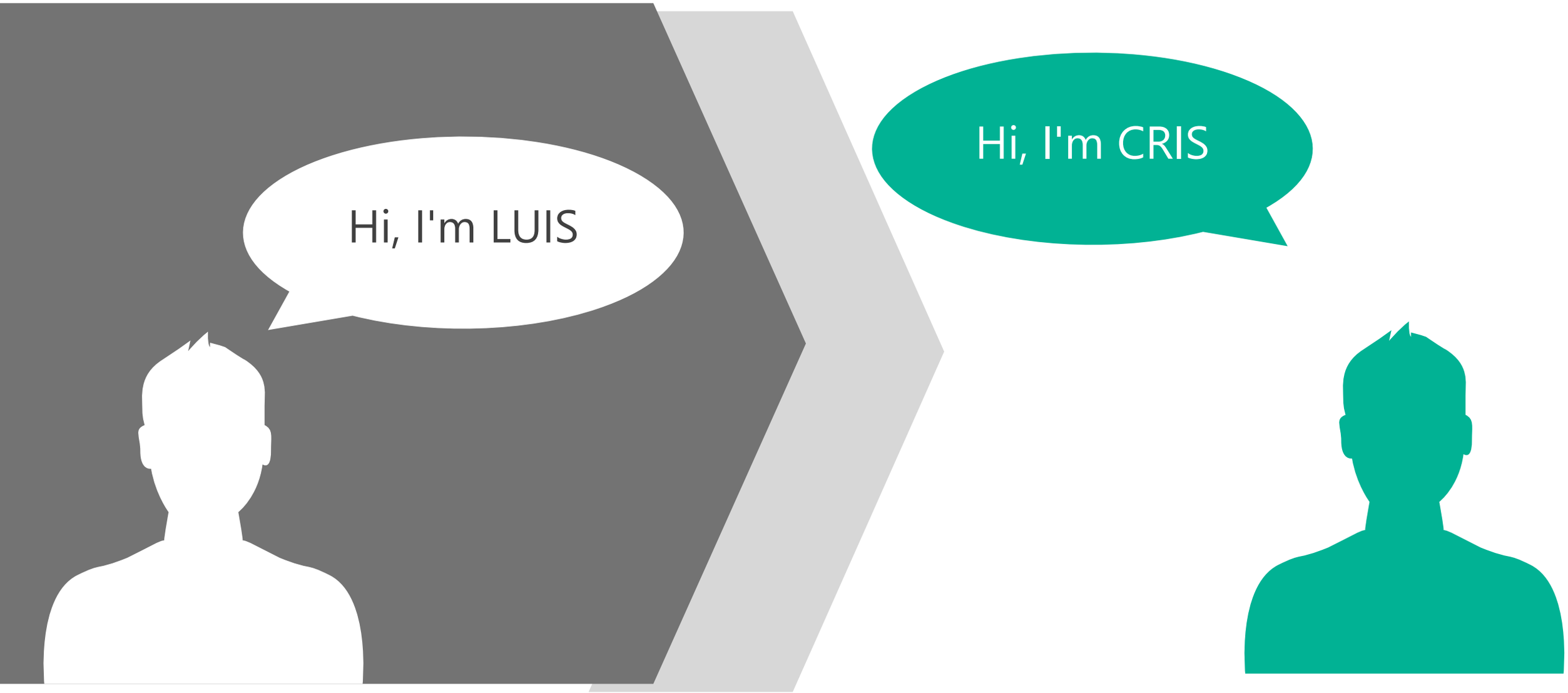
Windows 10 has Speech APIs built in

Voice Recognition Modes

	Short Form	Long Form
Duration of Audio	< 15 seconds	< 2 minutes
Final Result	n-best choice	Best Choice, delivered at sentence pauses
Partial Results	No	Yes

********* Final N-BEST Results *********
[0] Confidence=Normal Text="450 six St San Francisco."
[1] Confidence=Normal Text="For 50 six St San Francisco."
[2] Confidence=Normal Text="456th St San Francisco."
[3] Confidence=Normal Text="450 six St in San Francisco."
[4] Confidence=Normal Text="456 St San Francisco."

