The “Intelligent Kiosk”

a Cortana Intelligence Demo

Setup and Presenter’s Script

Last updated: 12/12/2016

This document is provided “as-is”. Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

© 2016 Microsoft Corporation. All rights reserved.

Table of Contents

[Getting Started 4](#_Toc454900462)

[Application Installation & Setup 6](#_Toc454900463)

[Presentation Script 6](#_Toc454900464)

[Introduction 6](#_Toc454900465)

[How-Old Demo 6](#_Toc454900466)

[Intelligent Mall Kiosk Demo 8](#_Toc454900467)

[Emotion API Playground Demo 12](#_Toc454900468)

[Face Identification + Training Demo 14](#_Toc454900469)

[Conclusion 20](#_Toc454900470)

# Getting Started

The Intelligent Kiosk is a Windows 10 Universal App that has a collection of demos highlighting the Perceptual Intelligence power of Azure Machine Learning and Microsoft Cognitive Services, which are part of the Cortana Intelligence Suite. The kiosk is an example of an intelligent application, one that can understand human faces and predict their age, gender and emotion. It can do so from live images from a web cam, Bing Images or local photos. It also allows you to very quickly create and train a model to perform face identification. Through the power of speech-to-text and text sentiment analysis, the kiosk also shows how an application can detect the sentiment of what someone is saying. Most the demos in the kiosk are hands-free and self-guided. They require only the web cam as input, using the live camera feed to detect when people are nearby and engaging with them through UI feedback.

**Form factor**

The kiosk can take many forms, from a simple Surface or laptop device to a large PPI display or a Surface Hub. The smaller form factors work well when you want to just show it during a presentation, while the large form factors work well when engaging with the audience in an event. The key thing is that the device needs to be running Windows 10, as the kiosk is a Windows 10 Universal App (yes, it works on Win10 phones too). Here are a few examples of the ideal kind of setup for event booths where people are often in crowds and standing in front of the screen:

  

**Display & camera requirements**

It will work on any display, and since the main form of input is the web cam itself, a touch screen is not a requirement. However, if you will be placing this in some event and want people to be able to interact with the screen to perform some actions in some demos that require button clicks (sharing of a photo, for instance), then you might want a touch screen. Also, you will want to have a top-mounted camera. We recommend the Microsoft LifeCam Studio camera, but you are welcome to try other cameras as well. The key thing is that the camera is top-mounted, as you want the users to have a similar experience as looking at a mirror when interacting with the kiosks. This can be an issue with the largest Surface Hub versions, as they have side cameras. On top of the camera setup, you will need to have some photos ready, depending on the demos you plan to show:

* 3 (or more) JPEG photos of yourself for the **Face API Training demo** (saved on the PC you will use for the demo). These photos should be ones of just you – where your face is clearly visible.
* For the **Intelligent Mall Kiosk Demo** you will need photos of the following (some stock photos will be provided to you as part of this demo packaging):
  + Child aged between 4-12
  + Two adults aged approximately between 30-40

During the demo these above two photos will need to be held up in front of the webcam – so you may wish to do one of the following:

1. Print them out on some good quality photo paper (A5 size is ideal). Or,
2. Load them onto your mobile phone - assuming your mobile phone’s screen size is large enough to be picked up by the webcam when held up.

**Kiosk App Setup**

You will need a set of API keys depending on the demos you plan to show below. We will walk you through on how to acquire them in the later sections of this guide.

* **Face API Key**
* **Emotion API Key**
* **Vision API Key** (only necessary for the CaptionBot demo)
* **Text Analytics Key** (only necessary for the Intelligent Mall Kiosk demo)

# Application Installation & Setup

## Camera setup

You will want to have a top-mounted camera. We recommend the Microsoft LifeCam Studio camera, but you are welcome to try other cameras as well. The key thing is that the camera is top-mounted, as you want the users to have a similar experience as looking at a mirror when interacting with the kiosks. This can be an issue with the largest Surface Hub versions, as they have side cameras.

