

You have already missed five turns. Why don't you ask Siri for directions?
Sophia will be late to her competition!





How does Siri work?

Let me explain you how Siri works.
That may help you in the competition

Concept # 1



Let's find out, how does
Siri record your voice
on iPhone and convert it
to digital format

Try this ...

Take your hand close to your mouth and say the word “**Boom.**” What did you experience?

When you speak, pressure waves get created and the microphone uses these **changes in air-pressure to record sound.** Your eardrum sort of works on a similar principle

Show me the waves..

Below is a typical sound wave, and in this case, I was asking Siri directions

How Long Is the Drive from Columbus to New York

Siri Not Available

You are not connected to the Internet

After this step, Siri converts your speech to text. The voice conversion requires a lot of computing power, and hence your voice is sent to cloud. That is the reason why Siri doesn't work without internet

Recap..

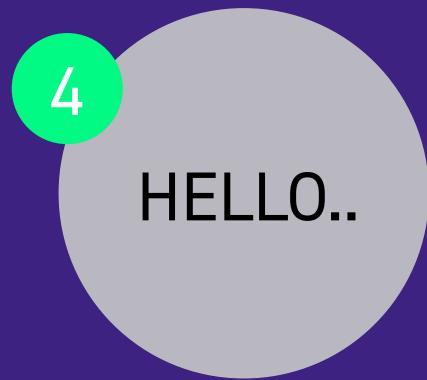
Let's see how Siri converts speech to text



you say "Hello"



Siri records your voice



speech converted to text



your voice is sent using internet

Concept # 2

Machine Learning (ML) algorithms are trained by using **lots of voice examples** and providing **accurate “text”** for each such sample. Very similar to how we teach kids by showing pictures..



This is
a “dog”

over 100,000 hours of audio is used for training



“my name”



“what”



“how long is”



“Columbus”

Humans do voice labeling,
and then it is provided to
algorithms as ground truth

**...equal to a employee
working over 50 years**

Apple has been in the news for privacy concerns..

Apple contractors 'regularly hear confidential details' on Siri recordings

Apple contractors 'regularly hear confidential details' on Siri recordings ... understand you the data "is used to help Siri and dictation ... understand you premise what you say".

Google workers were found to be 'paying human oversight of its automatic voice as revealed to employ staff to listen to some month, Google workers were found to be'

This is very important..

Training of an ML algorithm requires a lot of data that has been labeled by humans.

I highly recommend you read last three pages if this is not clear

Are you saying that during training, the algorithm predicts the text from voice samples and these predictions are compared against actual labels provided by human beings?



A cartoon illustration of a brown dog standing on its hind legs. The dog has a blue collar with a silver tag that reads "NEURON LEARNING". It has a worried expression with wide eyes and a slightly open mouth. A speech bubble originates from the top right, containing the text:

**Yes, you got it girl.
I am sure you will
win the competition**

A cartoon illustration of a man with brown hair and a blue cap being attacked by a large, tan-colored dog. The man is wearing a red shirt and a striped jacket. A speech bubble above him contains the text "How does the algorithm learn to make good predictions?"

**How does the algorithm learn
to make good predictions?**



That is similar to how I learn. When I do a good job you give me feedback and then I do more of that..

Character level predictions..

The Latest algorithms predict **character** in the audio file. This prediction is similar to how a non-English speaking person would spell the word “Wife” if it was said very slowly “WWIIIFFEE”. After that rules such as removing duplicate characters are applied to get the **most probable word**



