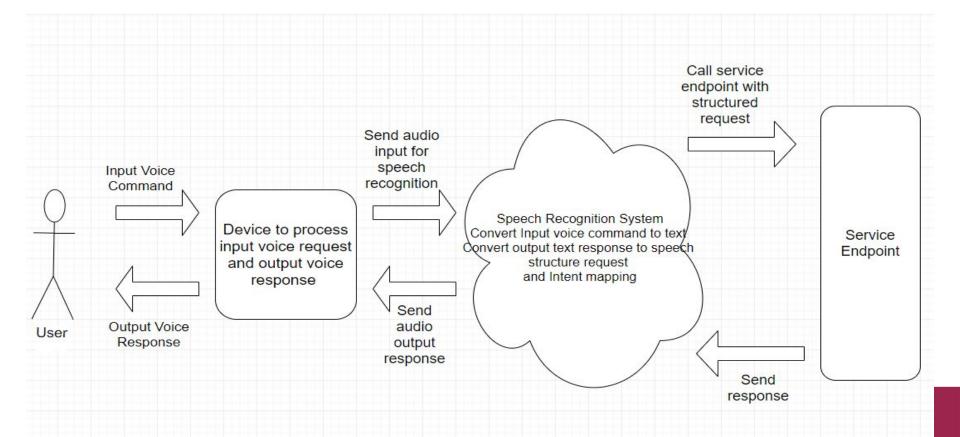
# Voice Activated Systems

- Sachin Bhalekar
- Sumit Jawale

## **Business Problem**

- Integrating voice capabilities into existing system.
- Enabling voice commands as an additional feature of interaction to existing system.

# Reference Architecture



## **Interaction Model**

- Invocation Name
- Utterances
- Slots
- Intents
- Endpoint

## **Invocation Name**

- Invocation name is the phrase (word) to activate the voice service for your system.
- Also known as the wake word for the system.
- Eg.: CSUN Calendar, Papa Johns, PayPal, Uber

## Utterances

- Utterances are the phrases that will be used to interact with the system.
- They can be the sentences used to request on a query.
- Eg.:
  - What are the upcoming events?
  - o Book me a ride!
  - Order a large pepperoni pizza with extra cheese.

## Slots

- Slots are the templates used to refer to a specific period, time or an instance in the utterance.
- It can be a number, date, time, period/range, etc.
- Eg:
  - O What are {tomorrow's} events?
  - Order a pizza at {8:30 PM}
  - Book a ride from {Home} to {Office}

## Intents

- Intents can be known as the actions taken upon the given utterances.
- They invoke the functions that process the request and return the response.
- Eg.:
  - FetchUpcomingEvents
  - BookRide
  - PlaceAnOrder

# **Endpoints**

- Endpoints are the system where the logic is built to handle the requests.
- Endpoints can be connected with an DB, or the internet or with third party
   APIs for processing the request.
- Endpoints can be a service or a HTTPS implementation.

## Voice Skill Models

- Custom Interaction
- Video/Visual cards
- Flash Briefings/Updates
- Smart Home

## **Custom Interaction**

- Build a personalized interaction model for your skill.
- Design a unique experience for the users.
- Create and build your own utterances, slots, intents and the endpoint.

## Video/Visual Cards

- Add some video/visual content to your skill.
- This helps in engaging users into the skill.
- Also used to provide additional information to the request.
- Eg.:
  - Play Interstellar.
  - Show me comedy movies.
  - o Open CNN.

# Flash Briefings/Updates

- Give users control of their news feed.
- This helps users control what updates they listen to.
- User can add the skill to their custom flash briefings for better experience.
- Eg.:
  - Give me my flash briefings.
  - O What are my updates?
  - What's in the sports news?

## **Smart Home**

- Gives users control of their smart home devices.
- This helps users turn off the lights and other devices without getting up.
- Eg.:
  - Turn off bedroom lights.
  - o Increase the volume of the home stereo.
  - Decrease the thermostat by 2 degrees.

# Security

- Security can be provided with the existing security module of the system for access to the user account. [Account linking]
- Provide a passcode/pin to access the features.
- Use voice id to provide security.

# **Testing**

- Skills need to be tested before publishing.
- Simulators to provide end-to-end testing with voice scripts.
- Manual testing with the system.
- Create test cases to test the actions individually.

# Skill Management and Publishing

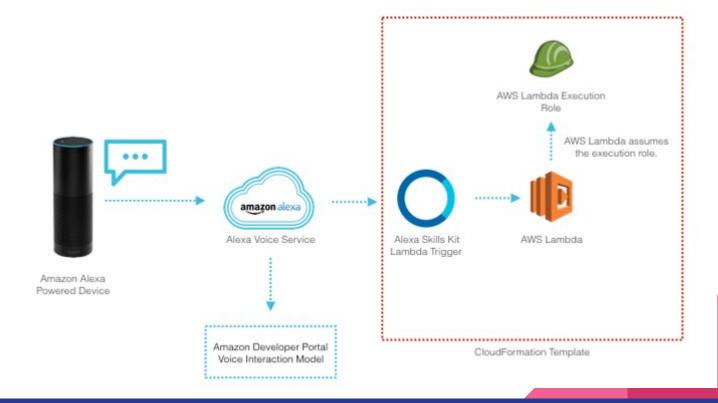
- A portal or SDK to develop and build your skills.
- Manage the skill for development and production separately.
- Skill can be published globally as a service.
- Rewards by providers based on user engagement.

# Implementation Guide

## **Amazon**: Alexa Voice Service

- Alexa Voice Service is a Amazon cloud service which helps in building Alexa skills and Alexa-enabled products easily and rapidly.
- Alexa is always getting smarter with new capabilities through machine learning.
- Complex technologies like Automatic Speech Recognition and Natural Language
   Understanding are handled in the cloud.
- AVS provides you easy-to-use APIs for core functionality like audio playback, volume control, or text-to-speech.
- AVS helps build your own custom skill with the Alexa Skills Kit (ASK).

