

The logo for AWS re:Invent, featuring the text "AWS re:Invent" in white. "AWS" is in a smaller, sans-serif font, "re:" is in a medium-sized sans-serif font, and "Invent" is in a large, bold, sans-serif font. The background is a dark blue gradient with a subtle, abstract pattern of vertical lines.

AWS re:Invent

The name "Rohit Prasad" in white. "Rohit" is in a large, bold, sans-serif font, and "Prasad" is in a medium-sized, sans-serif font. The background is a dark blue gradient with a subtle, abstract pattern of vertical lines.

Rohit Prasad

Vice President & Head Scientist
Amazon Alexa



Captain kirk talking to a computer without any interface and just by voice is an iconic image for voice computing as we know it. This is ***the holy grail of Conversational Artificial Intelligence***

It's not science fiction.

It's science fact.



Amazon has started building a computer that can installed into everything and made available everywhere, just by voice using Alexa, echo and other devices.



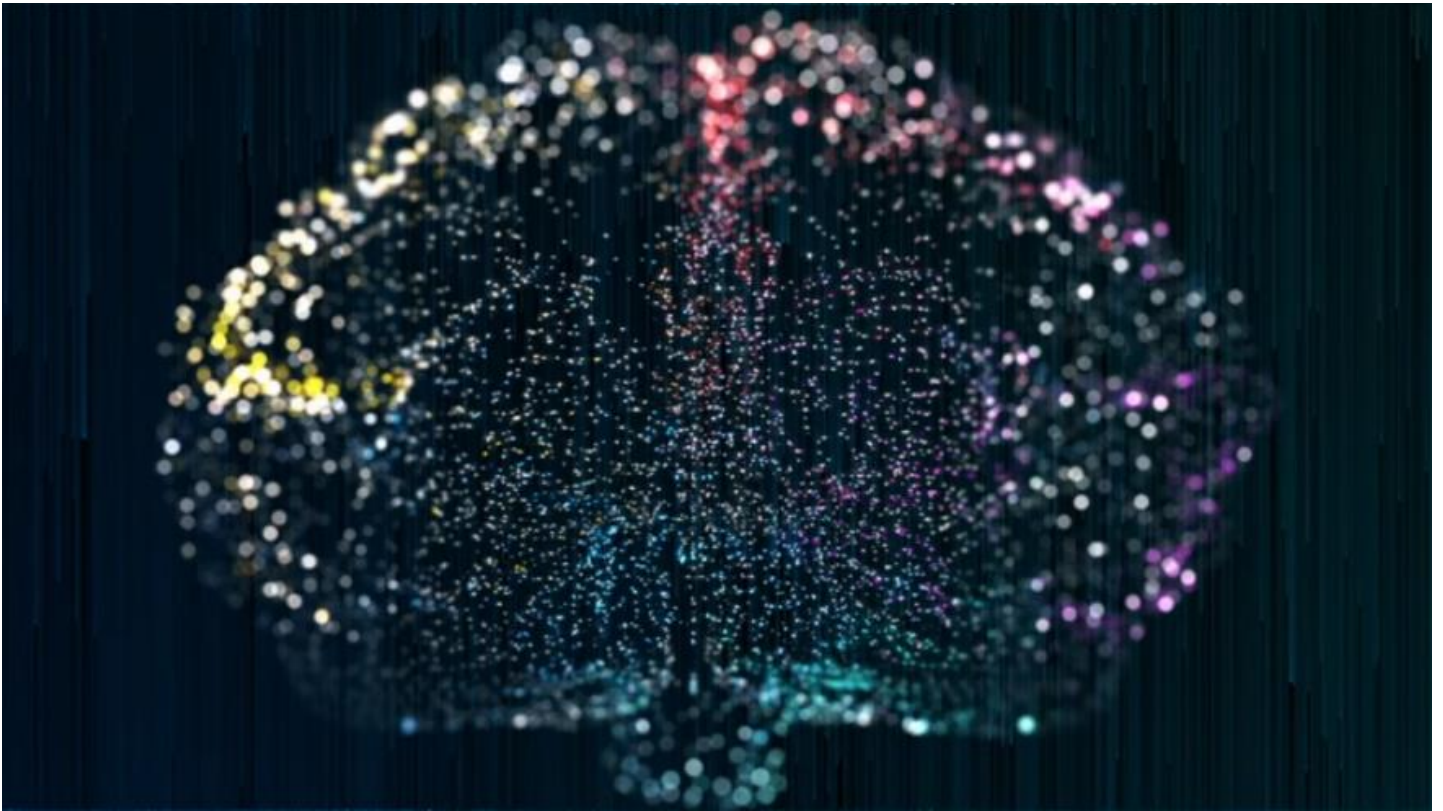
Amazon and other developers have added hundreds and hundreds of new skills to Alexa over the year. You can now call your family using ‘call mom’



You can now play music anywhere in the house using Alexa



There are now thousands of skills added using the Alexa Skills Kit, for pizza, hailing a ride, etc.

