

For a better experience on Facebook, [switch to our basic site](#) or [update your browser](#).

Sign Up

Join or Log Into Facebook

Email or Phone

haijunz1@gmail.com

Password

.....


[Forgot account?](#)

Log In

Do you want to join Facebook?

Sign Up

Building Jarvis

 MARK ZUCKERBERG · MONDAY, DECEMBER 19, 2016

My personal challenge for 2016 was to build a simple AI to run my home -- like Jarvis in Iron Man.

My goal was to learn about the state of artificial intelligence -- where we're further along than people realize and where we're still a long ways off. These challenges always lead me to learn more than I expected, and this one also gave me a better sense of all the internal technology Facebook engineers get to use, as well as a thorough overview of home automation.

So far this year, I've built a simple AI that I can talk to on my phone and computer, that can control my home, including lights, temperature, appliances, music and security, that learns my tastes and patterns, that can learn new words and concepts, and that can even entertain Max. It uses several artificial intelligence techniques, including natural language processing, speech recognition, face recognition, and reinforcement learning, written in Python, PHP and Objective C. In this note, I'll explain what I built and what I learned along the way.

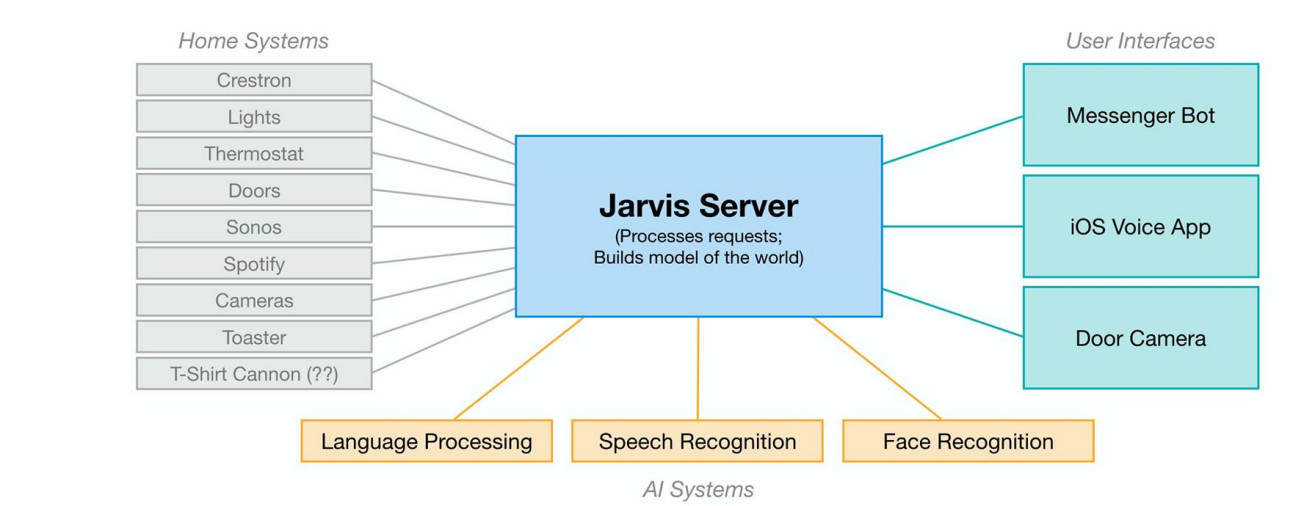


Diagram of the systems connected to build Jarvis.

Getting Started: Connecting the Home

In some ways, this challenge was easier than I expected. In fact, my running challenge (I also set out to run 365 miles in 2016) took more total time. But one aspect that was much more complicated than I expected was simply connecting and communicating with all of the different systems in my home.