WW_III_FF_EE



WIFE

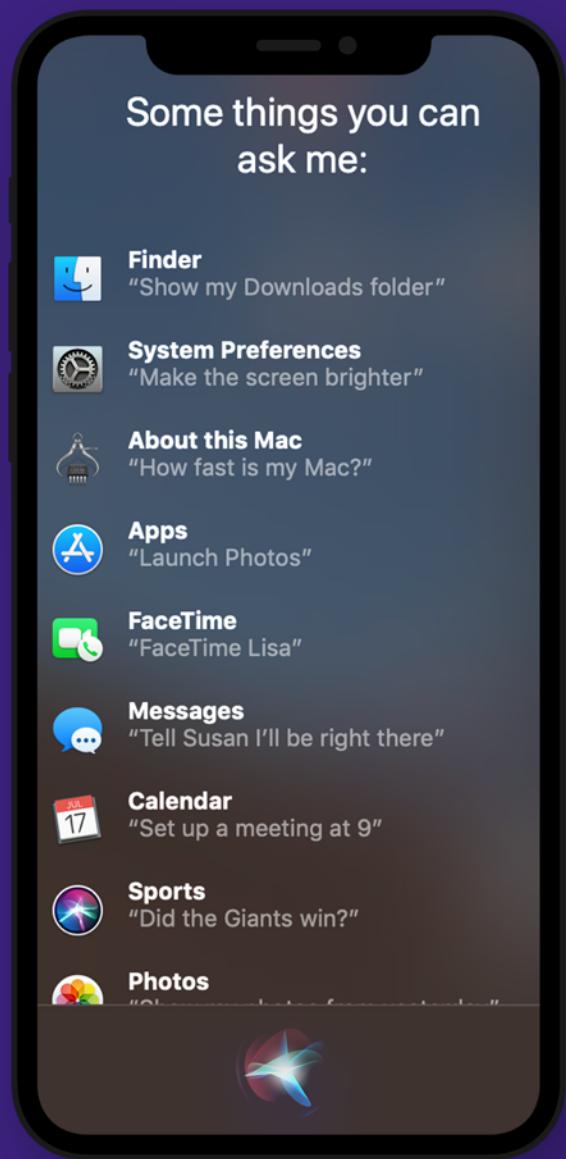
Intent behind the text

Let's assume a user says to Siri, "Call Wife". Based on our insights so far, we know Siri will convert it to text "Call Wife." How does Siri know **what do these word mean?**

That is covered in our last concept. You are almost there. Sit tight!

Concept #3

Siri doesn't try to figure out exactly what you said, but rather connect you to most relevant services/apps based on "trigger words"

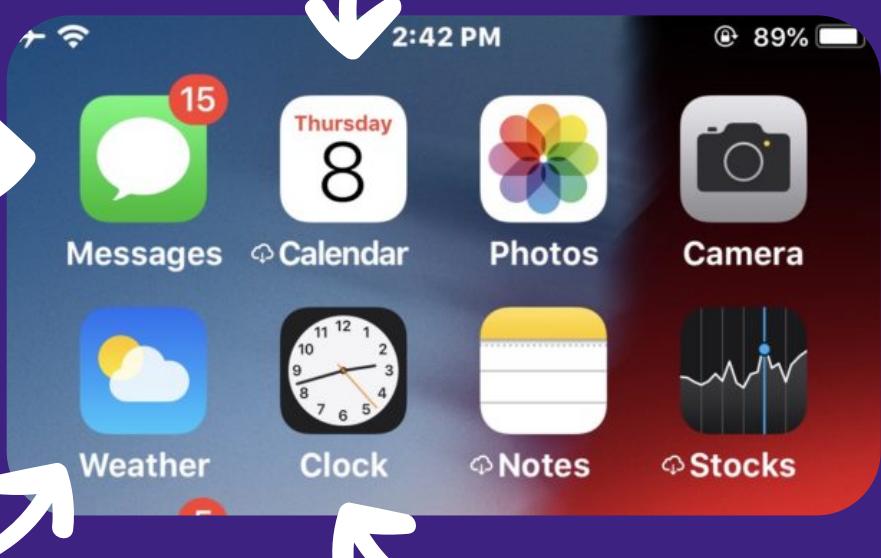


“Call Wife”

Based on the speech to text conversion, Siri will identify “Call” as a trigger word in “Call Wife”. After that it will invoke call placing application

Some trigger words..