## Kiosk App Setup

* A Windows 10 PC/Laptop/Surface with a web cam (we recommend the Microsoft LifeCam Studio camera)
* Install the application from <https://www.microsoft.com/store/apps/9nblggh5qd84>
* Launch the app (search for Kiosk in Windows, or look for recently installed apps)
* Use the Sign-in link at the top of the Overview page to sign in (this is an on-time step). You can either use a social account (Microsoft Account tied to Hotmail, outlook, etc) or create and use an account with a valid email address (follow the “Sign up now” link at the bottom of the Sign-In page if you need to create an account instead of using a social account).
* Enter the API keys in the Settings Page for the several services that the kiosk consumes. The settings are saved automatically in the roaming storage provided by Windows 10, and are loaded again every time the app launches.
* **Workspace Key:** This is a key concept internal to the tool, and is used as a way to separate Face Identification data (person groups, people in those groups, their face samples, etc) between different workspaces/people that share the same Face API Key. Enter something meaningful for the key (e.g. JoeKiosk), or generate a unique key through the tool and keep track of that value for your records. If you already have a key from a previous setup just re-use it, that way you will start with existing training data.
* **Face, Emotion, Vision and Text Analytics API Keys**: Microsoft employees can get access to free, high-throughput, keys by visiting <https://projectoxfordkeysignup.azurewebsites.net/>. These keys expire every 7 or 30 days, depending on the option you choose, so you will need to renew them when they stop working. Non-Microsoft employees can also get free keys from the public facing Cognitive Services website (<https://www.microsoft.com/cognitive-services/en-us/sign-up>), but they won’t be high throughput, so any of the realtime demos in the kiosk won’t work very well with them. Both of these links are available via the Settings Page in the kiosk by using the “Click here if you need keys” button in the Settings Page.

# Presentation Script

## Introduction

All major industries have significantly changed with the power of predictive analytics. For example, in the retail space, we have now the capacity to enhance sales and marketing efforts based on the consumer sentiments captured, which is then used to predict future behavior and preference.   
Cortana Intelligence enables business, such as retailers to take their consumer experience to the next level. Focusing on innovative Cognitive Services, the retailer can benefit by providing product recommendation in the form of personalized offers. The scenario below highlights how the use of intelligent kiosks enables the retailers to connect with their consumers like never.

## How-Old Demo

|  |  |  |
| --- | --- | --- |
| **Screenshots** | **Instructions** | **Narrative** |
|  | 1. Once the Intelligent Kiosk App has been launched (and you are shown the start page) - open the App’s Left menu Text Edit(7) (located in the top left of the App) 2. Click the ‘Demos’ menu item. | Hello everyone, my name is XXX and today we are going to show you some capabilities of Cortana Intelligence, using the Microsoft Cognitive Services.  For this demonstration, we built an Intelligent Kiosk. |
|  | 1. Click the ‘How-Old Kiosk’ icon from the list of available demos. | This first demonstration is the one you might have already seen on very famous website <http://how-old.net>. This capability was presented for the first time during the //build Conference in 2015.  What was meant to be a simple demonstration showcasing the Cognitive Services capabilities of Cortana Intelligence became a viral phenomenon, reaching more than 90 million people. As of today there has been over 600 million pictures uploaded to <http://how-old.net> . |
|  | 1. Center your face in the webcam viewer and hold still while the app begins its countdown.   The age calculation will show. | Why was this so popular? We realized that people weren’t really interested in knowing how old they were, but how old they looked…  The how-old demonstration makes use of the Face API offered via Microsoft Cognitive Services.  As you can see when I'm facing the camera my age and gender seems to magically appear but this is just one possibility of the capabilities of Cognitive Services. |

