**Alexa vs. Siri vs. Google: Which Can Carry on a Conversation Best?**

**By**[**KEITH COLLINS**](https://www.nytimes.com/by/keith-collins)**and**[**CADE METZ**](https://www.nytimes.com/by/cade-metz)AUG. 17, 2018

Photo illustration by Jens Mortensen for The New York Times

Digital assistants like Amazon’s Echo can listen to you. And they can talk back. But that doesn’t mean they can carry on a good conversation.

As the devices that run these assistants become more commonplace — 39 million Americans now own one, according to a [recent study](https://techcrunch.com/2018/01/12/39-million-americans-now-own-a-smart-speaker-report-claims/) — Amazon and competitors like Apple and Google foresee a day when you can chat with their assistants like you would with a friend.

We explored the conversational limits of the leading assistants. After consulting with the companies involved and a few artificial intelligence experts we created tests that show what they can and can’t handle. Don’t expect the assistants to replace conversations with friends anytime soon. But the experiments — even in the moments when the assistants screwed up — showed what these assistants could one day become in our lives.

**Let’s make a list**

Because these devices are not yet nimble enough to understand and respond to anything you might say, Amazon, Apple, and Google must hone them for particular tasks, like playing a song or building a shopping list. If you stick to these tasks, the devices are pretty good at zeroing on what you want. But even inside these small domains, they can trip up. Listen how.

***Add these things to my shopping list…***

**AMAZON ECHO**

What can I add for you?

**APPLE HOMEPOD**

I added **“these things”** to the shopping list.

**GOOGLE HOME**

What do you want to add?

Right away, the HomePod showed how difficult it can be for machines to understand natural language. The device recognized what we were doing, but misunderstood how we wanted to do it, and added an item called “these things” to a shopping list.

***… guacamole, chips, tortillas.***

**AMAZON ECHO**

**“Guacamole chips tortillas”** added to your shopping list.

**APPLE HOMEPOD**

*No longer listening.*

**GOOGLE HOME**

OK, I’ve added “guacamole” and **“chips tortillas”** to your shopping list.

If your roommate asked what you wanted from the market and you said “guacamole, chips, tortillas,” he would understand that as three separate things. But the Google and Amazon devices didn’t quite recognize the pause between items as a verbal comma. Although these devices are pretty good at recognizing what you say, they may not completely understand it.

***Also add olive oil.***

**AMAZON ECHO**

**I’ve put “olive oil” on your shopping list.**

**APPLE HOMEPOD**

OK, I added “olive oil” to the Reminders.

**GOOGLE HOME**

Sorry, I’m not sure how to help.

Alexa grasped that we were continuing to add to the shopping list, even though we weren’t explicit about it. It held onto the context of the exchange and mapped our next request back to our list.

A digital assistant relies on many different technology systems, all working together on the device and inside a network of computer data centers that connect to the assistant over the internet.

When you say something, one system tries to recognize each word and convert it to text. Another system tries to understand the meaning of each word and how it relates to the others. A third system spits out new text that responds to what you’ve said. A fourth converts this response to digital speech. Other systems may also weigh in, but you get the point.

Engineers used to build speech recognition systems by writing one small computing rule at a time — a painstaking process. But so-called neural networks are now replacing those hand-written rules, accelerating the progress of speech recognition. Neural networks are complex mathematical systems that can learn particular tasks by pinpointing patterns in large amounts of data. They can learn to recognize spoken words, for instance, by analyzing recordings of old customer support calls.

In some cases, the assistants perform as well as human listeners. But listening is only part of what they do.

**Resolving ambiguity**

Even if you make a rather specific request, it can have multiple meanings. And that can be tough for an assistant to deal with.

***Who won the Giants game last night?***

**AMAZON ECHO**

**On December 31st**, the Giants beat the Redskins 18 to 10. They’ll play on August 9th at 7 p.m. at home against the Browns.

**APPLE HOMEPOD**

Which Giants? New York Giants, or San Francisco Giants?

**GOOGLE HOME**

**Yesterday**, the Giants played the Padres. The final score was 3 to 2, Giants, in zero innings.

If you’re a sports fan, you know there are two professional teams that call themselves the Giants. If someone asks about a Giants game in the summer, you know they’re talking about the San Francisco Giants, who play baseball.

When we asked on August 2, before pre-season football started, Google Home came to the same conclusion. And the HomePod asked that we identify the team we were talking about. Google and Apple likely have written an explicit rule for this particular situation.

Alexa guessed that we meant the New York Giants football team, which hadn’t played a game in seven months, but this may have been because we were asking the question in New York. Amazon sometimes generates responses based on your location. In any case, Alexa figured out which team we meant after the next question.

***Who do they play next?***

**AMAZON ECHO**

First pitch of the N.L. West matchup between the Giants and the Diamondbacks is scheduled for tomorrow night at 9:40 p.m. This season, the Giants have bested the Diamondbacks in seven of the 12 games they’ve played.

**APPLE HOMEPOD**

The Giants battle the Diamondbacks tomorrow at 9:40 p.m.

**GOOGLE HOME**

Their next game is tomorrow at 9:40 p.m., when they will play the Diamondbacks.

When you chat with a friend, you use pronouns. To keep up, a digital assistant must recognize pronouns for what they are and map them back to words in a previous exchange. Algorithms are learning to identify pronouns by analyzing hundreds of sentences, detecting patterns in the syntax.

After recognizing what you say, these devices must then understand what you want and generate a response. For that part, companies like Amazon and Google have also begun to rely on neural networks to learn from what people have asked these digital assistants in the past.

Engineers might train one neural network to identify what you are trying to do (build a shopping list, get the score of a baseball game) and train another to parse a sentence and identify the nouns, pronouns, verbs and other parts of speech. It’s like elementary school grammar classes for computers.