## Alexa: Architecture

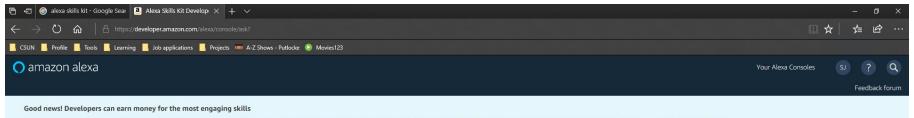


# Alexa Skill: Development

- Requirements
- Steps to build
- Testing
- Publishing
- Measure Statistics

## Requirements

- Alexa Skill Kit Developer Account
  - Portal to build and test your skills
- Amazon Web Service Account (AWS Lambda)
  - To host your endpoint of the skill on AWS Lambda
- AWS CLI
  - To deploy your code easily on AWS Lambda

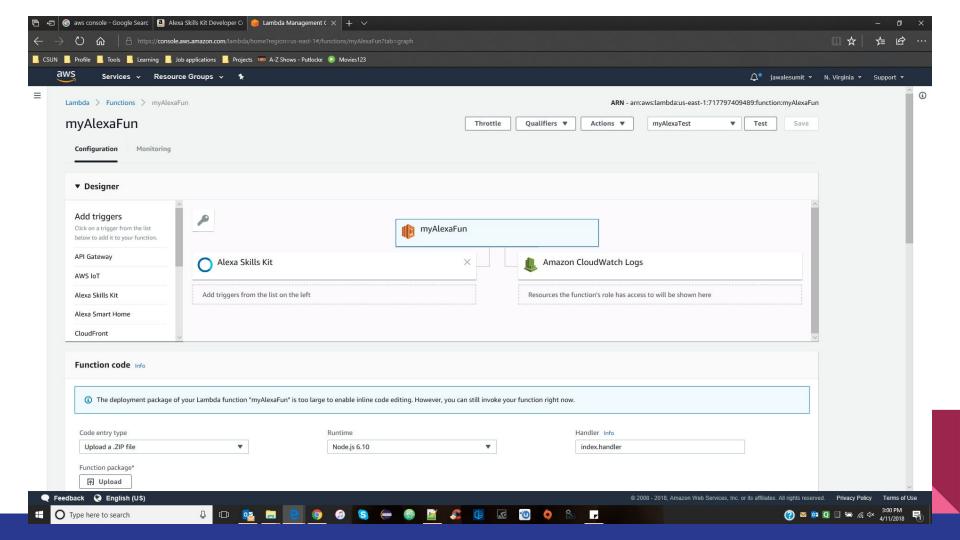


We're rewarding developers who design Alexa skills that customers love most! Developers can earn money each month for eligible skills that have the highest customer engagement in eligible skill categories. What's your next big idea? Learn more.

Welcome to the new Alexa Skills Kit Developer Console

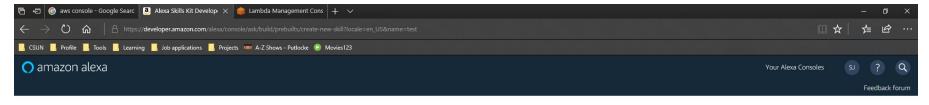
Curious about what's new? Watch the video overview or read about what's changed.

#### Alexa Skills SKILL NAME LANGUAGE TYPE MODIFIED STATUS ACTIONS CSUN Calendar Measure Edit Delete English (US) 2018-04-10 In Development Custom amzn1.ask.skill.a1aafc00-fb28-48c1-be51-2bc7db0d7ee7 TEST Measure Edit Delete English (US) Custom 2018-04-10 In Development amzn1.ask.skiil.5e8b410f-2da0-49e2-b004-bff4e2a66749



# Steps to Build

- Login to ASK
- Go to skills console and click on create new skill
- Give a name to skill
- Choose your model:
  - Custom
  - Flash Briefings
  - Smart Home
  - Video
- Skill Builder Checklist

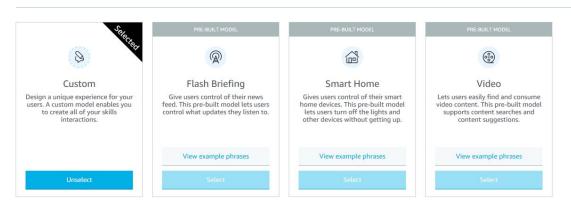


Create skill



#### Choose a model to add to your skill

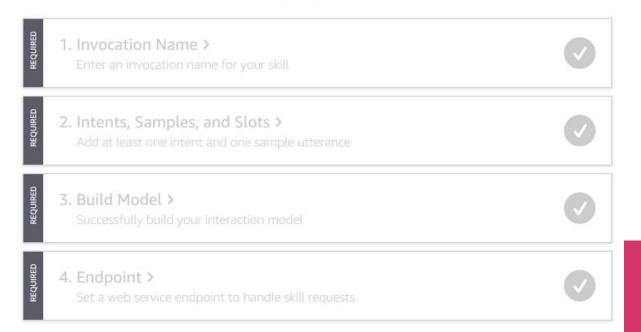
There are many ways to start building a skill, you can design your own custom model or start with a pre-built model. Pre-built models are interaction models that contain a package of intents and utterances that you can add to your skill.



## Skill Builder Checklist

#### Skill builder checklist

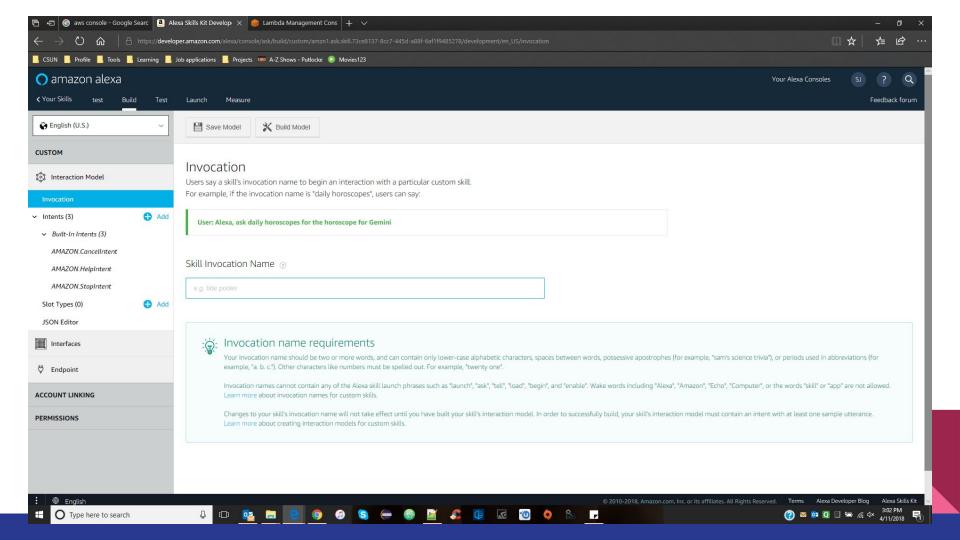
Complete these steps to be able to test your skill using the simulator in the test tab, or with your echo device.