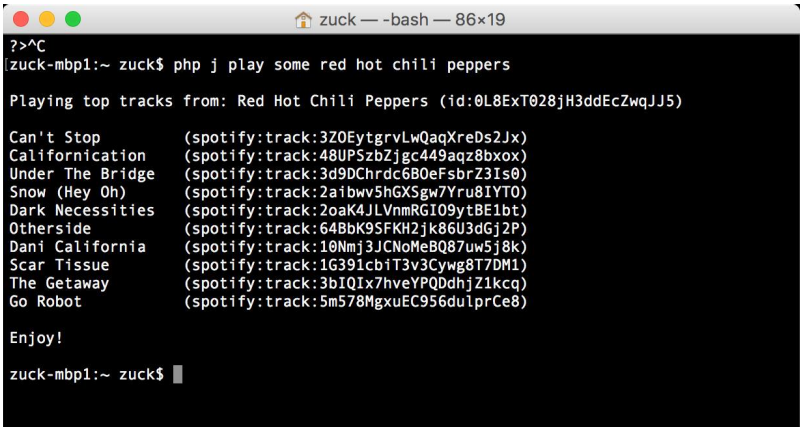
Before I could build any AI, I first needed to write code to connect these systems, which all speak different languages and protocols. We use a Crestron system with our lights, thermostat and doors, a Sonos system with Spotify for music, a Samsung TV, a Nest cam for Max, and of course my work is connected to Facebook's systems. I had to reverse engineer APIs for some of these to even get to the point where I could issue a command from my computer to turn the lights on or get a song to play.

Further, most appliances aren't even connected to the internet yet. It's possible to control some of these using internet-connected power switches that let you turn the power on and off remotely. But often that isn't enough. For example, one thing I learned is it's hard to find a toaster that will let you push the bread down while it's powered off so you can automatically start toasting when the power goes on. I ended up finding an old toaster from the 1950s and rigging it up with a connected switch. Similarly, I found that connecting a food dispenser for Beast or a grey t-shirt cannon would require hardware modifications to work.

For assistants like Jarvis to be able to control everything in homes for more people, we need more devices to be connected and the industry needs to develop common APIs and standards for the devices to talk to each other.

Natural Language

Once I wrote the code so my computer could control my home, the next step was making it so I could talk to my computer and home the way I'd talk to anyone else. This was a two step process: first I made it so I could communicate using text messages, and later I added the ability to speak and have it translate my speech into text for it to read.