The Adventures of LUIS and CRIS



LUIS
Language Understanding Intelligent Service

Determine Intent

Detect Entities

Improve Models

Language Understanding Intelligent Service

Lets you understand what your users are saying

Seamless integration with Speech Recognition

A few examples are enough to deploy an application

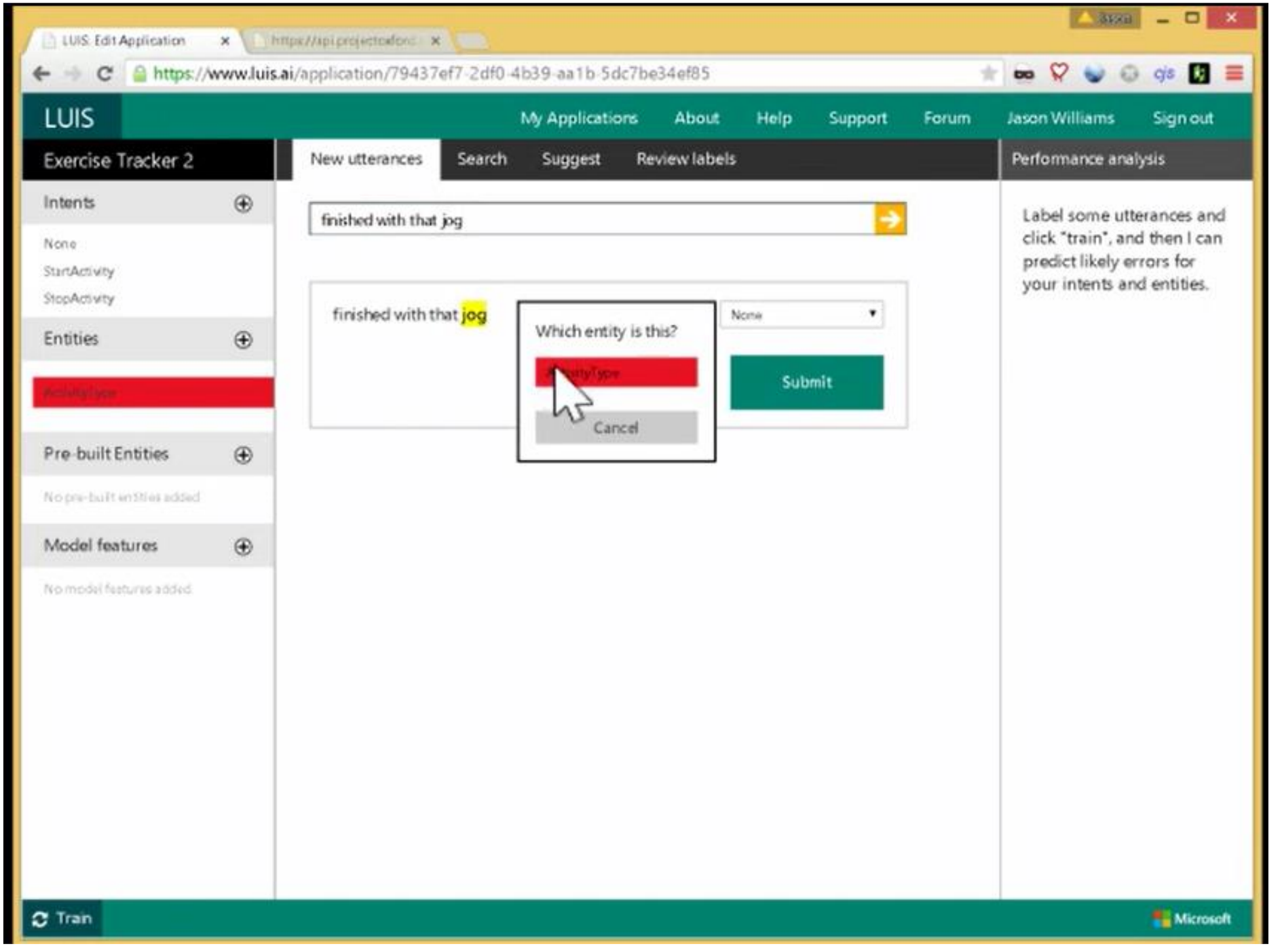
LUIS learns over time

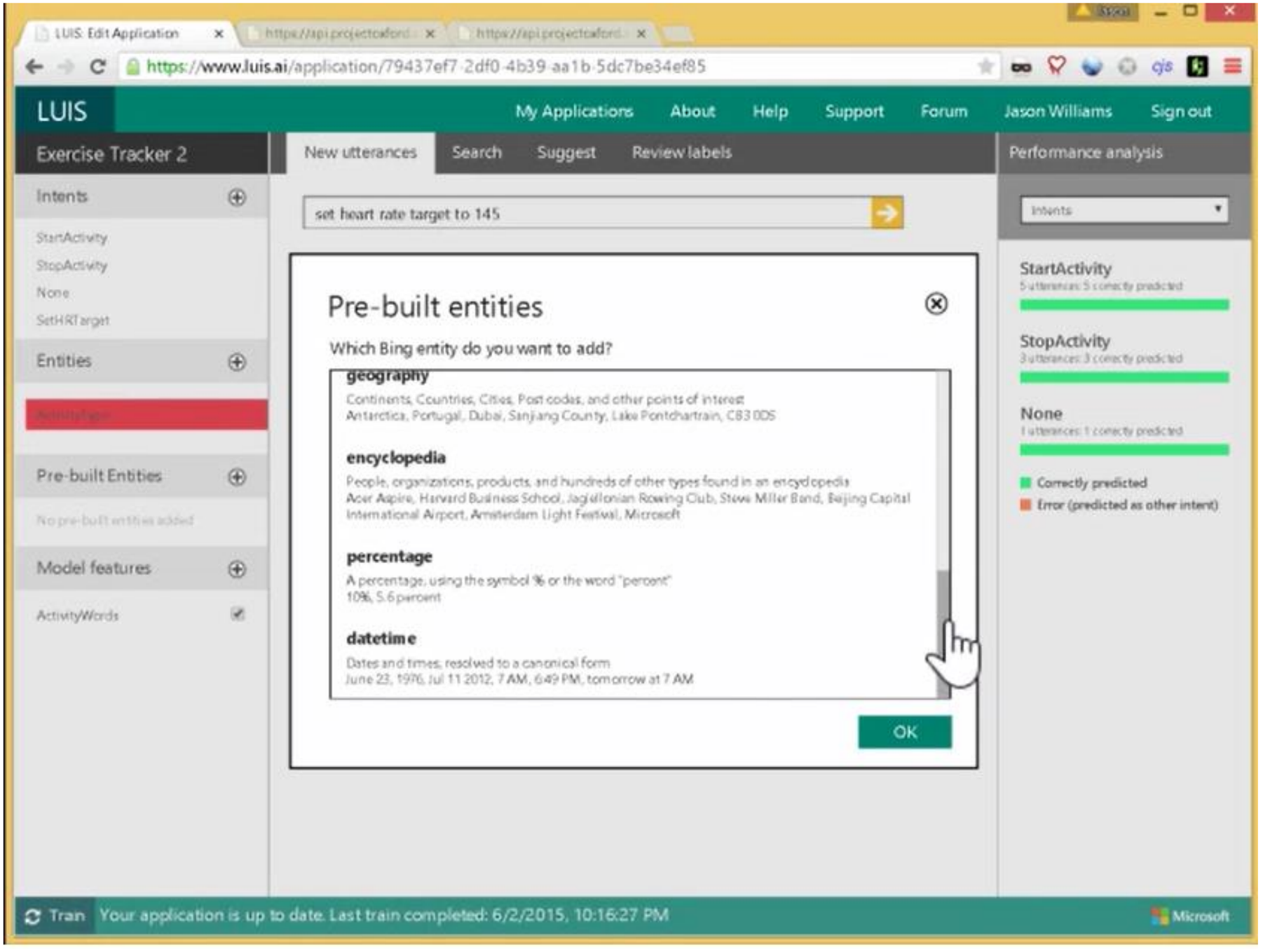
Define Concepts

Provide Examples

Deploy

Active Learning


LUIS: Edit Application
https://www.luis.ai/application/79437ef7-2df0-4b39-aa1b-5dc7be34ef85
LUIS
My Applications
About
Help
Support
Forum
Jason Williams
Sign out
Exercise Tracker 2
New utterances
Search
Suggest
Review labels
Performance analysis
Intents
None
StartActivity
StopActivity
Entities
Pre-built Entities
Model features
finished with that jog
finished with that jog
Which entity is this?
None
Submit
Cancel
Label some utterances and click "train", and then I can predict likely errors for your intents and entities.
Train
Microsoft