## 1. Invocation Name

- Users say a skill's invocation name to begin an interaction with a particular custom skill.
- Requirements:
  - 2 or more words.
  - only lower-case alphabetic characters, spaces between words, possessive apostrophes, or periods used in abbreviations.
  - characters like numbers must be spelled out. For example, "twenty one".
  - cannot be Alexa launch phrases: "launch", "ask", "tell", "load", "begin", and "enable".
  - Wake words including "Alexa", "Amazon", "Echo", "Computer", or the words "skill" or "app" are not allowed.

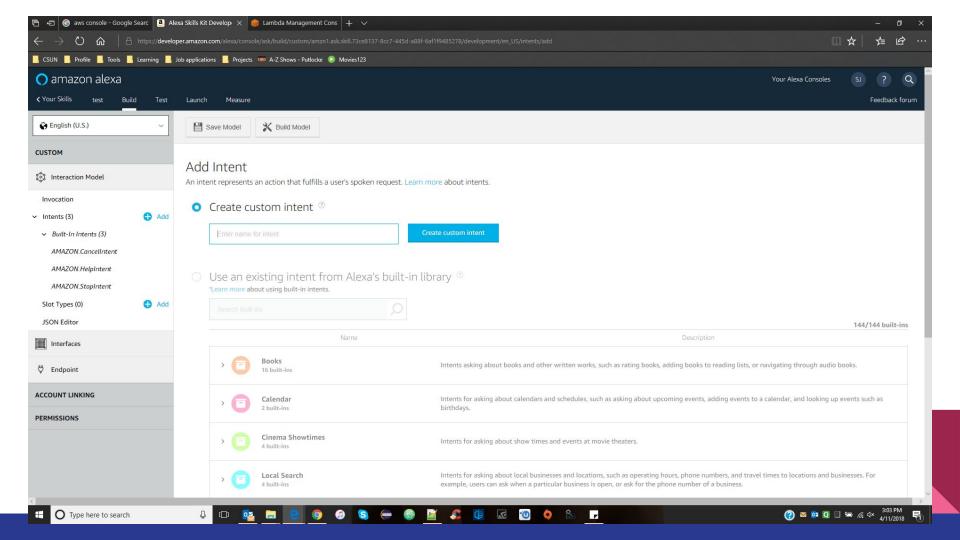




# 2. Intents, Samples and Slots

- An intent represents an action that fulfills a user's spoken request.
- Choose from Amazon's 144 built-in intents or create your own.
- Provide sample utterances to the intent
- Use slots for re-usability





## 3. Build Model

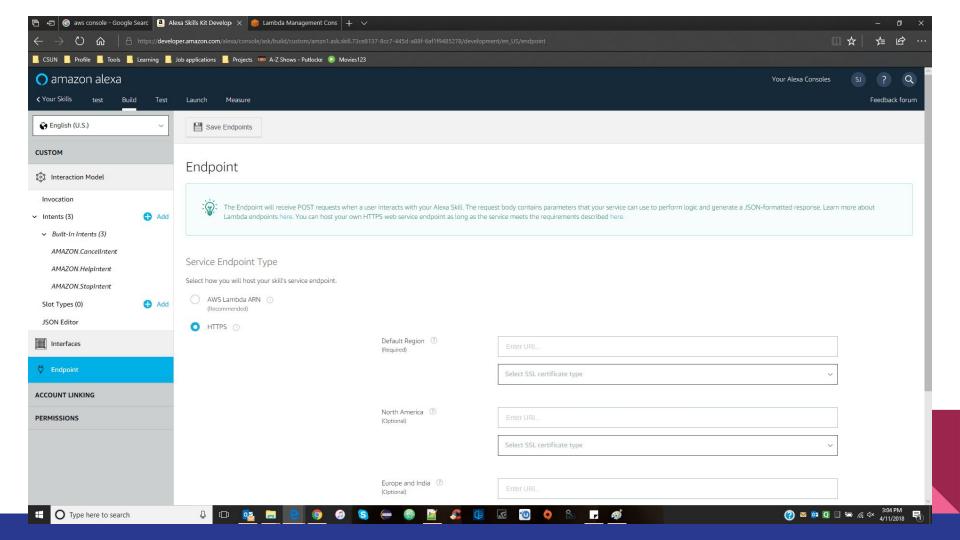
- Now build the interaction model.
- Check for any inconsistency or errors.
- Resolve and build again.



# 4. Endpoint

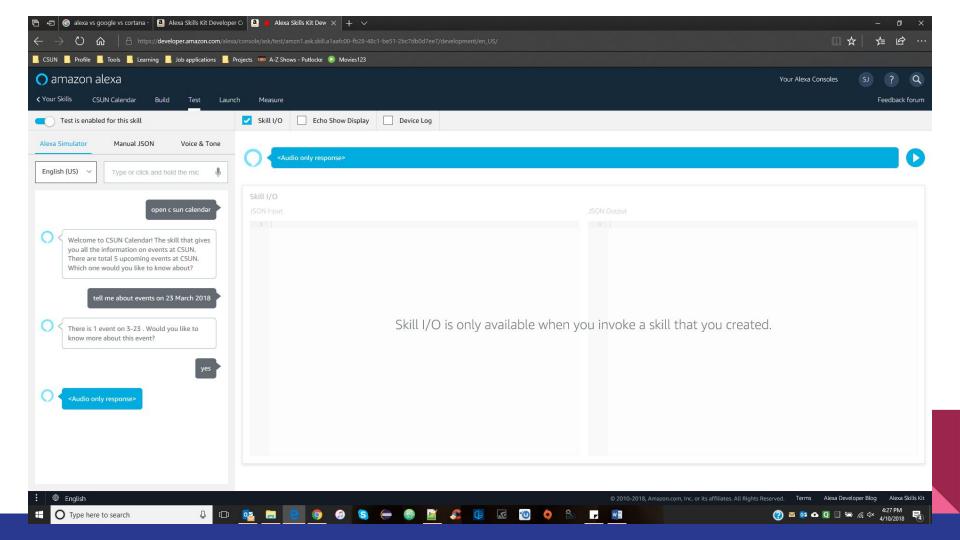
- The Endpoint will receive POST requests when a user interacts with your Alexa Skill.
- The request body contains parameters that your service can use to perform logic and generate a JSON-formatted response.
- Service Endpoint Type:
  - AWS Lambda ARN
  - HTTPS