Tell
Message
Text



Temperature
Umbrella
Weather

Set Appointment
Book Meeting

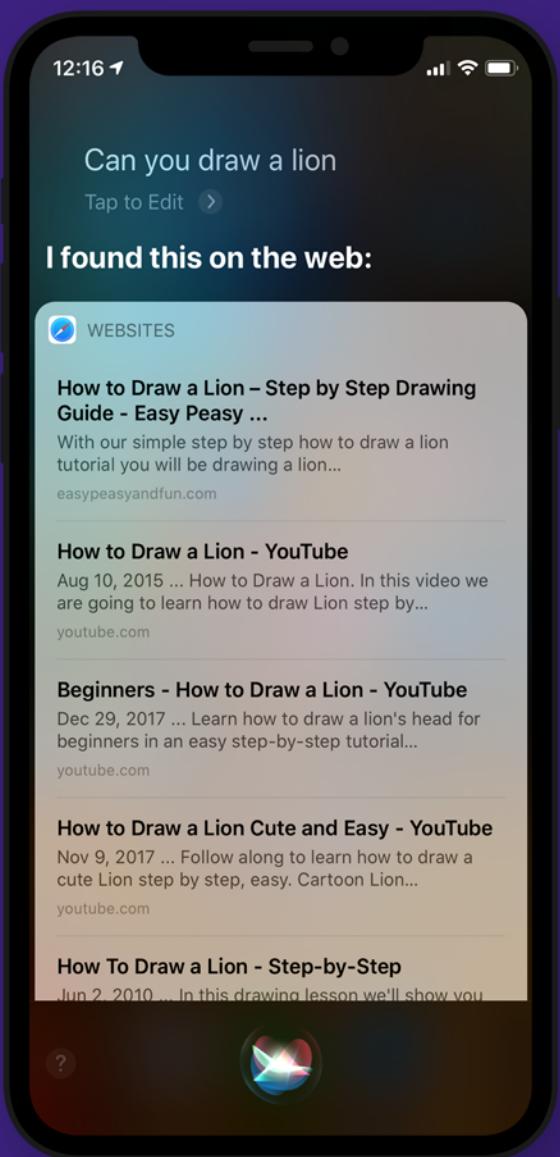
Alarm
Wake me

Given large volume
of voice samples, Siri has
a good sense for a different
number of ways people
ask for say “Weather”

Back to our “Call Wife” example

Once Siri realizes it has to place a “call”, it will get the name of a person to be called from text. In our case, if a contact “Wife” is saved then Siri will place the call and will complete the request

You have seen this default answer...



When Siri stumbles upon a request beyond its capability, it mostly invokes a web search