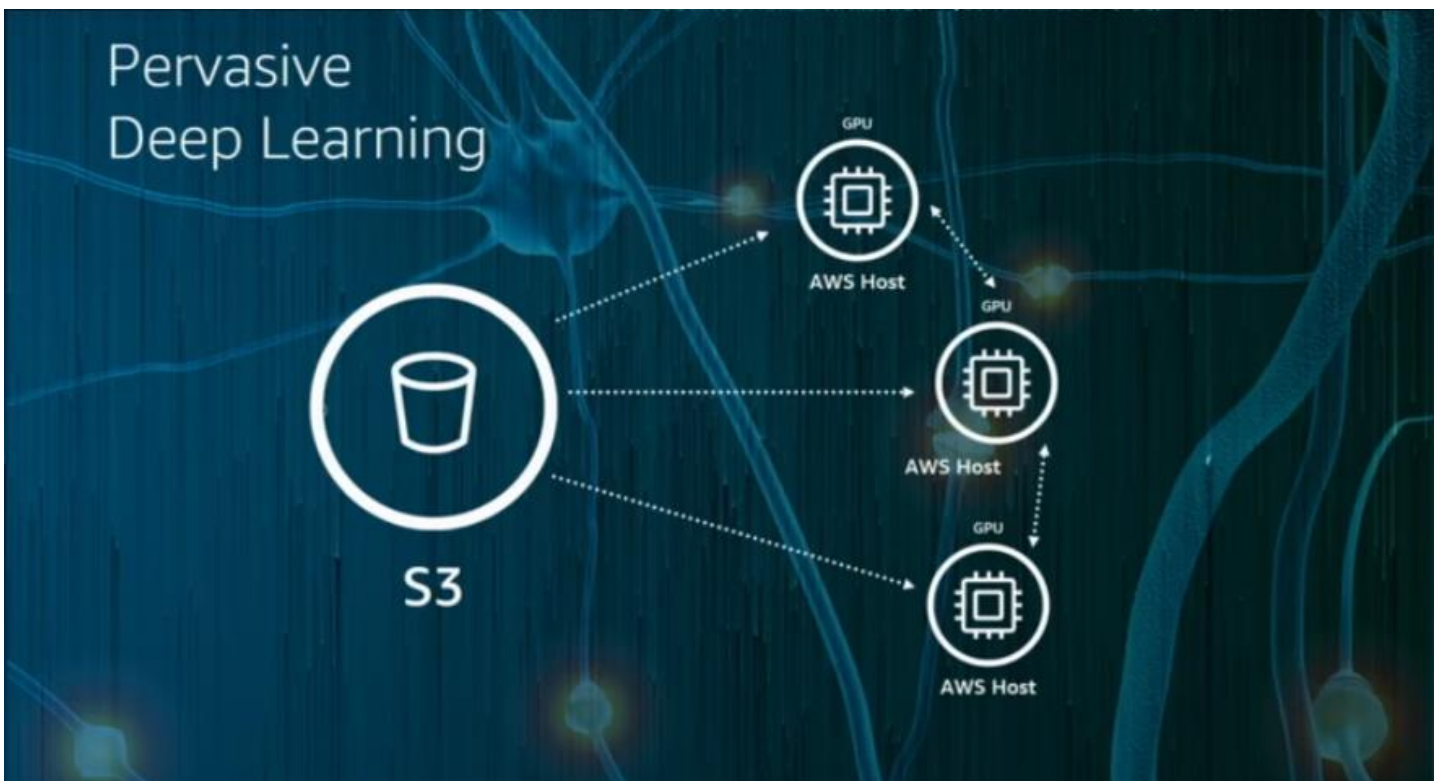


Let us now take a look at how we have been making Alexa smarter



This is a simplistic view of the **Alexa architecture**. Alexa listens for the **wake word recognition** which is 'Alexa', when it is confident the wake word was spoken it opens up an audio stream and the lights come up on the LED ring on the device, then the audio starts streaming to the cloud where what was said as audio bits gets converted by **Automatic Speech Recognition ASR** into sequence of words.

The next stage is for **Natural Language Understanding NLU**, to decipher/interpret meaning from these words and extract meaning from the utterance to decide what particular sets of skills are needed. This about figuring out the intent of the user and then issuing a directive for that intent like 'to play some music'.



Active Learning for Spoken Language Understanding

Compressed Time Delay Neural Network for Wake Word

Contextual Interpretation Reranking for Natural Language Understanding (NLU)

Domain-Specific Utterance End-Point Detection

Faster Large Vocabulary Conversational
Speech Recognition (LVCSR) Inference Engine

Faster Training and Inference in Long Short-Term Memory (LSTM)

Robust Speech Recognition via Anchor Word Representations

Unsupervised Speaker Adaptation via i-vectors

Transfer Learning for Neural Semantic Parsing

Zero-Shot Learning Across Heterogeneous Domains

These are the fundamental technology capabilities that AWS has been adding to advance the field of voice recognition already. Active learning is a technique in which Alexa itself figures out what things to learn and become better on its own.



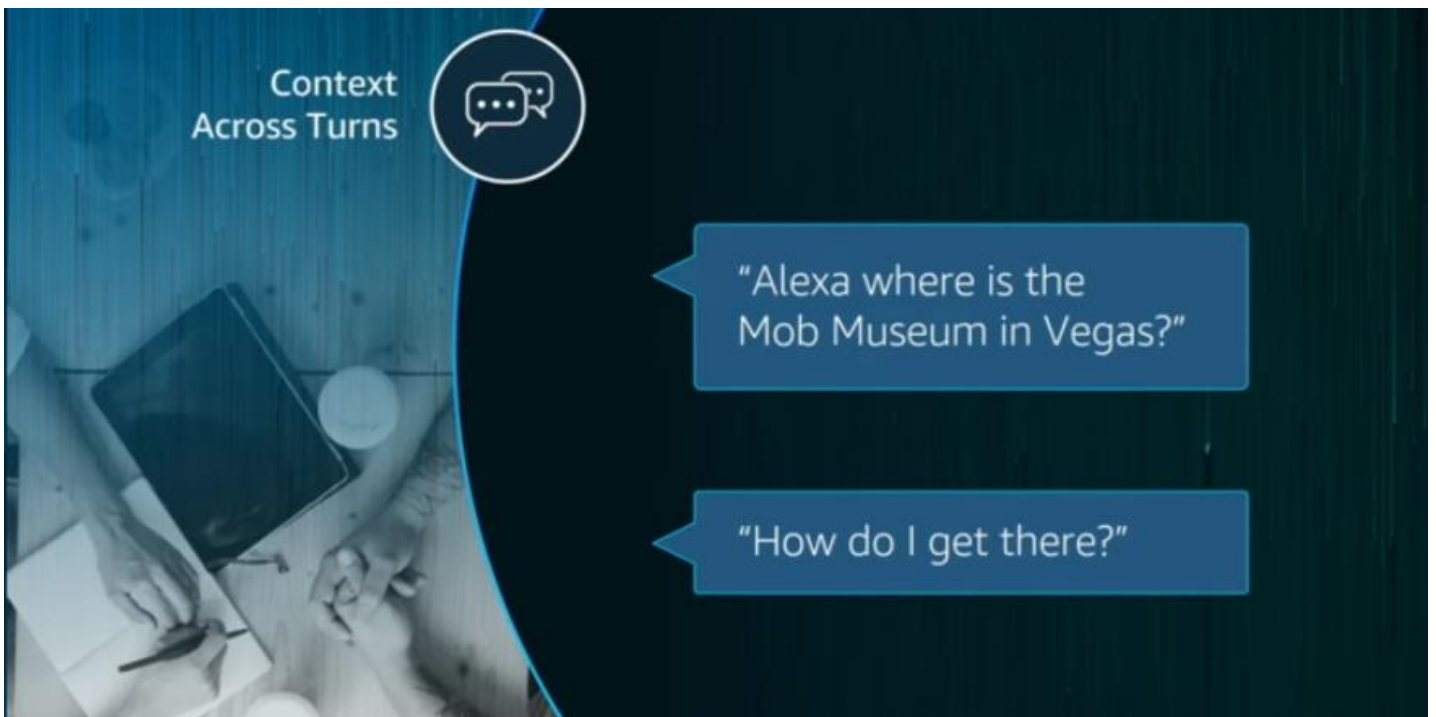
Alexa voice service can be put in different devices