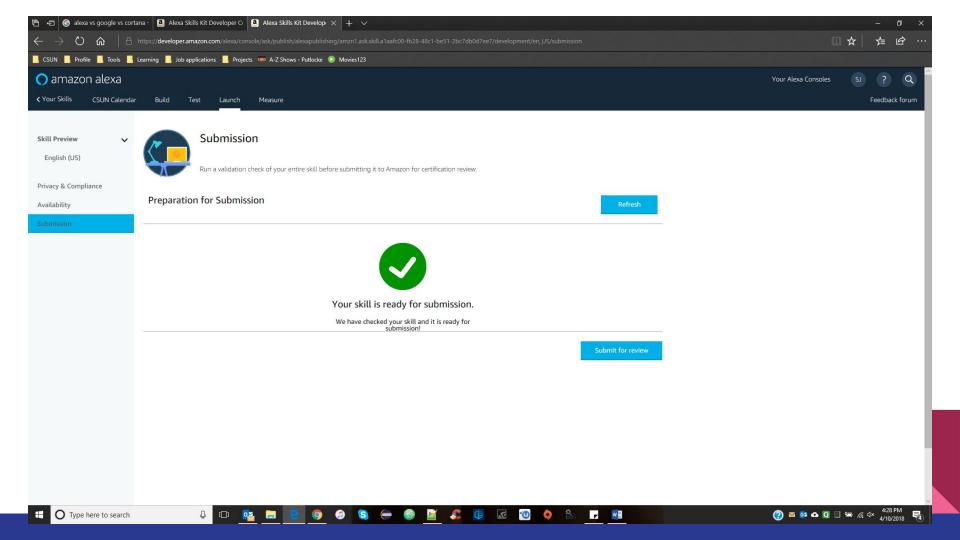
# **Testing**

- Create sample JSON requests and test the cases in ASK.
- Use Alexa Simulator to test the skill using voice-commands or by typing the requests.
- Create JSON request to test the AWS Lambda endpoint in AWS console.
- Use Voice & Tone feature to test Alexa's response.
- Directly test the skill by enabling it on Alexa product. [The skill in development will require the Alexa product logged-in with the developer account credentials]



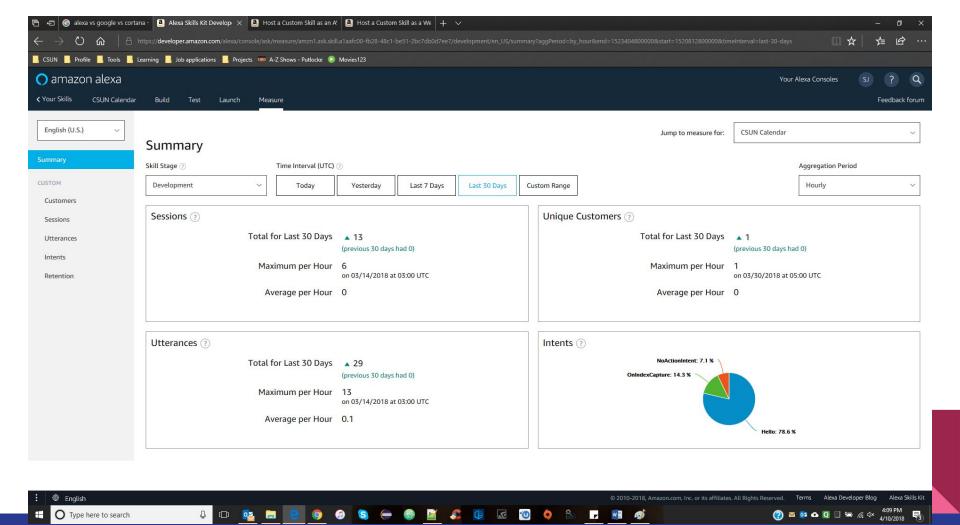
## Publishing

- Skill preview in store:
  - Required: Public name, description, example phrases, small and large icons, category
  - Optional: keywords, privacy policy, terms of use
- Privacy and Compliance:
  - Requirement of purchases, user information, age limit, advertisement
- Availability:
  - Public or Business
  - Global or specific regions
- Submit for review



### **Statistics**

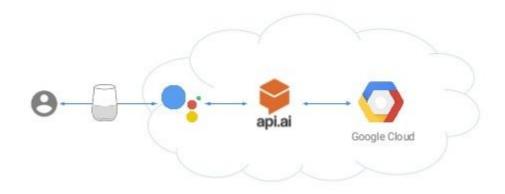
- Alexa Skills Kit provide a feature to measure the statistics of your skill in development or in production.
- It gives a summary regarding :
  - sessions
  - customers
  - utterances
  - intents



### Google: Google Assistant

- The Google Assistant is a virtual personal assistant developed by Google that is primarily available on mobile and smart home devices.
- Google Assistant uses Dialogflow for natural language understanding and machine learning.
- Dialogflow's Actions on Google integration allows you to reach users on voice-activated speakers like Google Home, eligible Android phones and other services.