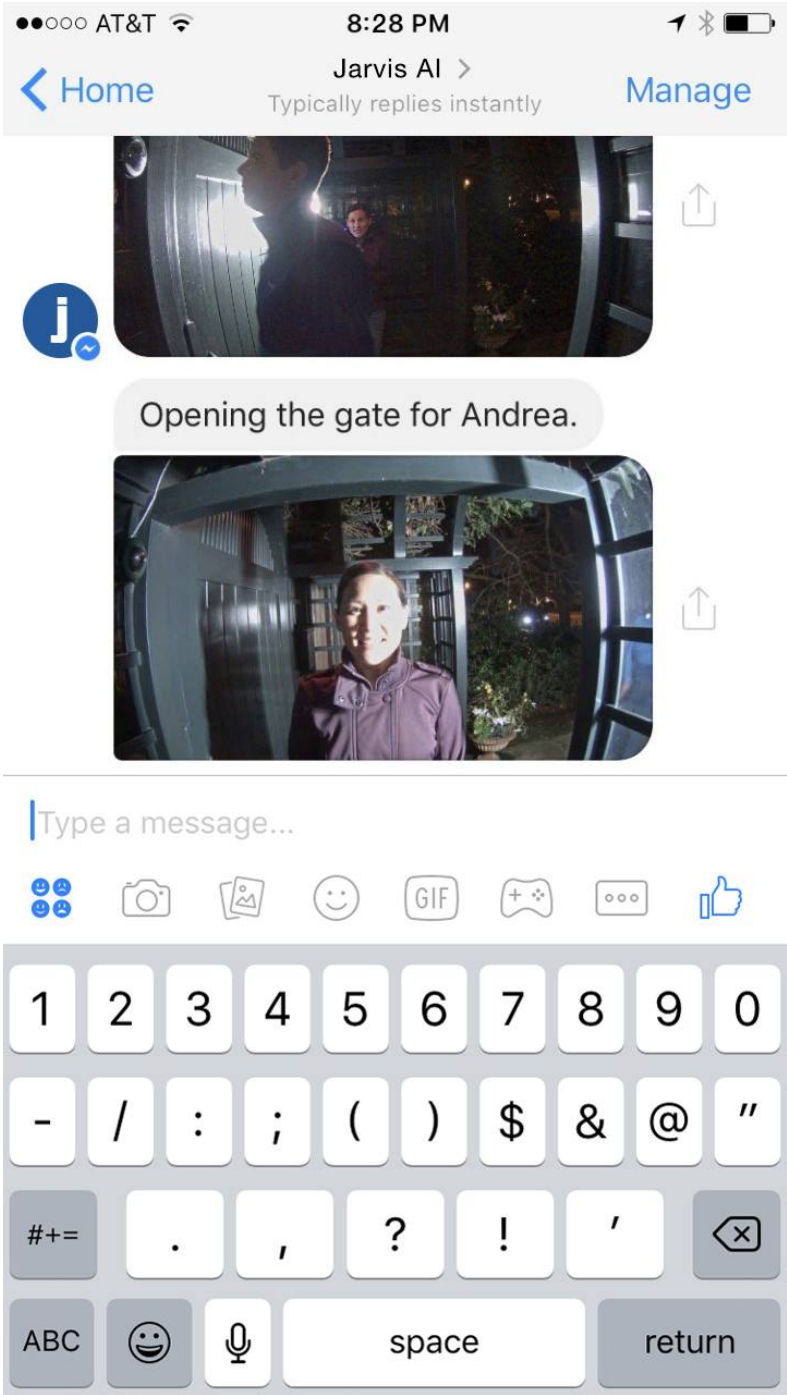
An example natural language request from command line.

It started simple by looking for keywords, like "bedroom", "lights", and "on" to determine I was telling it to turn the lights on in the bedroom. It quickly became clear that it needed to learn synonyms, like that "family room" and "living room" mean the same thing in our home. This meant building a way to teach it new words and concepts.

Understanding context is important for any AI. For example, when I tell it to turn the AC up in "my office", that means something completely different from when Priscilla tells it the exact same thing. That one caused some issues! Or, for example, when you ask it to make the lights dimmer or to play a song without specifying a room, it needs to know where you are or it might end up blasting music in Max's room when we really need her to take a nap. Whoops.

Music is a more interesting and complex domain for natural language because there are too many artists, songs and albums for a keyword system to handle. The range of things you can ask it is also much greater. Lights can only be turned up or down, but when you say "play X", even subtle variations can mean many different things. Consider these requests related to Adele: "play someone like you", "play someone like adele", and "play some adele". Those sound similar, but each is a completely different category of request. The first plays a specific song, the second recommends an artist, and the third creates a playlist of Adele's best songs. Through a system of positive and negative feedback, an AI can learn these differences.

The more context an AI has, the better it can handle open-ended requests. At this point, I mostly just ask Jarvis to "play me some music" and by looking at my past listening patterns, it mostly nails something I'd want to hear. If it gets the mood wrong, I can just tell it, for example, "that's not light, play something light", and it can both learn the classification for that song and adjust immediately. It also knows whether I'm talking to it or Priscilla is, so it can make recommendations based on what we each listen to. In general, I've found we use these more open-ended requests more frequently than more specific asks. No commercial products I know of do this today, and this seems like a big opportunity.



Jarvis uses face recognition to let my friends in automatically and let me know.