## Intelligent Mall Kiosk Demo

|  |  |  |
| --- | --- | --- |
| **Screenshots** | **Instructions** | **Narrative** |
|  | 1. Open the left menu and click on the ‘**Demos’** menu item. 2. Then click the ‘**Intelligent Mall Kiosk**’ item. | If we consider a retail store scenario, we can see how the Microsoft Cognitive Services technologies could be combined with digital signs, self-help kiosks, and customer tracking to provide more targeted and engaging experience for their customers. |
|  | 1. Center the webcam on a person so they are visible in the webcam viewer (shown in bottom right of App). 2. Click the Text Edit(3) Icon in the bottom right of the webcam viewer to proceed with the capture. | Imagine you're standing in front of a Kiosk - not just any Kiosk, but an ‘Intelligent Kiosk’.  As you can see, when I **click on the button (step 2)**, my age and gender is guessed thanks to the Face detection capability. |
|  | 1. Interact with the web page contents which are shown to show it’s a ‘real’ web page. | Based on the age + gender detected, the kiosk can provide a recommendation, in the form of a website with products relevant to me. |
|  | 1. Press the Text Edit icon to re-enable the webcam. 2. Now, center the webcam on two adults.  Alternatively, hold up a printed photo (of supplied stock photo) /or/ show it on your mobile device.   Click the C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit(3).png Icon again. | The interesting point is that if I'm with another person in front of the camera, another type of recommendation would be shown that could be better suited to something we could do together.  In this example, the Face API system is recognizing both of us – so it is showing us a restaurant where we can both go and eat. |
|  | 1. Press the Text Edit icon to re-enable the webcam. 2. Now, center the webcam on a person but this time hold up a photo of a child so it’s they both appear in the webcam preview.   Click the C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit(3).png Icon again. | Now another interesting example - imagine I’m not with my kid however I have the picture of my child with me. When the webcam detects this another recommendation magically appears – such as this playground.  But you can see that it’s a very expensive playground… |
|  | 1. Click the Text Edit(5) icon to show the speech recognition dialog. 2. Speak clearly into the demonstration PC’s microphone and say   ‘I think that’s really expensive’ 3. Then click on the square when finished speaking. | So imagine for one second, if this kiosk could not only determine the age + gender of the customer - but it could also understand what the customer is saying.  So if I click on that button and if I say “I think that’s really expensive”, this kiosk understands that the product being shown is above my desired budget.  To keep me interested and engaged – the kiosk can then show another similar product that’s cheaper. |
|  | 1. (Optional) you can click the refresh button and show how other phrases are being analyzed (i.e. ‘I love this item’) etc. | Another interesting feature of this demo app is that below the Speech to text translation, there’s a bar measuring the emotion of what is being said. This represents an analysis of the sentiment of my sentence – from negative to positive sentiments.  In this case the kiosk considered that what I said was pretty negative. |
|  |  | So based on this demonstration, we could put together not one, but three machine learning API’s combined in a very easy way. The **Face API** (to determine the gender & age), the **Speech to Text Translation API** (to understand what was being said), and then finally the **Text to Sentiment Analysis API** (to detect the sentiment in their phrase). |

## Emotion API Playground Demo

|  |  |  |
| --- | --- | --- |
| **Screenshots** | **Instructions** | **Narrative** |
|  | 1. Open the left menu and click on the **‘Demos’** item. 2. Then click the ‘**Emotion API Playground**’ item in the list of demos. | But what if we could go one step further?  What if we could identify how people would react to these recommendations.  What was science fiction yesterday is real today – as another Microsoft Cognitive Service has the ability to recognize emotions. |
|  | 1. To initiate the webcam viewer, press the Text Edit(4) icon in the top right of the screen 2. Center yourself on the webcam and then and smile / look happy. 3. Once you are ready to take a webcam photo, press the Text Edit icon. 4. Then when your done with your emotion, press C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit(3).png to get your analysis. | Here I just need to click on a button and my emotion is detected.  The Emotion API takes a facial expression in an image as an input, and returns the confidence across a set of emotions for each face in the image, as well as a bounding box for the face. |
|  | 1. Repeat this process but this time try to look surprised (or another distinct emotion). | The emotions detected are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise.  These emotions are understood to be cross-culturally and universally communicated with particular facial expressions. |