# Google Assistant: Architecture



### Google Assistant App: Development

- Requirements
- Steps to build
- Testing
- Publishing
- Measure Statistics

### Requirements

- Google Assistant Actions developer account
  - Portal to build and test your skills
- Google dialog flow developer account
  - To host your natural language processing intelligence

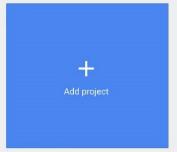
### Steps to build

- Login to actions console
- Select a new project
- Give a name to the app
- Choose your model:
  - Custom (using Dialogflow)
  - Smart Home
  - Templates:
    - Trivia
    - Personality Quiz
    - Flash cards

#### Welcome to Actions on Google

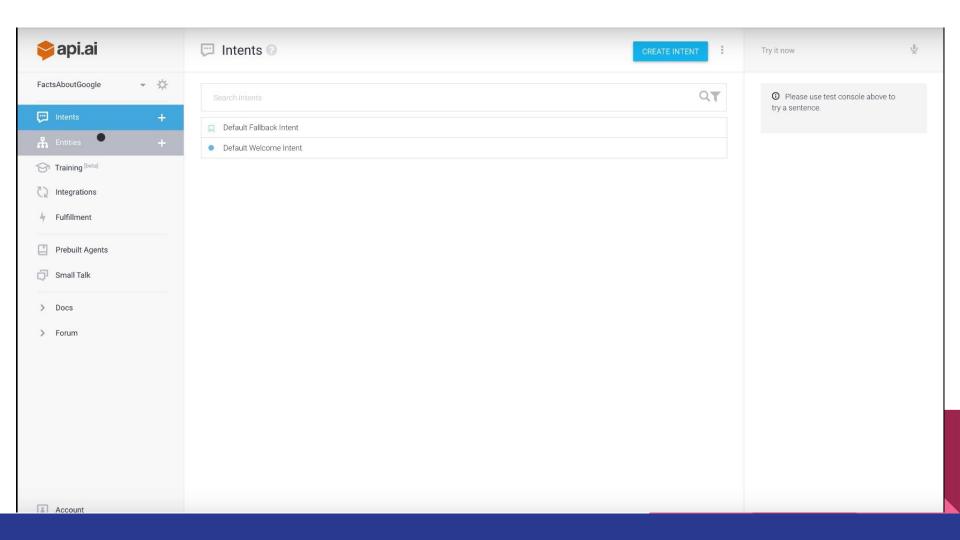
Actions on Google is the platform for developers to extend the Google Assistant. Join this emerging ecosystem by developing actions to engage users on Google Home today, and in the future, on Pixel, Allo, and many other experiences where the Google Assistant will be available. Learn more

Your projects with an Assistant app



### **Conversation Design**

- Select a persona:
  - Personas can range from happy, sad, self-deprecating, formal, and anything in between.
- Write dialogs:
  - Construct the statements that will be used by user as a requests.
- Define intents with entities:
  - Create the intents for the dialogs and add entities as required.
- Define actions:
  - Map the intents to actions/fulfilment (endpoint/webhook).
- Build the model



## **Testing**

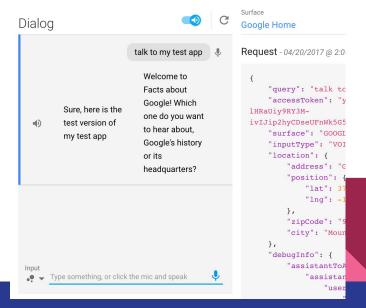
#### Use real devices:

 This lets you experience your apps the way users will and gives you a better idea of how well your app's user experience is designed.

#### Use Actions simulator:

 The simulator also lets you switch virtual surfaces easily (such as a phone or voice-activated speaker) to see how the experience works on different devices and lets you specify user input with text and voice.





## Publishing

- Skill info:
  - Required: Public name, pronunciation, description, example phrases, small and large icons,
     category
  - Optional: keywords, privacy policy, terms of use
- Privacy and Compliance:
  - Requirement of purchases, user information, age limit, advertisement
- Availability:
  - Public or Business
  - Global or specific regions
- Submit for review





Dialogflow actions

App information

English, Assistant app name: My Awesome App, Pronounciation: awesome app, Descr...

Location targeting

All 213 countries (recommended)

Surface capabilities

No required capabilities

5 Account linking (optional)

Set up authentication so that users can link their accounts to your app. Learn more

ADD

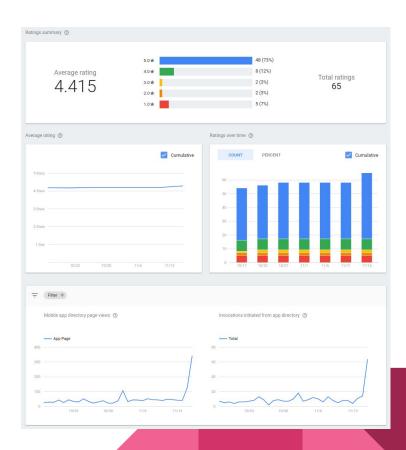
Press TEST DRAFT to test your app in the simulator or on your devices. Submit your draft for review when you are ready to publish your app.

TEST DRAFT

SUBMIT DRAFT FOR REVIEW

### Statistics - Google Analytics

- Once you've published a version of your app, you'll start seeing analytics data in the Actions console within 24 hours.
- These analytics provide information about usage, health, and discovery.

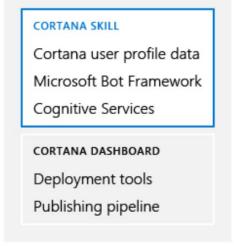


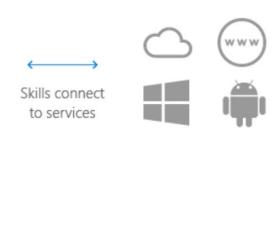
### Microsoft: Cortana

- Cortana and Bing uses Language Understanding Intelligent Service (LUIS) as service.
- LUIS is a machine learning-based service to build natural language into apps, bots, and IoT devices.
- LUIS helps building model by integrating with Azure Bot Service to implement a voice activated system.
- Once the model starts processing input, LUIS begins active learning, allowing you to constantly update and improve the model.

### Cortana: Architecture







CORTANA SKILLS KIT

# Creating a Great Skill from Scratch

The Bot Framework enables deep integration with Cortana's experiences.