Deeper Contextual Learning

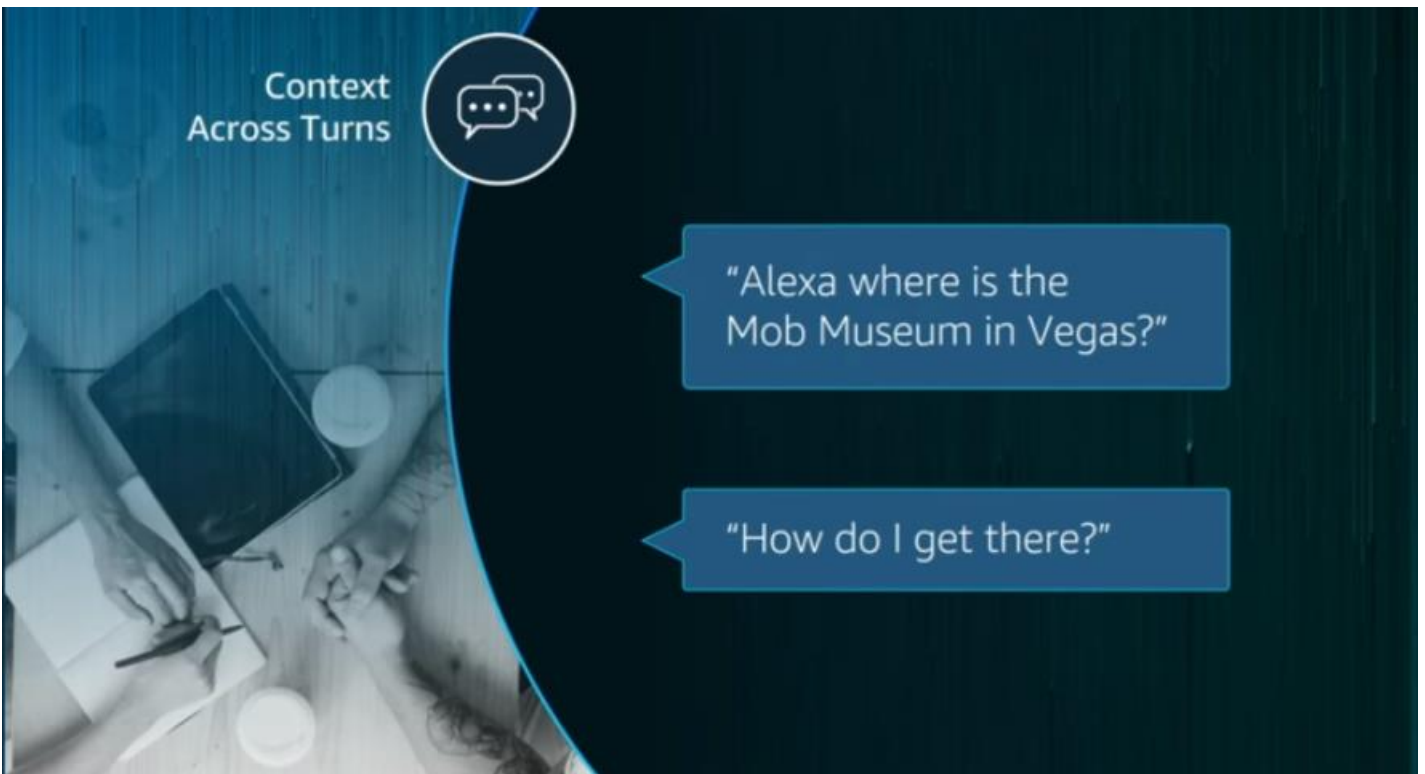
Context comes in many forms



Humans ask follow up questions to previous dialogues, like 'where is a museum', 'how can I get there?'



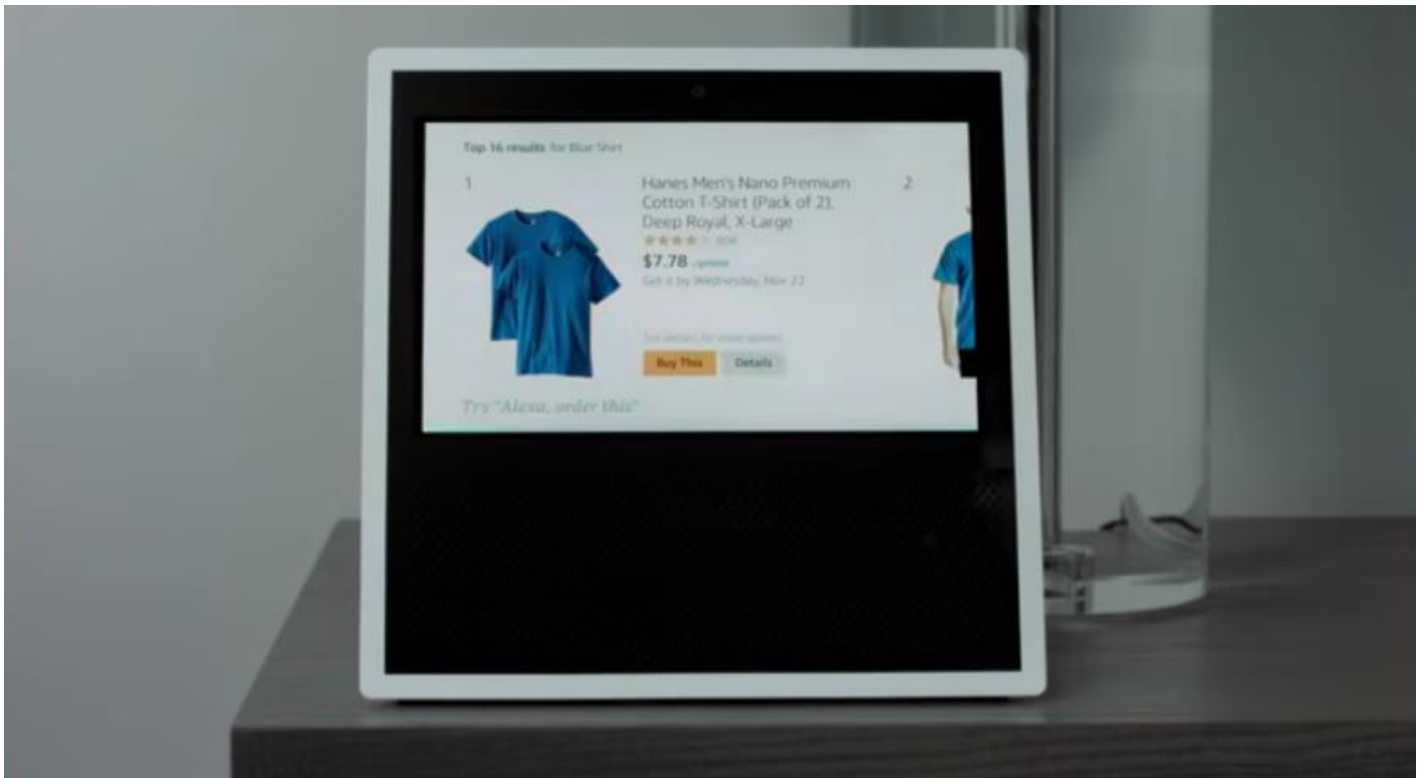




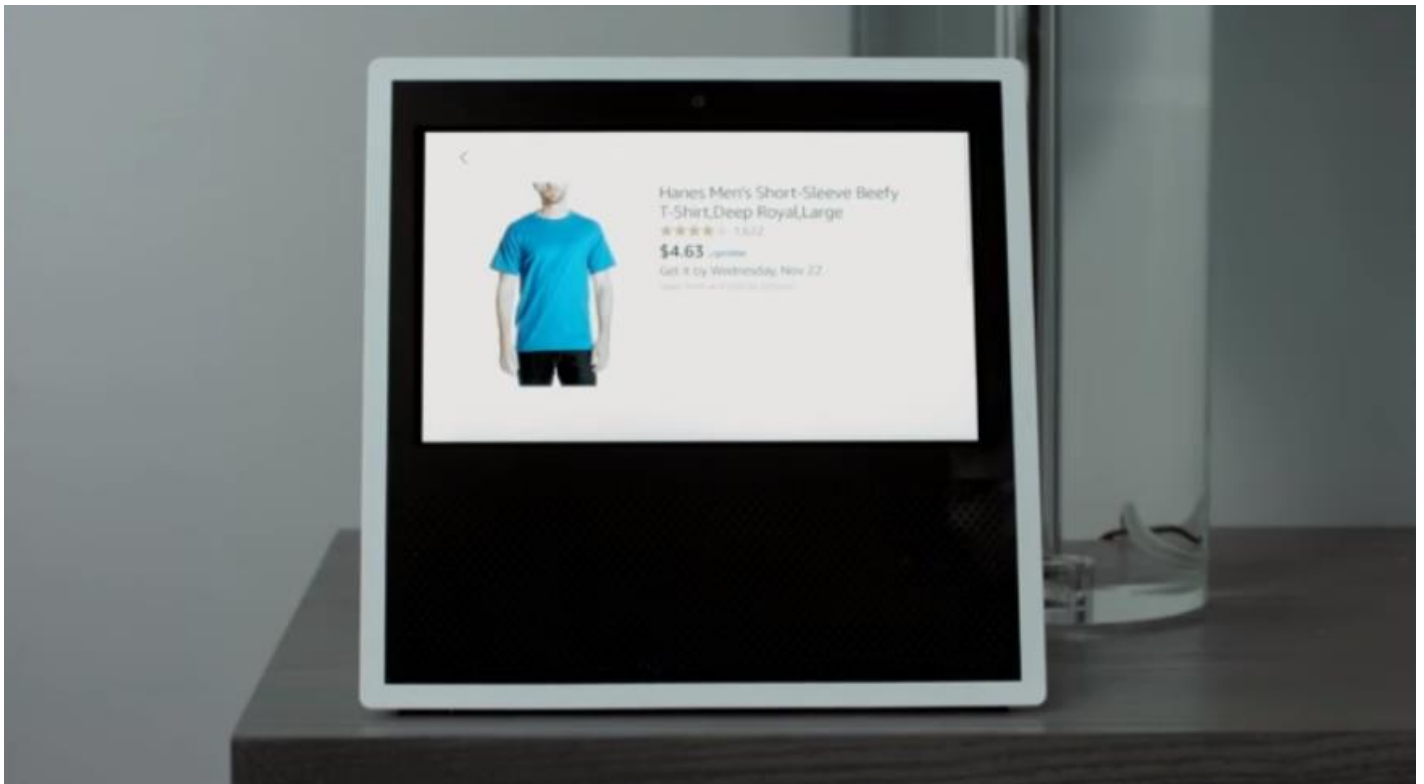
Note that there were no specifics about the Mob Museum in the 2nd utterance, this happens when we do natural dialogue. The way we solve this problem is to carry the intent from the 1st utterance to the next utterance in a probabilistic approach where we can merge the intent from the previous utterance to make a new decision.