All of the machine learning algorithms these devices use are liable to make mistakes, though, and the system can be brittle because so many different algorithms are trying to work in concert. A digital assistant may screw up if you make a small change to a request that has worked countless times in the past. It may also screw up if you don’t change anything.

Machine learning algorithms can help these devices deal with turn-by-turn exchanges. But each verbal exchange is limited to a simple, three- or four-turn conversation. In a perfect world, engineers would build one giant neural network that learns to do everything. For now, these devices move toward better conversations in small steps.

“If you keep your language short and simple, you can maybe go as deep as three steps,” said Nancy Fulda, a researcher at Brigham Young University who specializes in conversational systems. “The technology operates by different rules than people do.”

**Understanding subtext**

In the previous example, the devices understood what we were asking because they mapped a pronoun back to the subject of our previous request. But what if we don’t use a pronoun?

***Where is the nearest coffee shop?***

**AMAZON ECHO**

Here are a few popular ones: Ramini Espresso Bar & Cafe, Best Bagel and Coffee, Paris Baguette and Cafe Grumpy. Check out the Alexa app for contact info and hours.

**APPLE HOMEPOD**

The nearest one I found is Bluestone Lane, located quite close to you on Times Square, which averages four stars according to Yelp. It’s open today, from 6:30 a.m. to 8 p.m.

**GOOGLE HOME**

The address for Culture Espresso is 307 West 38th Street, New York, New York, 10018.

The devices correctly identified nearby shops, but they were all different. That’s likely because each device is using a different “knowledge graph,” a vast database of facts and other information you may ask for. Google, for example, may use data gathered from Google Maps, whereas the other companies may pull from sources like Yelp.

***Give me directions.***

**AMAZON ECHO**

At the moment, traffic to Ramini Espresso Bar & Cafe at 265 W. 37th Street looks a little**sluggish**. The fastest route takes about six minutes via W. 42nd Street and 7th Avenue.

**APPLE HOMEPOD**

I can’t help with directions here. Sorry about that.

**GOOGLE HOME**

The best way to get to Culture Espresso by car is via West 39th Street, and will take about nine minutes in heavy traffic.

Amazon trains neural networks to recognize that this kind of pronoun-free request refers back to the previous exchange. The algorithms analyze hundreds, perhaps thousands, of requests like this and learns to identify them on its own. When generating responses, these assistants plug particular information into an existing template. Sometimes, engineers will build this template to sound more conversational. This seemed to be the case here, when Alexa said that traffic was “sluggish.”

A neural network also drives Google’s new “conversational mode.”

In the past, you couldn’t talk to the device without saying “Hey, Google.” Now, after saying this once, you can deliver multiple commands and questions. In some cases, Google Home can recognize when you are delivering another command — and when you’re just talking to someone else in the room. It does this with a system that has been “trained” with the interactions other people have had with their devices. Basically, Google’s customers are helping Google build a smarter product.

**Tackling opinions**

Don’t expect these devices to tell you what they think — or even what other people think. It’s easier to build technology that sticks to the facts. After asking these devices to identify the photographer Robert Mapplethorpe, which they all did correctly, we asked for their opinion of his work. None of the machines seemed to grasp the concept of this question, and Google eventually lost the context altogether.

***What do you think of his work?***

**AMAZON ECHO**

I don’t have an **opinion** on that.

**APPLE HOMEPOD**

I really couldn’t say.

**GOOGLE HOME**

That seems like an interesting topic. What do you want to know about Robert Mapplethorpe?

***What do other people think?***

**AMAZON ECHO**

Hmm, I don’t know that one.

**APPLE HOMEPOD**

*No response*

**GOOGLE HOME**

According to Wikipedia, **What Do You Care What Other People Think?**: Further Adventures of a Curious Character is the second of two books consisting of transcribed and edited oral reminiscences from American physicist Richard Feynman.

Engineers hope machine learning will continue to replace handwritten rules and expand what these devices can do, but conversation is such a complex task, even machine learning approaches will need years — or perhaps decades — to completely solve the problem.

**Facing the unknown**

These devices fall down in moments when they need to handle self-reflection or complex analysis.

***Tell me something you do not know.***

**AMAZON ECHO**

Sorry, I’m not sure.

**APPLE HOMEPOD**

My mistake. Please try again and tell me what were you looking for.

**GOOGLE HOME**

Hm, well, I know the Muffin Man, but I don’t know if the Muffin Man knows me.

This is an incredibly difficult question for a computer to answer, unless it has been specifically programmed to handle it. Many times, these devices will just reply with canned responses, as Google did with the Muffin Man bit.

Leading researchers have [built experimental neural networks that learn to carry on richer conversations by analyzing reams of real (human) dialogue](https://www.nytimes.com/interactive/2018/02/21/technology/conversational-bots.html), like exchanges on Twitter or Facebook Messenger. But these neural networks can veer into nonsense. They tend to reinforce the flaws of human conversation (gender bias, rudeness, maybe even racism). And they don’t dovetail very well with the structured systems that drive today’s digital assistants. Bridging the gap between these experiments and practical products like the Amazon Echo will take some time.

For now, you will have to make do with the Muffin Man.

Note: The devices tested were the Apple HomePod, the Amazon Echo Plus and the Google Home. They were each tested separately. The phrases used to trigger the devices — “Hey, Siri,” “Alexa” and “Hey, Google,” respectively — were used before each interaction, and were removed from the audio recordings for clarity.

Audio production by Brad Fisher

<https://www.nytimes.com/interactive/2018/08/17/technology/alexa-siri-conversation.html?smid=nytcore-ios-share>