LUIS: Edit Application
https://www.luis.ai/application/79437ef7-2df0-4b39-aa1b-5dc7be34ef85
LUIS
My Applications
About
Help
Support
Forum
Jason Williams
Sign out
Exercise Tracker 2
New utterances
Search
Suggest
Review labels
Performance analysis
Intents
StartActivity
StopActivity
None
SetHRTarget
Entities
Pre-built Entities
No pre-built entities added
Model features
ActivityWords
set heart rate target to 145
Pre-built entities
Which Bing entity do you want to add?
geography
Continents, Countries, Cities, Post codes, and other points of interest
Antarctica, Portugal, Dubai, Sanjiang County, Lake Pontchartrain, CB3 0DS
encyclopedia
People, organizations, products, and hundreds of other types found in an encyclopedia
Acer Aspire, Harvard Business School, Jagiellonian Rowing Club, Steve Miller Band, Beijing Capital International Airport, Amsterdam Light Festival, Microsoft
percentage
A percentage, using the symbol % or the word "percent"
10%, 5.6 percent
datetime
Dates and times, resolved to a canonical form
June 23, 1976, Jul 11 2012, 7 AM, 6:49 PM, tomorrow at 7 AM
OK
StartActivity
5 utterances: 5 correctly predicted
StopActivity
3 utterances: 3 correctly predicted
None
1 utterances: 1 correctly predicted
Correctly predicted
Error (predicted as other intent)
Tran
Your application is up to date. Last train completed: 6/2/2015, 10:16:27 PM
Microsoft

LUIS: My Applications | https://api.projectoxford.

https://api.projectoxford.ai/luis/v1/app[illegible]bon?id=6512b0fc-bb63-415e-ab0d-7b2f6a8c55f0&subscriptio


```
{
  "query": "start tracking a run",
  "intents": [
    {
      "intent": "StartActivity",
      "score": 0.9999995
    },
    {
      "intent": "None",
      "score": 0.0262200516
    },
    {
      "intent": "StopActivity",
      "score": 0.022188127
    },
    {
      "intent": "SetHRTarget",
      "score": 0.00241672155
    }
  ],
  "entities": [
    {
      "entity": "run",
      "type": "ActivityType"
    }
  ]
}
```

Language Understanding Models


```
{
  “entities”: [
    {
      “entity”: “flight_delays”,
      “type”: “Topic”
    }
  ],
  “intents”: [
    {
      “intent”: “FindNews”,
      “score”: 0.99853384
    },
    {
      “intent”: “None”,
      “score”: 0.07289317
    },
    {
      “intent”: “ReadNews”,
      “score”: 0.0167122427
    },
    {
      “intent”: “ShareNews”,
      “score”: 1.0919299E-06
    }
  ]
}
```

Custom Recognition Intelligent Service

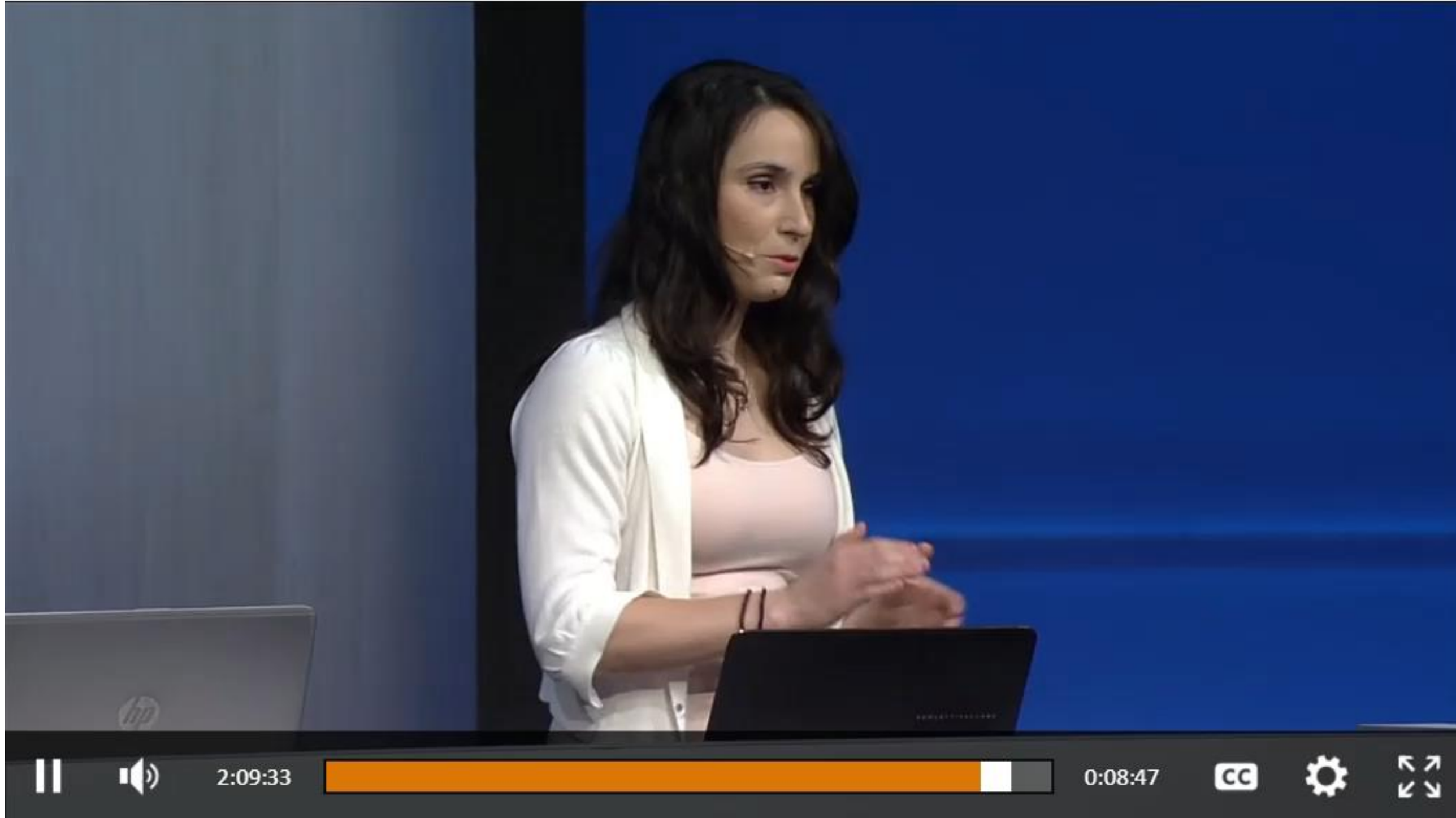
Lets you overcome speech recognition barriers like speaking style, background noise, and vocabulary.