When we introduced the **Echo Show** which has a camera and a screen, the dynamics of interacting with Alexa changed. *Let us see below an example of how a user interacts with echo Show leveraging the camera and the screen*



'Alexa, show me a blue shirt' ...here, there is an intent to buy a blue shirt. Then 'Alexa, select the second one'



'Alexa, show more'



You now get more information



This new dimension is about the fact that Alexa now needs to understand what is being shown on the screen, interpret it, merge it with the current utterance, to generate the new context, intent, and decision. The interaction has now changed into a multimodal context.



The 3rd type of context is that Alexa now needs to be aware about what each device it is inside is capable of, this is the evolving device context.



‘Alexa, play the last city of Zi’...Alexa knows this is an audio book since the user does not have a screen and says ‘getting you a book from Audible, resuming the last city of Zi’...then it starts from where you left off from the audiobook



Now if another user that has Alexa connected to their FireTV also says ‘Alexa, play the last city of Zi’...Alexa says ‘getting the last city of Zi from FireTV’



And starts playing the movie from where they stopped. Alexa is context aware that it is connected through an Echo device to a fireTV and can dynamically change its interpretation from not playing the audiobook to playing the movie instead.



Alexa is a device in a home that lets multiple people interact with the same Alexa device, Alexa can separate who is speaking and then customize its reaction to them individually. This feature is called 'Your Voice', this can be used in things like auto-signing a text message with the user's name if Alexa knows them, you can get your own playlist. Users can enroll into Alexa. You only need to set up your enrollment profile once since the speaker recognition is happening in the cloud. Then there is no pass phrase needed once you do this, this is called 'Text Independent Recognition' using feature set vectors.



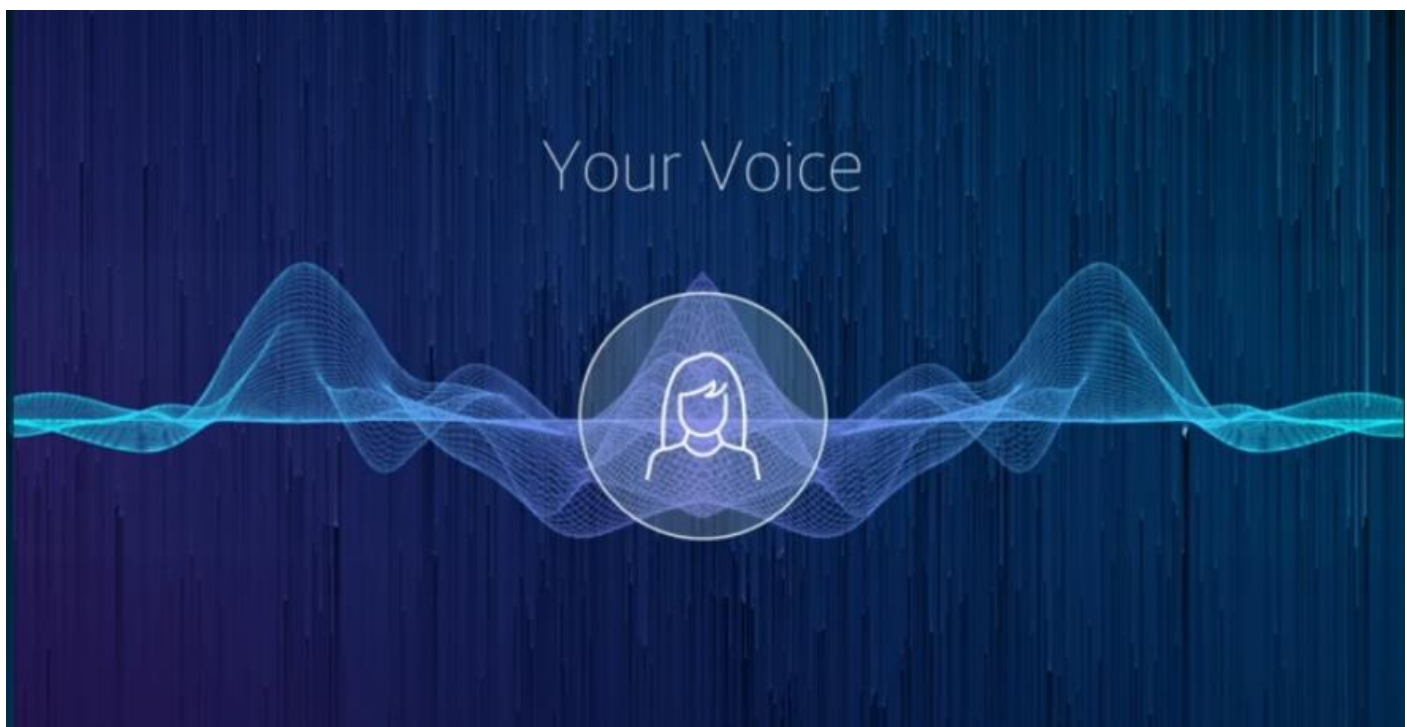
Developers can use 2 sets of APIs. The **Alexa Voice Service (AVS)** lets you integrate Alexa into your devices and applications with a simple set of APIs. The **Alexa Skills Kit (ASK)** lets you to build unique skills on Alexa using a set of APIs available to you.



There is a skill for every occasion now being built by developers




Weekly active users for skills is growing exponentially too



The 'You Voice' feature is a speaker recognition capability to personalize your interactions to the device. Developers can now use this feature in their own skills to further personalize their actions to user responses

Notifications

When can Alexa be more proactive for things like getting notifications after shopping for something? The developer preview for the Notifications feature is being expanded

A photograph of a family consisting of a mother, a young boy, and a father, all smiling and engaged in playing a board game on a table. The image is partially obscured by a dark blue curved overlay on the right side.