Vision and Face Recognition

About one-third of the human brain is dedicated to vision, and there are many important AI problems related to understanding what is happening in images and videos. These problems include tracking (eg is Max awake and moving around in her crib?), object recognition (eg is that Beast or a rug in that room?), and face recognition (eg who is at the door?).

Face recognition is a particularly difficult version of object recognition because most people look relatively similar compared to telling apart two random objects -- for example, a sandwich and a house. But Facebook has gotten very good at face recognition for identifying when your friends are in your photos. That expertise is also useful when your friends are at your door and your AI needs to determine whether to let them in.

To do this, I installed a few cameras at my door that can capture images from all angles. AI systems today cannot identify people from the back of their heads, so having a few angles ensures we see the person's face. I built a simple server that continuously watches the cameras and runs a two step process: first, it runs face detection to see if any person has come into view, and second, if it finds a face, then it runs face recognition to identify who the person is. Once it identifies the person, it checks a list to confirm I'm expecting that person, and if I am then it

will let them in and tell me they're here.

This type of visual AI system is useful for a number of things, including knowing when Max is awake so it can start playing music or a Mandarin lesson, or solving the context problem of knowing which room in the house we're in so the AI can correctly respond to context-free requests like "turn the lights on" without providing a location. Like most aspects of this AI, vision is most useful when it informs a broader model of the world, connected with other abilities like knowing who your friends are and how to open the door when they're here. The more context the system has, the smarter it gets overall.

Messenger Bot

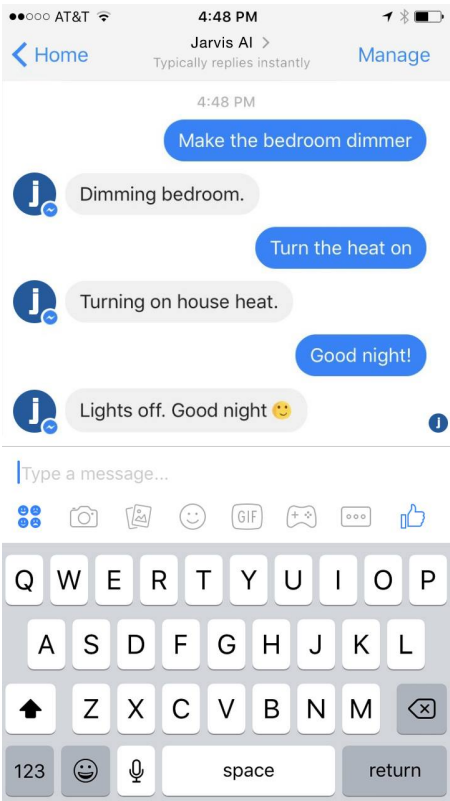
I programmed Jarvis on my computer, but in order to be useful I wanted to be able to communicate with it from anywhere I happened to be. That meant the communication had to happen through my phone, not a device placed in my home.

I started off building a Messenger bot to communicate with Jarvis because it was so much easier than building a separate app. Messenger has a simple framework for building bots, and it automatically handles many things for you -- working across both iOS and Android, supporting text, image and audio content, reliably delivering push notifications, managing identity and permissions for different people, and more. You can learn about the bot framework at messenger.com/platform.

I can text anything to my Jarvis bot, and it will instantly be relayed to my Jarvis server and processed. I can also send audio clips and the server can translate them into text and then execute those commands. In the middle of the day, if someone arrives at my home, Jarvis can text me an image and tell me who's there, or it can text me when I need to go do something.

One thing that surprised me about my communication with Jarvis is that when I have the choice of either speaking or texting, I text much more than I would have expected. This is for a number of reasons, but mostly it feels less disturbing to people around me. If I'm doing something that relates to them, like playing music for all of us, then speaking feels fine, but most of the time text feels more appropriate. Similarly, when Jarvis communicates with me, I'd much rather receive that over text message than voice. That's because voice can be disruptive and text gives you more control of when you want to look at it. Even when I speak to Jarvis, if I'm using my phone, I often prefer it to text or display its response.