## Face Identification + Training Demo

|  |  |  |
| --- | --- | --- |
| **Screenshots** | **Instructions** | **Narrative** |
|  | 1. Open the left menu and click on the ‘**Face Identification Setup**’ item. | In these previous demonstrations, we had the opportunity to see what could be an ideal recommendation based on the age, gender, and even emotions.  But there we were doing more guessing than actually knowing who the people are.  So imagine that you could provide an even better recommendation, or in fact the most accurate recommendation ever, based on identifying who is actually in front of the Kiosk.  We can do that too thanks to the Face API identification technology.  So how does this work?  Part of the Face API, the Face Identification capability helps us search and identify faces in one or more photos. You can create a known ‘person’ in the system and then upload one or more photos of them to ‘train’ the system. Once the system has been trained, it will be able to recognize them in new photos it’s never seen before. |
|  | 1. Click the C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit(2).png in the top right to add a new ‘Person Group’. 2. Type ‘Customer Loyalty Program’ in the textbox and click ‘Add’. 3. Select this group when it appears in the main content area (after it’s created). | When we are training the Face API, it allows us to group sets of people together in what is known as ‘Person Groups’. These groups are saved in the system so they can be accessed by other Kiosks in your business too.  So let’s start by creating a group for our favourite customers called ‘Customer Loyalty Group’. |
|  | 1. Click the Text Edit(2) in the top right corner to add a new Person. 2. Type your name in the textbox and click ‘Add’. 3. Once it’s saved to the system, select the name you just entered from the list that appears. | Now we need to add a person to this group that we want to try and recognize.  In a real life scenario, you would likely add many other customers to this person group as well - however for the purposes of this demo we’ll just add the one. |
|  | 1. Click the C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit(2).png icon in the top right corner to ‘Add Faces’ to this person. 2. The app will automatically search Bing for searches that match the name. 3. Otherwise, when the dropdown picker is shown, click the Text Edit(6) icon in the top right corner. 4. Navigate to a folder on your PC which has one or more photos of you. Select the photos to use for training and click ‘Open’. | For this new person we have added, we can add one or more existing photos from our hard drive which contains this person.  As with the person groups + names, information about these photos will also be stored in the Cognitive Services system for later use by this Kiosk or other connected kiosks in your business. |
|  | 1. The images you selected in the previous step will now appear in the popup window –select each one you want to import (so it is highlighted with an orange border) and click on the  icon to complete the import.   *During the final import you will see a progress icon as images are uploaded to the Cognitive Services service. Once this is complete the images will be shown. Note - depending on your connection speed you may only be shown preview icons for each image instead of photos.* | As with the person groups + names metadata, these photos will also be stored in the Cognitive Services system for later use. |
|  | 1. Open the left Menu again and click the Back Button to return back to the ‘Person Group’ you have created. 2. Click the C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit.png icon in the top right corner to ‘Train Group’. 3. The progress ring will be shown for a few moments while the training is performed – wait until this disappears (and training is complete). | Now that we have our person group, person and some photos setup, it’s time to ‘train’ the system as to what these people look like.  We’ll use the photos we have already provided here to achieve this. |
|  | 1. Open the Left Menu and click on the ‘**Demos**‘ item. 2. Then click the ‘**Face API Playground**’ item. | Ok so now we have trained the system, let’s validate it by checking if it can recognize me via the webcam. |
|  | 1. Press the C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit(4).png icon in the top right of the screen and center yourself in the webcam viewer that appears. 2. Once you are ready to take a webcam photo press the C:\Users\suppo\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Text Edit(3).png icon. | Now you can see that the system has not only guessed my age – it’s actually recognized who I am.  You’ll note that next to my name, as with the emotion detection there’s also a ‘confidence level’ shown here which has been returned from the system.  If we’ve got a high score then we can be (mostly) certain that it’s a correct match. |

## Conclusion

In closing, today we have shown you how easy it is to connect several Cognitive Services such as Face, Speech, Text Analytics, and Emotion APIs in one demonstration and highlight in the future several other components of Cortana Intelligence such as data ingestion & visualization capabilities. While the technology is new, we can imagine a future where such features will be used in many applications in various industries.