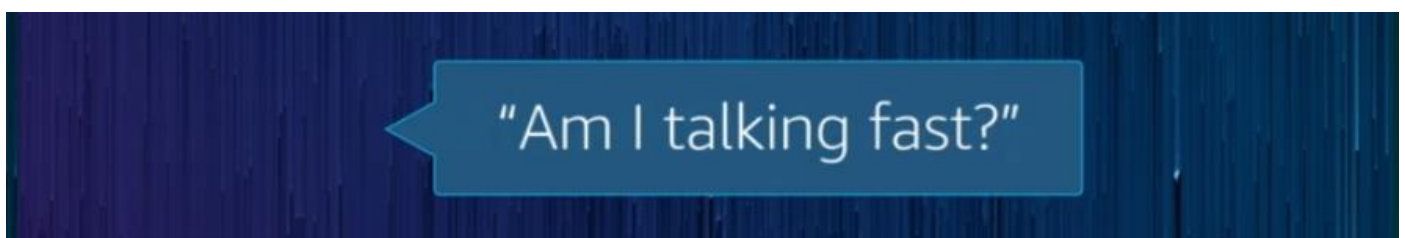
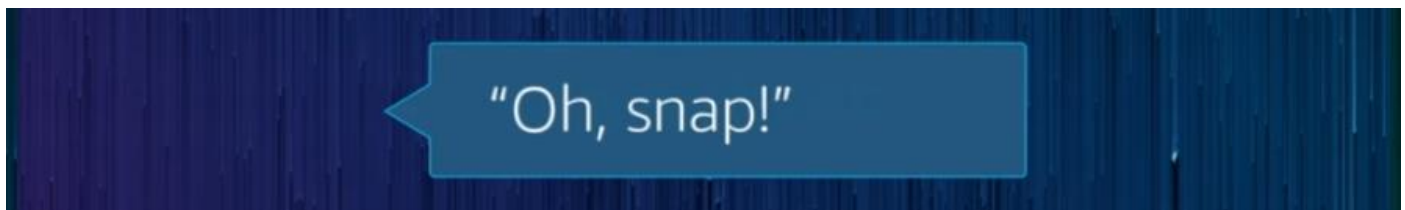
Skills made **easier, faster, magical**



There have been many things done to allow developers build more skills easily and quickly. The skill builder tool has a large library of built in semantic elements, prebuilt intent and entity types (like food items) that you can use in any skill you are building. There are also skill samples in our repos.



Speechcon is our speech-to-text engine





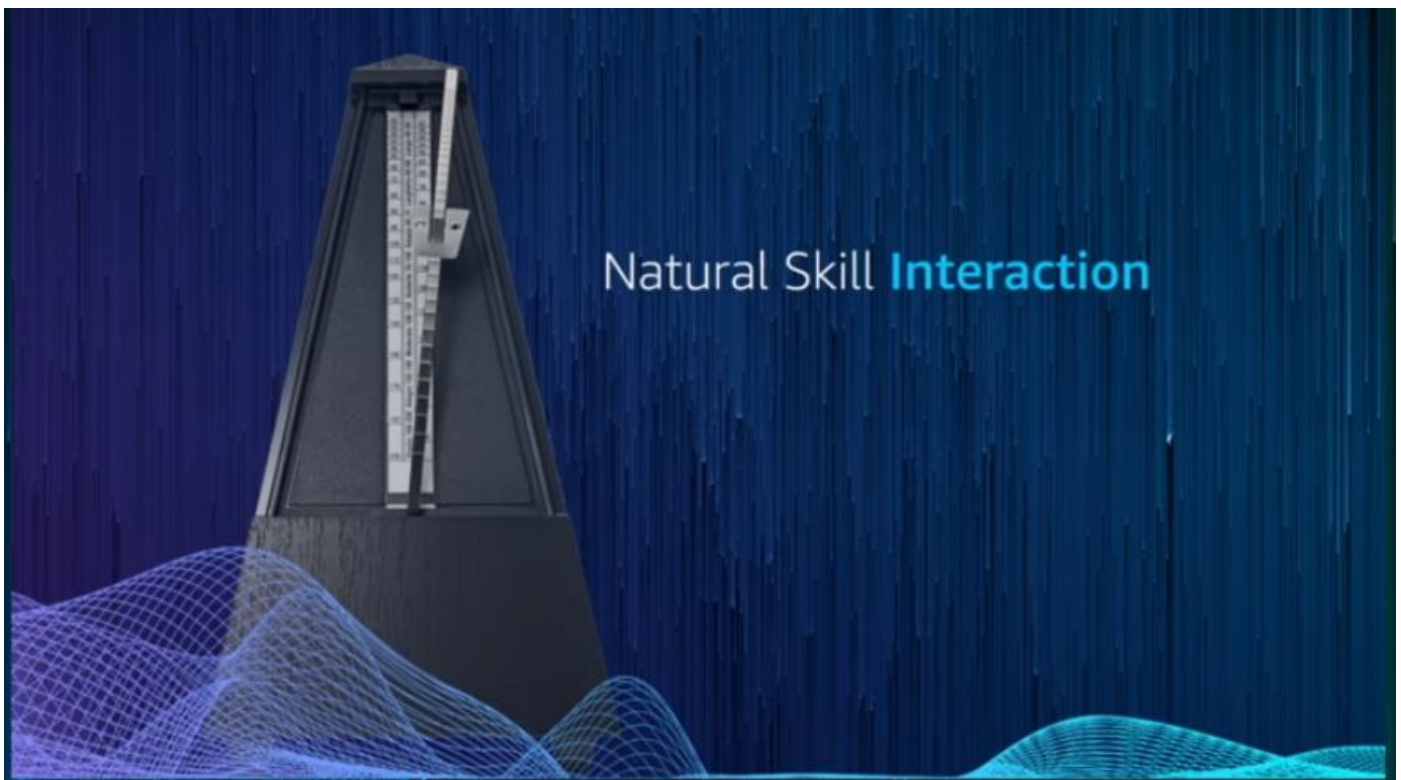
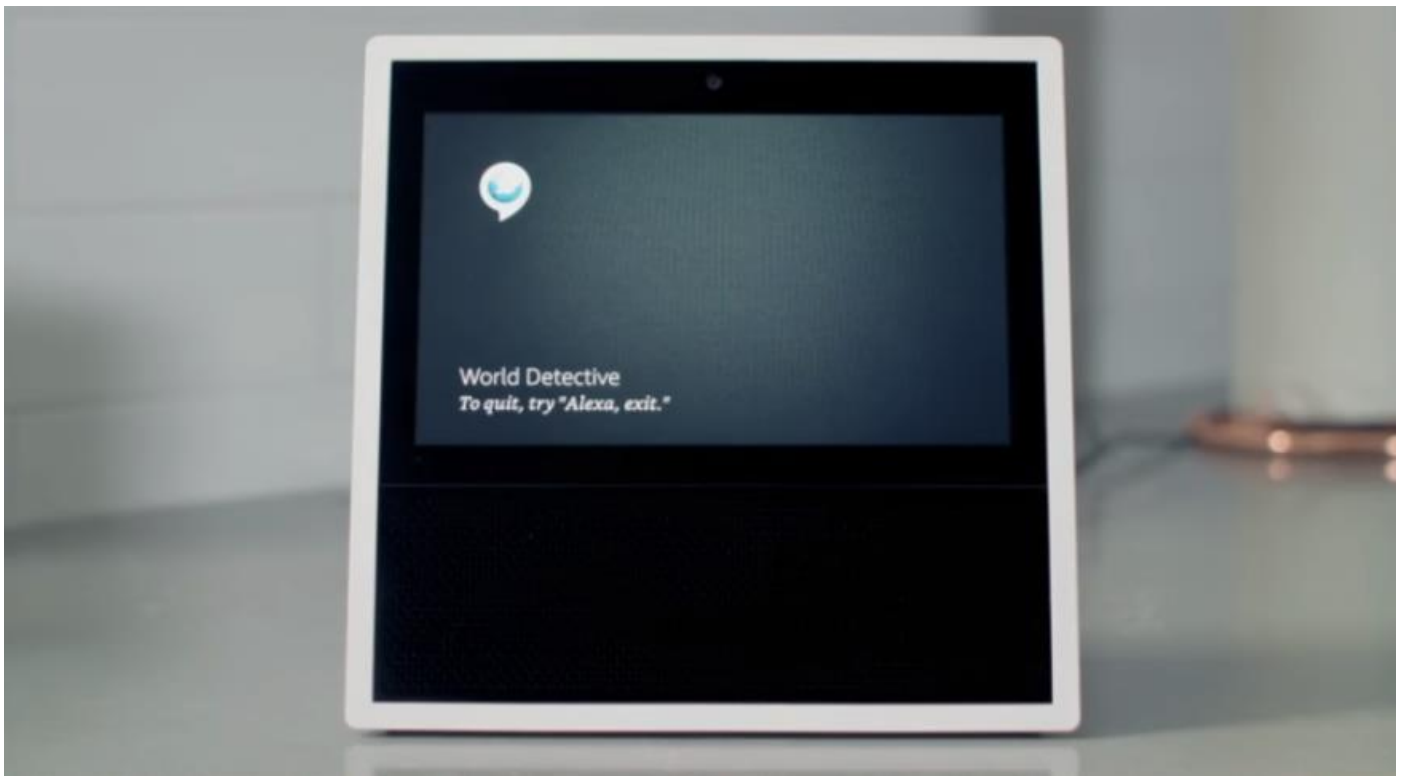
Skill **Discovery**