This preference for text communication over voice communication fits a pattern we're seeing with Messenger and WhatsApp overall, where the volume of text messaging around the world is growing much faster than the volume of voice communication. This suggests that future AI products cannot be solely focused on voice and will need a private messaging interface as well. Once you're enabling private messaging, it's much better



I can text Jarvis from anywhere using a Messenger bot.

to use a platform like Messenger than to build a new app from scratch. I have always been optimistic about AI bots, but my experience with Jarvis has made me even more optimistic that we'll all communicate with bots like Jarvis in the future.

Voice and Speech Recognition

Even though I think text will be more important for communicating with AIs than people realize, I still think voice will play a very important role too. The most useful aspect of voice is that it's very fast. You don't need to take out your phone, open an app, and start typing -- you just speak.

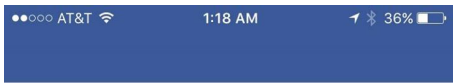
To enable voice for Jarvis, I needed to build a dedicated Jarvis app that could listen continuously to what I say. The Messenger bot is great for many things, but the friction for using speech is way too much. My dedicated Jarvis app lets me put my phone on a desk and just have it listen. I could also put a number of phones with the Jarvis app around my home so I could talk to Jarvis in any room. That seems similar to Amazon's vision with Echo, but in my experience, it's surprising how frequently I want to communicate with Jarvis when I'm not home, so having the phone be the primary interface rather than a home device seems critical.

I built the first version of the Jarvis app for iOS and I plan to build an Android version soon too. I hadn't built an iOS app since 2012 and one of my main observations is that the toolchain we've built at Facebook since then for developing these apps and for doing speech recognition is very impressive.

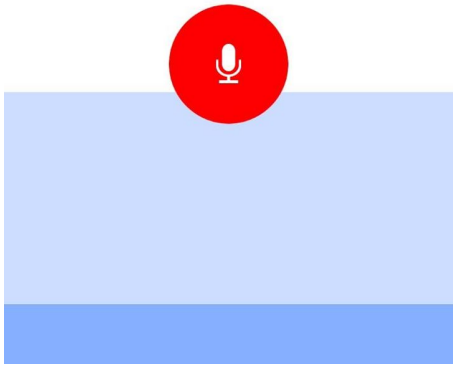
Speech recognition systems have improved recently, but no AI system is good enough to understand conversational speech just yet. Speech recognition relies on both listening to what you say and predicting what you will say next, so structured speech is still much easier to understand than unstructured conversation.

Another interesting limitation of speech recognition systems -- and machine learning systems more generally -- is that they are more optimized for specific problems than most people realize. For example, understanding a person talking to a computer is subtly different problem from understanding a person talking to another person. If you train a machine learning system on data from Google of people speaking to a search engine, it will perform relatively worse on Facebook at understanding people talking to real people. In the case of Jarvis, training an AI that you'll talk to at close range is also different from training a system you'll talk to from all the way across the room, like Echo. These systems are more specialized than it appears, and that implies we are further off from having general systems than it might seem.

On a psychologic level, once you can speak to a system, you attribute more emotional depth to it than a computer you might interact with using text or a graphic interface. One interesting observation is that ever since I built voice into Jarvis, I've also wanted to build in more humor. Part of this is that now it can interact with Max and I



shoot me a new gray t-shirt



Jarvis uses speech recognition in my iOS app to listen to my request for a fresh t-shirt.

want those interactions to be entertaining for her, but part of it is that it now feels like it's present with us. I've taught it fun little games like Priscilla or I can ask it who we should tickle and it will randomly tell our family to all go tickle one of us, Max or Beast. I've also had fun adding classic lines like "I'm sorry, Priscilla. I'm afraid I can't do that."