Custom language models

Custom acoustic models

Access Endpoint

Any Device


2:09:33
0:08:47
CC

MyMoustache.net

74,319 faces analyzed and counting #MyMoustacheRobot

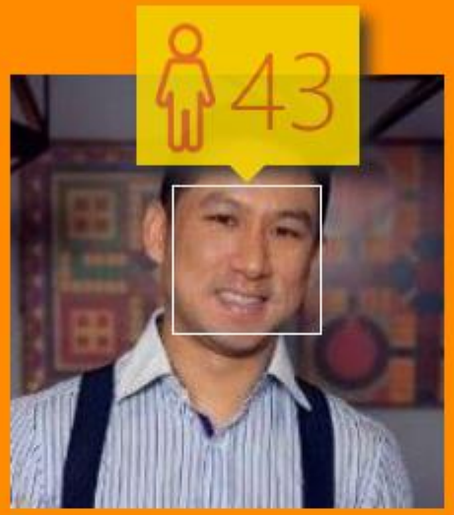
De-stache Me! **Try Again!**

Sorry if we didn't quite get the results quite right - we are still improving this feature..

It doesn't look like you have a moustache! #NOSTACHE

How-Old.net

How old do I look? #HowOldRobot



Sorry if we didn't quite get it right - we are still improving this feature..

Try Another Photo!

Microsoft

P.S. We don't keep the photo

Share 2.3M Tweet

The magic behind How-Old.net

Privacy & Cookies | Terms of Use | View Source

How-Old.net

How old do I look? #HowOldRobot

Sorry if we didn't quite get it right - we are still improving this feature.

Try Another Photo!

Microsoft

P.S. We don't keep the photo

Share 2.3M Tweet

The magic behind How-Old.net

Privacy & Cookies | Terms of Use | View Source

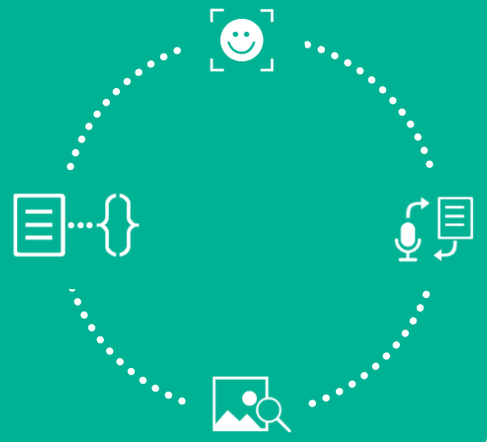
Kids

KidProVision

Demo

Lets see the online API

Microsoft Cognitive Services



A portfolio of REST APIs and SDKs which enable developers to write applications which understand the content within the rapidly growing set of multimedia data