Create your skill logic using the Bot Framework

Add the Cortana channel

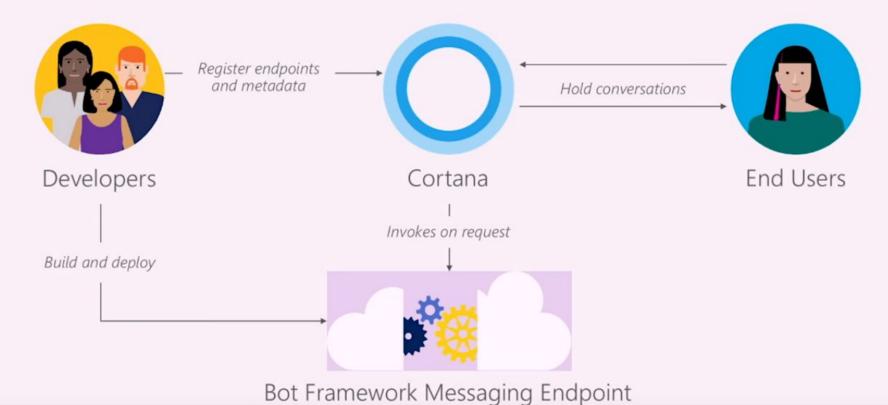
Publish your skill

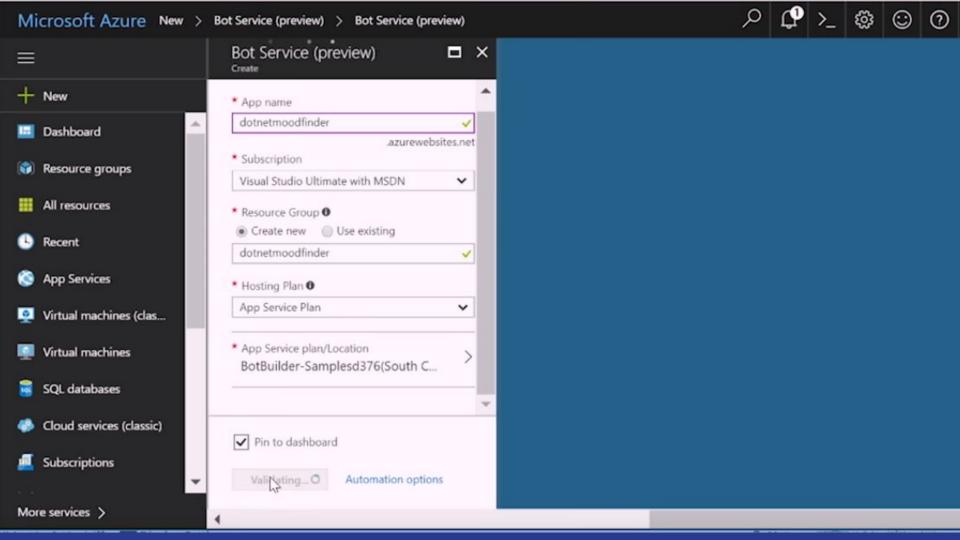
- Build a bot using the Bot Builder SDK
- Use Cognitive Services (LUIS) for natural language understanding

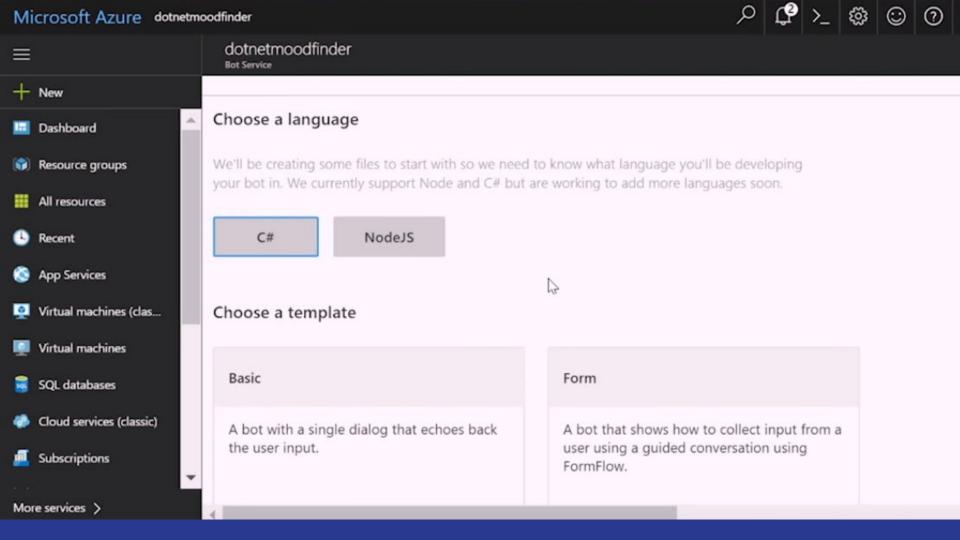
- Register your bot as a Cortana skill
- Configure Cortanaspecific data and features

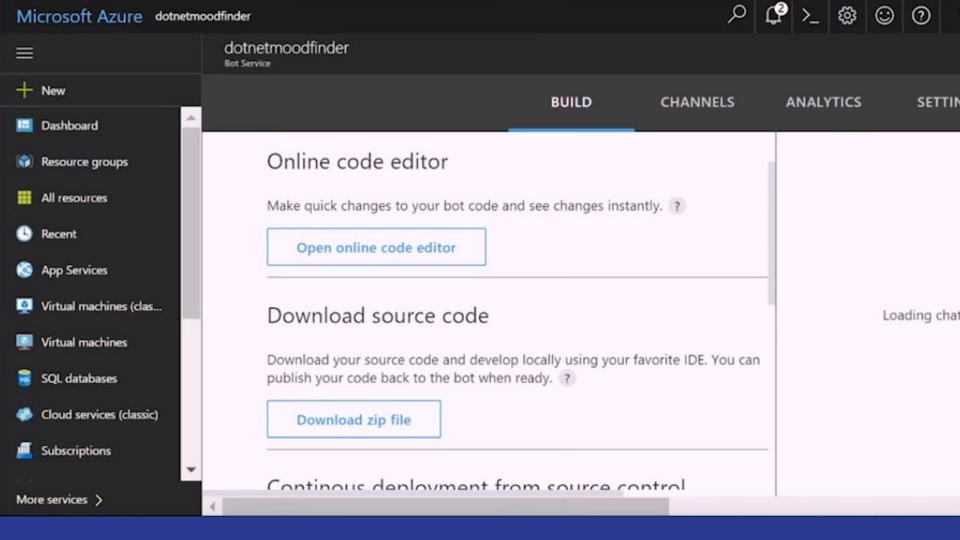
 Make your skill discoverable by all Cortana users