There are several skills in the skills store now. You can just say 'Alexa, plat jeopardy' and the jeopardy skills will be downloaded and you can play it. You don't have to say 'Alexa, enable jeopardy' anymore to get that skill on your Alexa device.



"Alexa, let's play a game."







‘Alexa, play me beats at 90 beats per second’

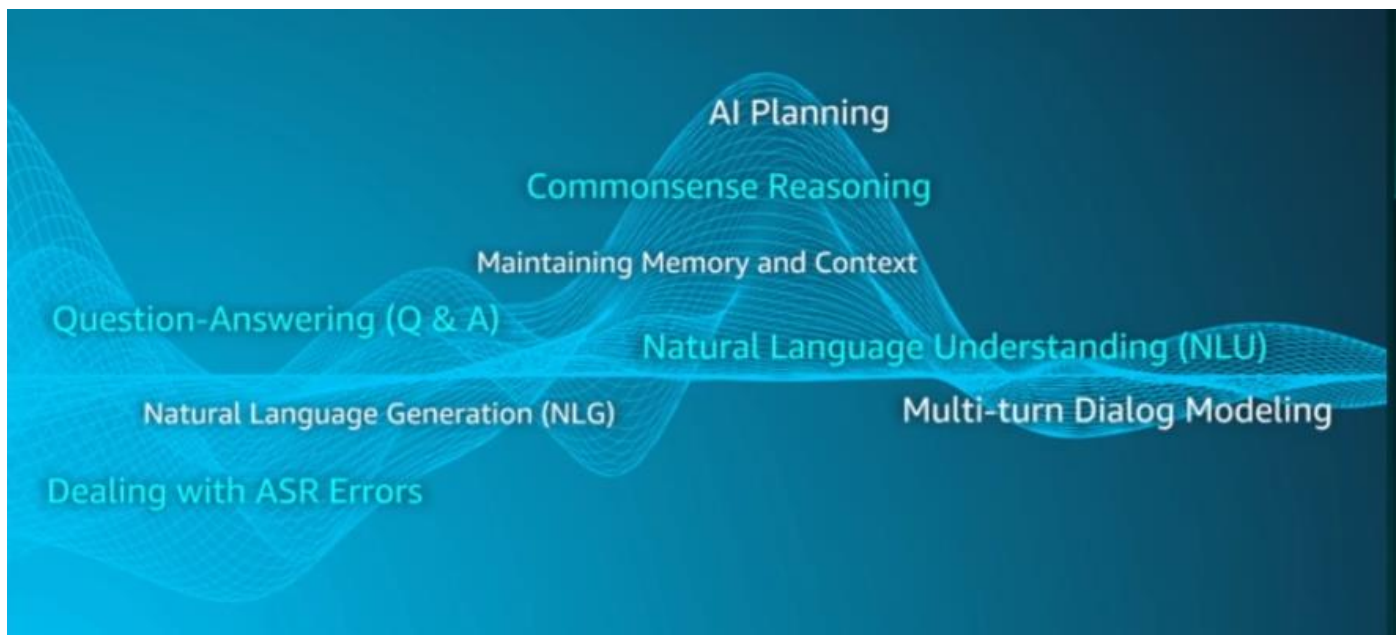


We are using deep learning models that can select a subset of the skills that can answer a question, then we can rank and select the best skill to use

amazon alexa prize



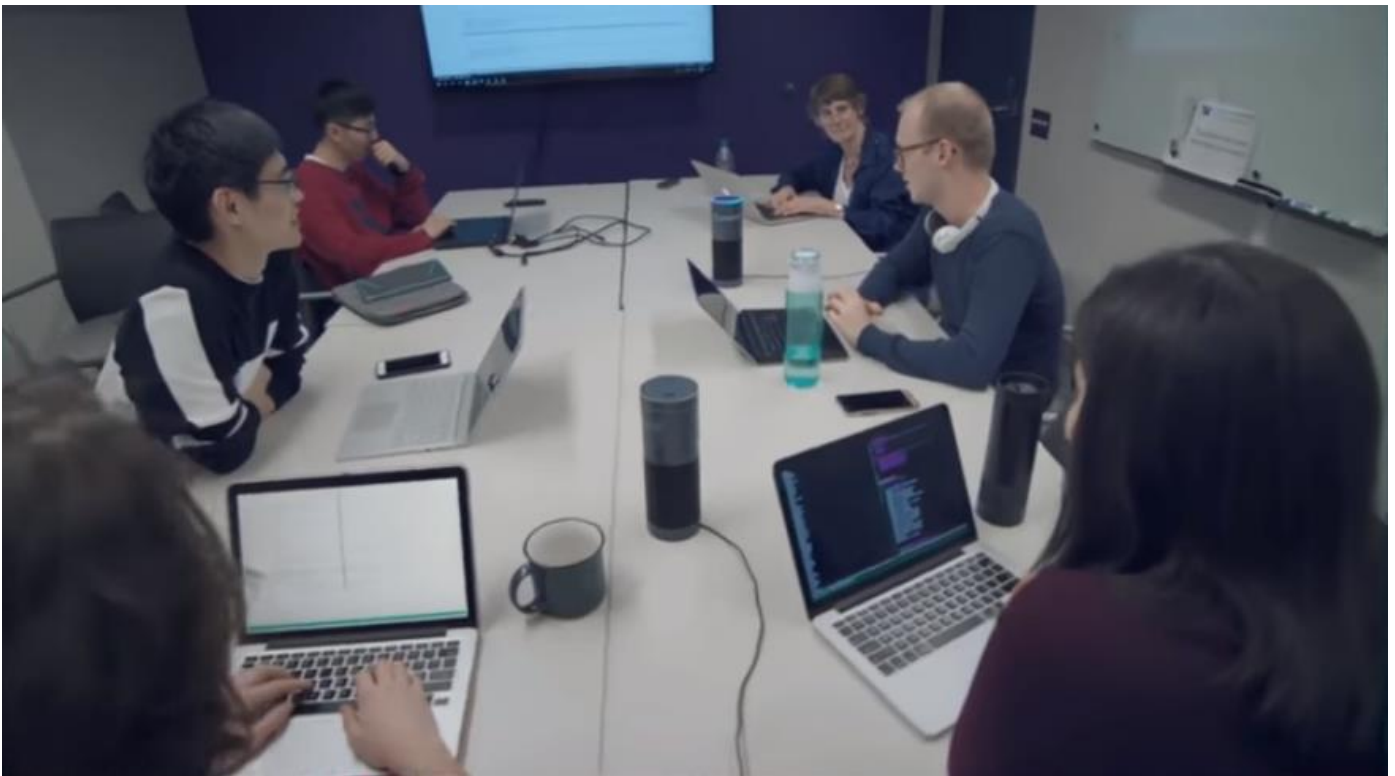
Competition to create Alexa skills for social interaction with customers in conversation that can last 20 minutes. This problem involves developing conversational skills that are robust to errors. Using 'Alexa, let's chat'.