There's a lot more to explore with voice. The AI technology is just getting good enough for this to be the basis of a great product, and it will get much better in the next few years. At the same time, I think the best products like this will be ones you can bring with you anywhere and communicate with privately as well.

Facebook Engineering Environment

As the CEO of Facebook, I don't get much time to write code in our internal environment. I've never stopped coding, but these days I mostly build personal projects like Jarvis. I expected I'd learn a lot about the state of AI this year, but I didn't realize I would also learn so much about what it's like to be an engineer at Facebook. And it's impressive.

My experience of ramping up in the Facebook codebase is probably pretty similar to what most new engineers here go through. I was consistently impressed by how well organized our code is, and how easy it was to find what you're looking for -- whether it's related to face recognition, speech recognition, the [Messenger Bot Framework](#) [[messenger.com/platform](#)] or iOS development. The open source [Nuclide](#) [[github.com/facebook/nuclide](#)] packages we've built to work with GitHub's Atom make development much easier. The [Buck](#) [[buckbuild.com](#)] build system we've developed to build large projects quickly also saved me a lot of time. Our open source [FastText](#) [[github.com/facebookresearch/fastTex...](#)] AI text classification tool is also a good one to check out, and if you're interested in AI development, the whole [Facebook Research](#) [[github.com/facebookresearch](#)] GitHub repo is worth taking a look at.

One of our values is "move fast". That means you should be able to come here and build an app faster than you can anywhere else, including on your own. You should be able to come here and use our infra and AI tools to build things it would take you a long time to build on your own. Building internal tools that make engineering more efficient is important to any technology company, but this is something we take especially seriously. So I want to give a shout out to everyone on our infra and tools teams that make this so good.

Next Steps

Although this challenge is ending, I'm sure I'll continue improving Jarvis since I use it every day and I'm always finding new things I want to add.

In the near term, the clearest next steps are building an Android app, setting up Jarvis voice terminals in more rooms around my home, and connecting more appliances. I'd love to have Jarvis control my Big Green Egg and help me cook, but that will take even more serious hacking than rigging up the t-shirt cannon.

In the longer term, I'd like to explore teaching Jarvis how to learn new skills itself rather than me having to teach it how to perform specific tasks. If I spent another

year on this challenge, I'd focus more on learning how learning works.

Finally, over time it would be interesting to find ways to make this available to the world. I considered open sourcing my code, but it's currently too tightly tied to my own home, appliances and network configuration. If I ever build a layer that abstracts more home automation functionality, I may release that. Or, of course, that could be a great foundation to build a new product.

Conclusions

Building Jarvis was an interesting intellectual challenge, and it gave me direct experience building AI tools in areas that are important for our future.

I've previously predicted that within 5-10 years we'll have AI systems that are more accurate than people for each of our senses -- vision, hearing, touch, etc, as well as things like language. It's impressive how powerful the state of the art for these tools is becoming, and this year makes me more confident in my prediction.

At the same time, we are still far off from understanding how learning works. Everything I did this year -- natural language, face recognition, speech recognition and so on -- are all variants of the same fundamental pattern recognition techniques. We know how to show a computer many examples of something so it can recognize it accurately, but we still do not know how to take an idea from one domain and apply it to something completely different.

To put that in perspective, I spent about 100 hours building Jarvis this year, and now I have a pretty good system that understands me and can do lots of things. But even if I spent 1,000 more hours, I probably wouldn't be able to build a system that could learn completely new skills on its own -- unless I made some fundamental breakthrough in the state of AI along the way.


In a way, AI is both closer and farther off than we imagine. AI is closer to being able to do more powerful things than most people expect -- driving cars, curing diseases, discovering planets, understanding media. Those will each have a great impact on the world, but we're still figuring out what real intelligence is.

Overall, this was a great challenge. These challenges have a way of teaching me more than I expected at the beginning. This year I thought I'd learn about AI, and I also learned about home automation and Facebook's internal technology too. That's what's so interesting about these challenges. Thanks for following along with this challenge and I'm looking forward to sharing next year's challenge in a few weeks.