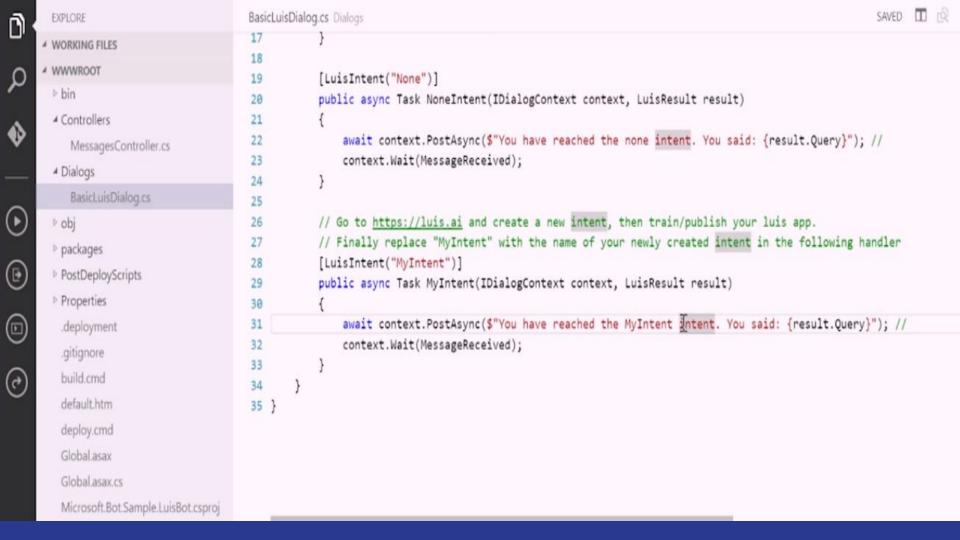
### How do Skills Work?



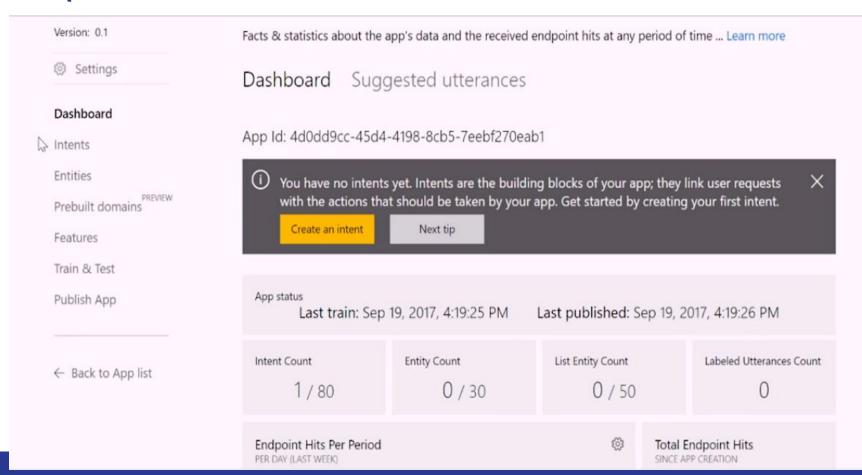








### https://www.luis.ai/home



#### (i) Predicting utterance

#### dotnetmoo...

Version: 0.1

Settings

Dashboard

#### Intents

Entities

Prebuilt domains PREVIEW

Features

Train & Test

Publish App

← Back to App list

#### iviyimtem

Here you are in full control of this intent; you can manage its utterances, used entities and suggested utterances ... Learn more