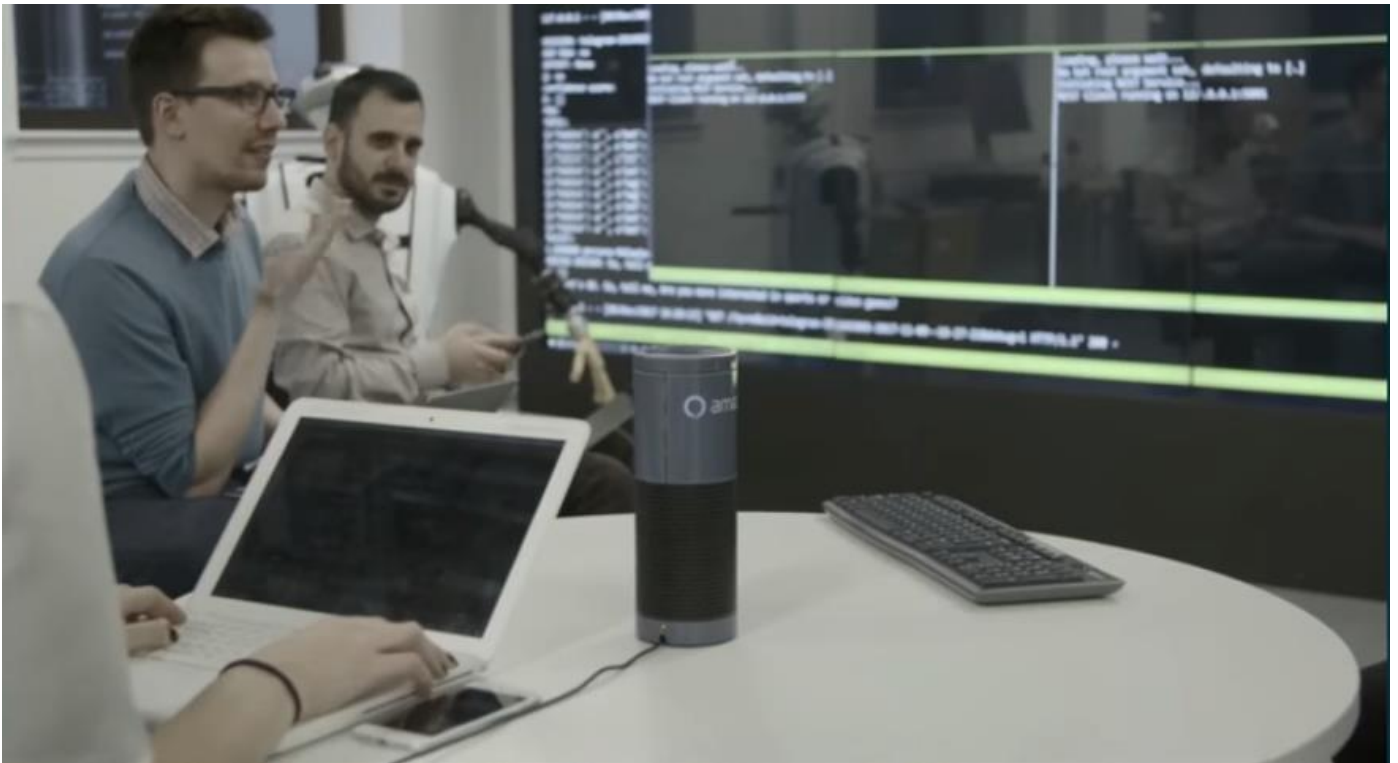


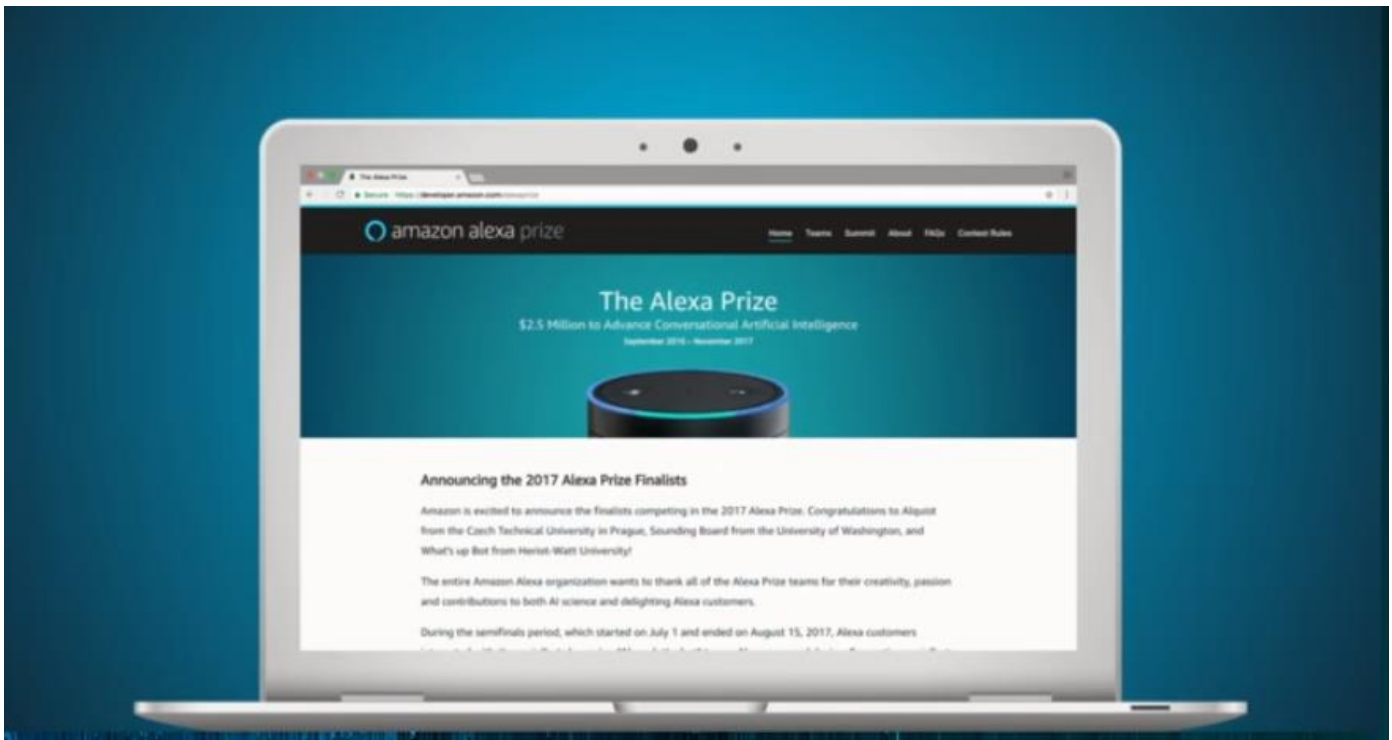
The social bots have to stay in touch with the current news everyday in order to interact with customers.