Priscilla Chan, Jessica E. Lessin, Chris Hughes and 229,604 others like this.

28,933 shares


17K Comments



Janos Keresztes

Great job ! Could you post a short video demo? As you perform many tasks such as voice and face recognition, did you encounter any lag?

670 · December 19, 2016 at 9:19am



Mark Zuckerberg

Yup, coming later today or tomorrow.

2,166 · December 19, 2016 at 9:32am

View more replies

8 of 10

2017年04月27日 18:47



Sajan Sj Wish jarvis could read me a summary of this long note !
459 · December 19, 2016 at 9:11am



Mark Zuckerberg I'll post a fun video summary later today or tomorrow.
1,047 · December 19, 2016 at 9:31am
[View more replies](#)



Pedram Keyani Well done! Are you going to make the iron man suit in 2017?
377 · December 19, 2016 at 9:32am
[View previous replies](#)



Mark Zuckerberg No, different kind of challenge each year.
497 · December 19, 2016 at 2:11pm
[View more replies](#)



Dan Gerson We already know Ironman and Iron Patriot, what will you call your suit?
96 · December 19, 2016 at 10:49am



Mark Zuckerberg The Iron Hoodie 😊
1,640 · December 19, 2016 at 10:29pm
[View more replies](#)



Udbhav Bhatnagar Android app FTW! You could rig up many cheap android tablets to act as terminals in your home much like the way finders and meeting rooms at mpk
197 · December 19, 2016 at 9:13am · Edited



Mark Zuckerberg That's the plan!
373 · December 19, 2016 at 9:20am
[View more replies](#)



Zizhuang Yang Impressive. Have you built proactive features beyond the door into Jarvis yet (ie. automatically playing music, managing lights) or do you prefer keeping the trigger to be upon human-initiated commands?
157 · December 19, 2016 at 9:53am



Mark Zuckerberg The toaster!
502 · December 19, 2016 at 10:19am
[View more replies](#)



Tanzim Saqib We are close, yet far away. AI needs a quantum leap. May be there's a hardware gap to fill in. Cloud is not the answer here. Our fundamental approach towards instruction execution is just too old; perhaps a non-binary new kind of hardware executor could be a key.
186 · December 19, 2016 at 9:23am



Mark Zuckerberg That's the thing about AI. It's sort of like magic. We only call things AI that we don't understand yet. Once we understand something, it's just math.

But if you'd asked someone 30 years if a computer system that you could talk to, that could see your... [See More](#)
971 · December 19, 2016 at 10:30am
[View more replies](#)



Thabelo Morobe Interesting times indeed. I regret not doing programming in my youth.
472 · December 19, 2016 at 9:24am
[View previous replies](#)



Mark Zuckerberg Also, it's never too late to start yourself 😊
550 · December 19, 2016 at 9:28am
[View more replies](#)



Gabor Nagy Nice results! I'm build an "semi-intelligent" house with lights, heating, doors/gate control. I'm looking for such system that I can manage by my iPhone. I want to use cameras, motion sensors, wi-fi system to locate my position (and my family members) ... [See More](#)
84 · December 19, 2016 at 2:58pm



Mark Zuckerberg I don't think you need any special brand of lights, thermostat or doors for this to work. I recommend using a system like Crestron as a common interface to all those basic devices to simplify things though. The part you'll need to pay closer attention ... [See More](#)
429 · December 19, 2016 at 10:40pm
[View more replies](#)



Erica Cohen A little more complex than the days of Synapse AI? 😊
108 · December 19, 2016 at 9:48am · Edited
[View previous replies](#)



Mark Zuckerberg Just a little 😊
208 · December 19, 2016 at 10:17am
[View more replies](#)

[View more comments](#)

10 of 17,539