Utterances Entities in use Suggested utterances





Dashboard

#### Intents

Entities

Prebuilt domains

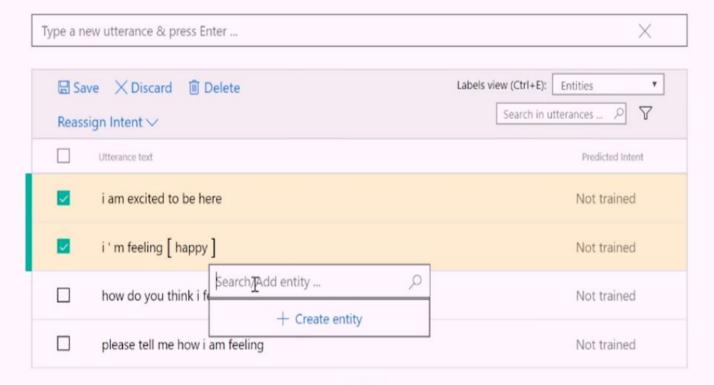
Features

Train & Test

Publish App

← Back to App list

### Utterances (2) Entities in use Suggested utterances



Dashboard

Intents

Entities

Prebuilt domains

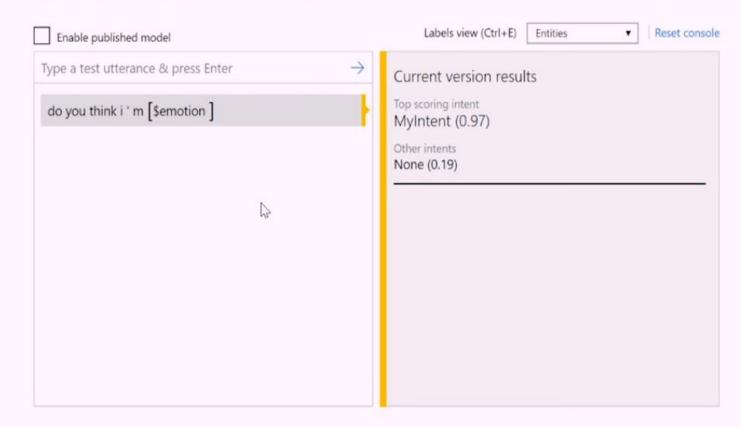
Features

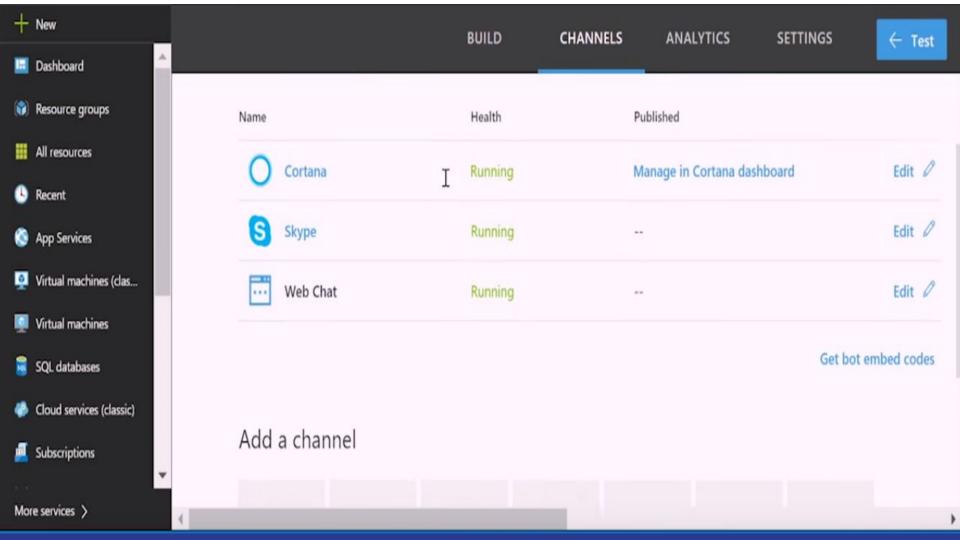
Train & Test

Publish App

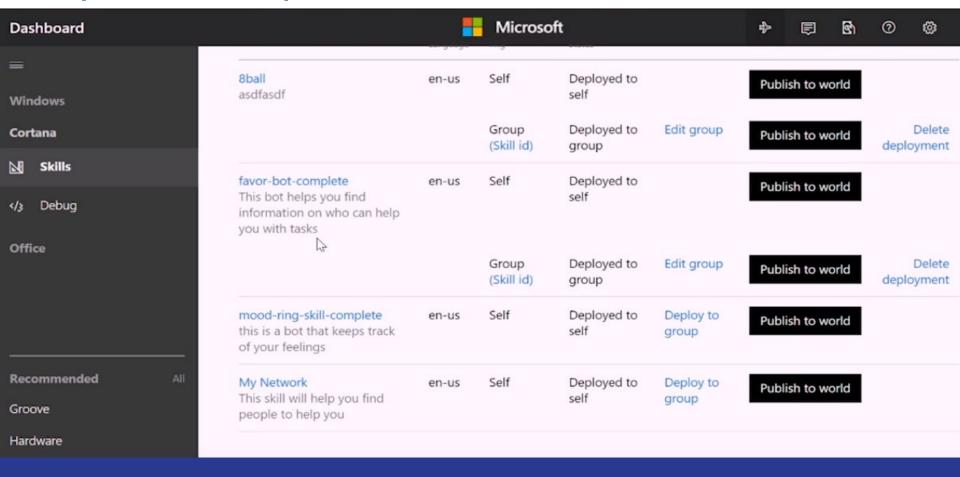
← Back to App list

### Interactive Testing Batch Testing





## https://developer.microsoft.com/en-US/cortana/



## Comparison of Providers

- Support for Languages
- Home automation
- Flexibility
- Integration
- Pricing
- Pros
- Cons

## **Supported Languages**

Google	Amazon	Microsoft
10 Languages:	4 Languages:	8 Languages:
English, Hindi, Dutch, French, German, Italian, Japanese, Korean, Portuguese, Spanish.	English, French, German, Japanese.	English, French, German, Italian, Japanese, Portuguese, Spanish, Chinese.

By year end, Google is willing to release support for over 30 more languages.

### **Home Automation**

Google	Amazon	Microsoft
Full-fledged with Google Services (Youtube, GMaps, etc)	Full-fledged with Amazon Services (Prime, Music, Amazon.com, etc)	Limited home automation      Windows PCs     XBox     Smart Home Hubs     Lighting     Switches, Dimmers, and Outlets     Heating and Cooling

# Flexibility

Google	Amazon	Microsoft
<ul> <li>Primarily available on Android phones and google services.</li> <li>Can be integrated with custom systems.</li> </ul>	<ul> <li>Primarily available on Alexa devices.</li> <li>Can be integrated with custom systems.</li> </ul>	<ul> <li>Primarily available on Windows devices that are Cortana-enabled.</li> <li>Can be integrated with custom systems.</li> </ul>

# Integration

Google	Amazon	Microsoft
<ul> <li>Create apps with devices enabled with Google assistant.</li> <li>Additional cost for custom integration.</li> </ul>	<ul> <li>Create skills with devices enabled with Alexa.</li> <li>No additional cost for custom integration.</li> </ul>	<ul> <li>Create skills with devices enabled with Cortana.</li> <li>Additional cost for custom integration.</li> </ul>

# Pricing

Google	Amazon	Microsoft
<ul> <li>Google's dialog flow is free if used to develop apps using Google Assistant skill kit.</li> <li>To use dialog flow API,         <ul> <li>Free: limit to 15000 requests per month</li> <li>Enterprise: unlimited request per month; with \$ 0.0065 for every 15 secs</li> <li>Maximum audio length is 60 sec for both</li> </ul> </li> </ul>	<ul> <li>No charges for Alexa skill or Alexa SDK integration into custom devices.</li> <li>Charges are incurred for AWS Lambda endpoint as follows:         <ul> <li>\$ 0.20 per 1 million request, monthly {Free 1M request per month}</li> </ul> </li> </ul>	<ul> <li>LUIS has two editions:         <ul> <li>Free: 10,000 requests per month</li> <li>Basic: \$ 1.50 per 1000 request with 50 requests per sec</li> </ul> </li> </ul>

### Pros

Google	Amazon	Microsoft
<ul> <li>Support for more languages.</li> <li>Primarily available with Google services.</li> </ul>	<ul> <li>Easy integration with Alexa skill kit and Alexa SDK for custom devices.</li> <li>Less cost for resources.</li> <li>No charges for custom integration.</li> </ul>	Support for Chinese language.

### Cons

Google	Amazon	Microsoft
Incurs charges to use API for custom integration.	Fewer language support.	<ul> <li>Limited device support.</li> <li>Incurs charges to use API for custom integration.</li> </ul>

# Questions?