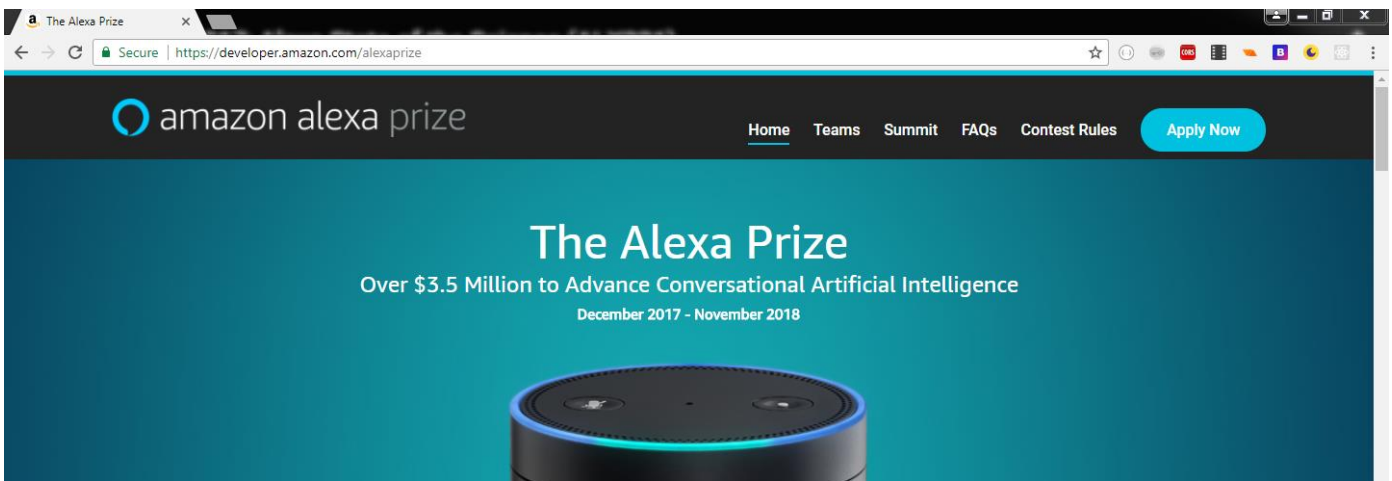
 amazon alexa prize







Is a research testbed for universities and other researchers to push the boundaries of Conversational A.I, NLU, Dialog design framework, speech recognition fields?



Applications are open for the 2018 Alexa Prize. Go [here](https://developer.amazon.com/alexprize) to apply now.

Congratulations!

Amazon is excited to announce the winners of the 2017 Alexa Prize.

Winner: Team Sounding Board, University of Washington, Seattle, WA, USA

With an average score of 3.17 and average duration of 10 minutes and 22 seconds, Team Sounding Board will

<https://developer.amazon.com/alexprize> contains technical papers by each team in the competition.



Deep Dive Session and Finals Film Screening


1:30 p.m. @ JG Steakhouse

Panel discussion with
Alexa Prize finals participants

1st

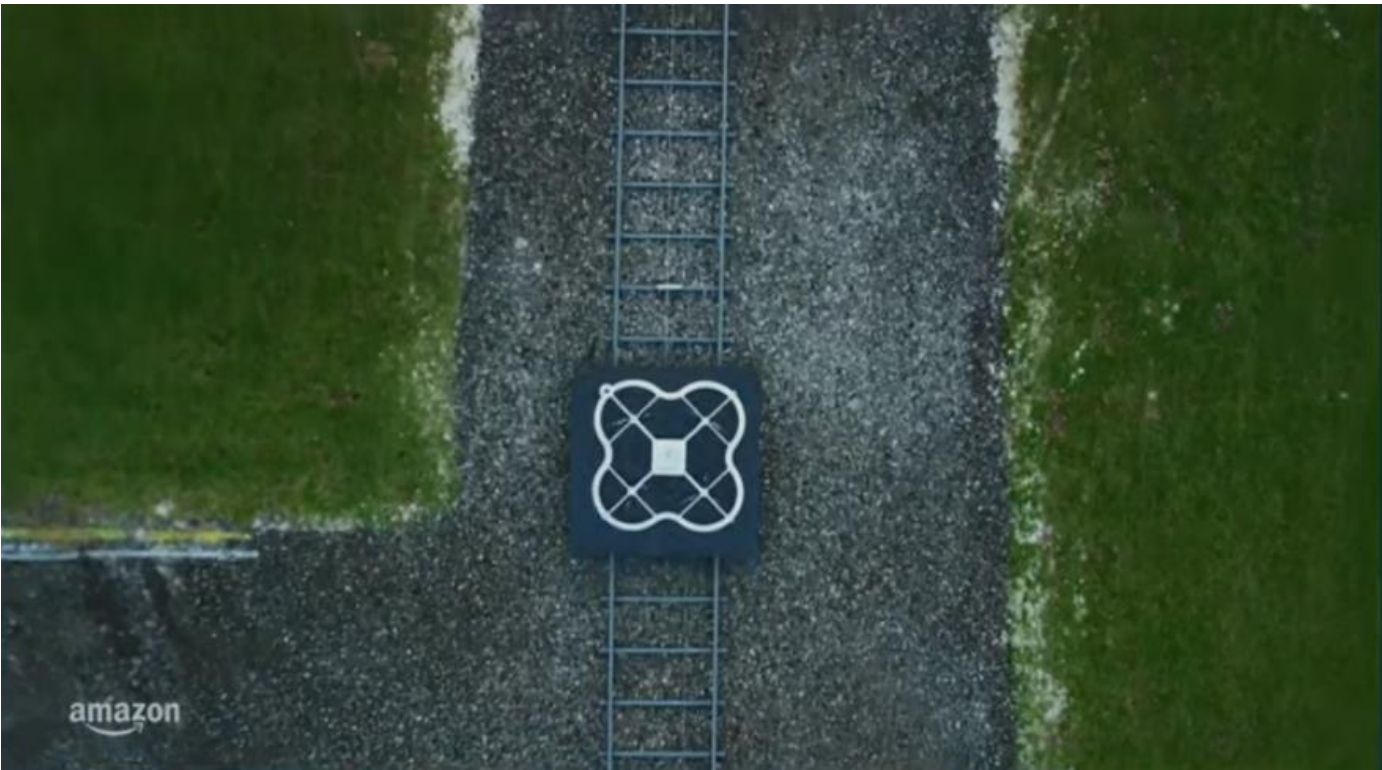


W
UNIVERSITY *of*
WASHINGTON

 amazon alexa prize

www.alexaprize.com

amazon

amazon
Prime Air

ACTUAL FLIGHT FOOTAGE
NOT SIMULATED





ACTUAL FLIGHT FOOTAGE
NOT SIMULATED



ACTUAL FLIGHT FOOTAGE
NOT SIMULATED









echo show

amazon

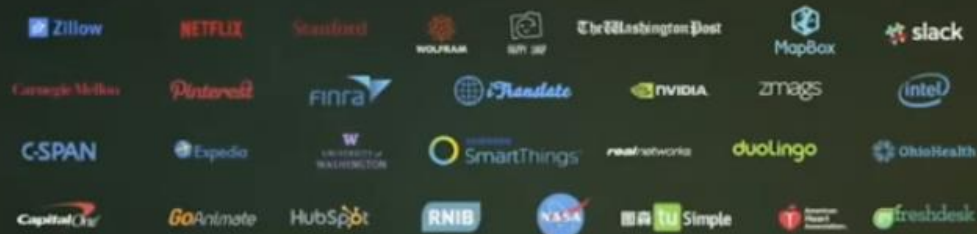




amazon**machine**learning



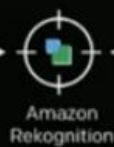
Some Of The Customers Using Amazon AI In Production Today



amazonrekognition



Amazon Rekognition: Images In, Categories and Facial Analysis Out



Objects & Scenes
Car
Outside
Daytime
Driving

Faces
Female
Smiling
Sunglasses



Rekognition: Facial Search



amazonsearch