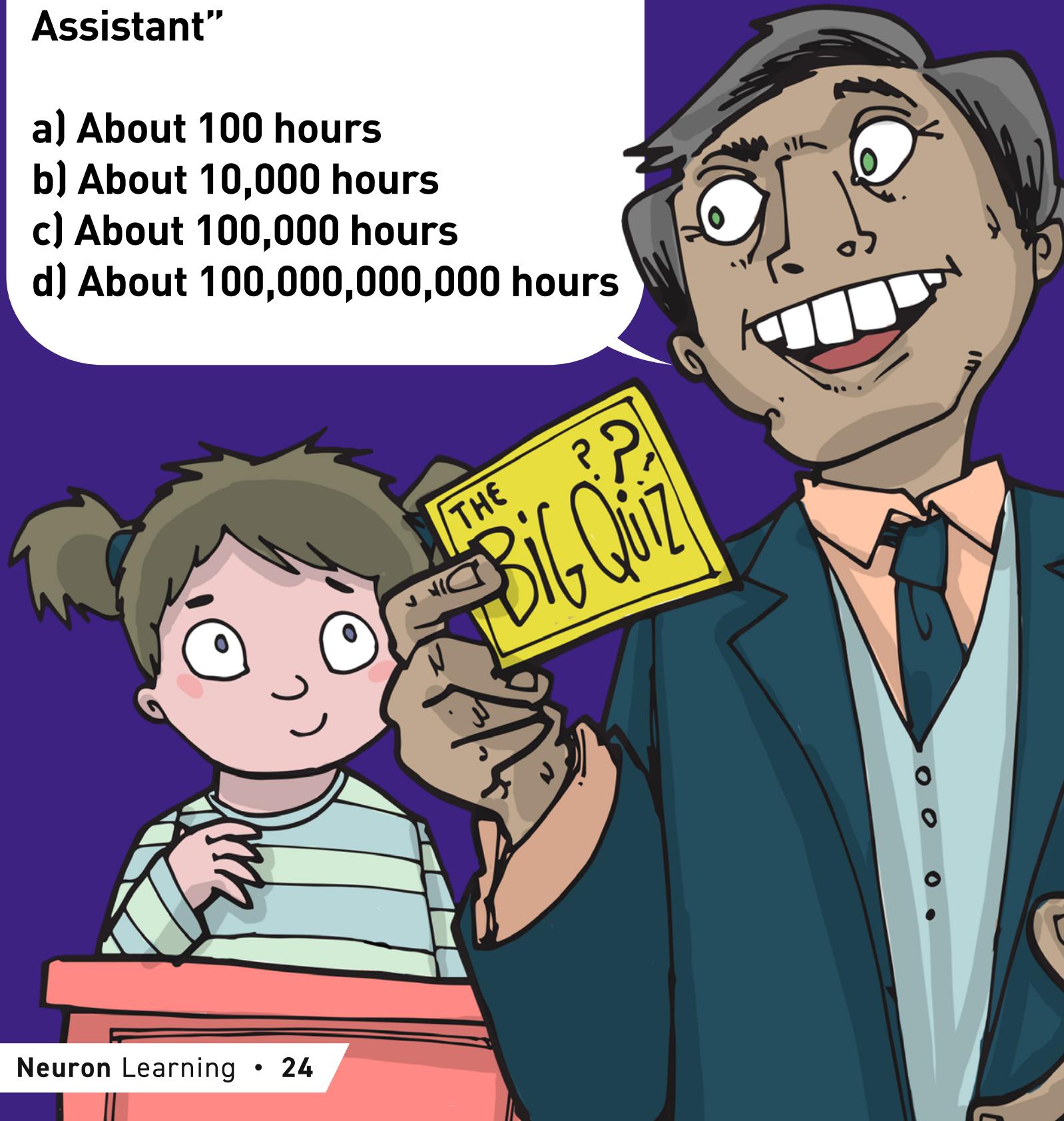
**OMG. You are the world's
best talking dog.I finally
understand how Siri works.**

**See you after the
competition**

**This is question for \$10,000 in
Level 5.**

**How many hours of audio are
used by Siri, Alexa and Google
Assistant”**

- a) About 100 hours**
- b) About 10,000 hours**
- c) About 100,000 hours**
- d) About 100,000,000,000 hours**



Congratulations. You are right Sophia





Follow Amol Palekar on LinkedIn to get next edition of **Neuron Learning**

For email notification,
subscribe:
NeuronTimes.com

Resources

Natural Language Processing by
National Research University Higher School
of Economics on Coursera

Sequence Models by deeplearning.ai
on Coursera

Adam Coates. - Deep Speech at BAMMF
on Youtube <https://www.youtube.com/watch?v=hyZCH3xU42E&list=PLes5o2b5ie2XM4SLb7drQjQEPnalgJLl>

Adam Coates – Deep Learning
for Speech Recognition on YouTube
<https://www.youtube.com/watch?v=g-sndkf7mCs&t=1980s>

TheGurdian.com

These are my personal opinions and for educational purposes only.
Content has been simplified to convey core ideas to non data scientists