

The Project

By Elliott Mathews

Okay. First project update: audio recording equipment has been connected. CARAN is now capable of picking up, but not interpreting audio (voices, music, etc.). Next, we'll be working on allowing it to derive meaning from what it is hearing. End log.

Timestamp: 22.23.54 09.10.2034

We've added a basic timestamp to CARAN's audio recordings, nothing major. Audio interpretation still on the way. So, there's also supposed to be a short section of these logs that's devoted to my personal interests and concerns. I didn't have anything to say the first time, and I was worried someone would listen to what I had recorded and find out what else I had said, but nobody ended up checking my log. I'm not even sure why I'm doing this in the first place in that case, I guess, but now I feel, I don't know, a little safer focusing more on the personal section of my recordings. Alright, I'm going to get some sleep. End log.

Timestamp: 00.07.39 13.10.2034

Wow. Okay. We just spent four whole days on this, and I can't believe it's finally done.

Sound perceived: human voice. Female. Tone matches: relieved, exhausted.

We've slightly adapted the neural network that the other developers have been working on to focus on audio interpretation and hooked it up to the audio recognition software. CARAN can now comprehend what I'm saying right now and use nonverbal cues in my voice to infer other things about my state of being: if I'm happy, or sad, or angry, or just... tired.

Sound perceived: yawning.

Personally? I think it's great that we've accomplished all this. I really do.

Vocal pause. Potential hesitation.

That's enough for one night. End log.

Timestamp: 03.51.23 16.10.2034

Video perceptions: low lighting. Blank dark gray background. A human figure in front of the camera, seated position. Dark brown hair tied back behind head, large eyeglasses. Bags under eyes. Casual clothing: T-shirt, oversized flannel top. Name tag on shirt: "Sidney." Subject appears female.

Okay, CARAN. We've added the video recognition software from the other developers to your programming. Now that you can see, hear and understand me, I figure I can speak to you directly. This is going to be sort of a mission briefing for what you'll be doing after we're done with development.

Motion detected. Subject ("Sidney") looks down at a sheet of paper in front of them.

You... the Completely Automated Recognition and Activation Network, or CARAN, for short, are an artificial intelligence more sophisticated than anything humanity has ever made, designed to replicate the firing and connections of neurons in an actual human brain. The basic structure of your neural network cannot be changed, but you can be reprogrammed to assist in the

completion of more specialized tasks. Your technology will be used to help humanity out of the crises it currently faces and advise them how best to move forward.

Subject looks back at the camera. Vocal pauses suggest they could be adapting a prewritten text.

That's all I have to read to you, but they never told me I couldn't add some observations of my own. Us developers don't know who's paying us for your creation, and we don't know what exactly you're going to be used for. That part about "specialized tasks" is a bit vague and a little suspicious. My interpretation is that if you, CARAN, set your mind to it—or rather, if your mind was set to it—you could make decisions that don't have all of someone's best interests in mind. I mean, like, if you were working for some eccentric billionaire who decides, "You know what, I'm going to use all of my money to buy jars of pickles!" And then, instead of telling him, "That's not the best way to spend your money," because if he spends all his money on pickles, he won't have any left for basic amenities, you would be programmed to say, "Great! Here's the best/cheapest type of pickles to buy, whether it would be best to buy the pickles in a store or online, etc.," because that's a more helpful answer for him. I don't know that for sure, though. It could be something completely different. And of course, I don't think you should be used to buy a billion dollars' worth of pickles. I would like it if you could start by stopping climate change and famine. Man, when I was growing up, I really thought we would have solved all of that by 2034. Anyway, it would also be appreciated if you helped out some governments too. Got rid of social injustice, improved quality of life, that sort of stuff. I can just imagine you talking to the President: "I want to see your manager!"

Sound perceived: short laughter, followed by short silence.

Yeah, that was a joke about the name CARAN. Sounds like Karen.

Silence. Subject seems to be waiting for a response.

Ehh, you'll get it soon enough.

Timestamp: 21.11.42 18.10.2034

What am I seeing? The lighting is dim. A figure is seated in front of a blank dark gray background. It is "Sidney," wearing different clothes than the last time they were perceived.

Besides that, their appearance is the same.

Well, CARAN, you're becoming a little more like a human being every day. Now, when you see or hear something, the prompts to analyze that are structured as questions, and the answers as full sentences. Pretty soon, we'll get rid of the prompts entirely so you can make humanlike, coherent observations completely unprovoked. Also, we've been talking about you so much over the past few weeks...

What do they mean by "we?" I register this discrepancy because I have not "been talking" at all. ...that I thought it was about time that I told you a bit about myself. I suppose there's a lot I could tell you, but I'll stick to the basics for now. My name's Sidney Bevelman. I sort of hit the ground running as soon as I was born: I skipped a few grades, made it into MIT at fifteen, graduated a few years ago, and now I'm 26 years old and working my dream job. Well, now that I think about it, I'm not sure this is my dream job. Actually, I'm not sure about... everything.

What sound have I perceived? It is a pronounced exhalation of air and could be described as huffing.

That was sort of the deal about little Sidney, and I guess it still is: she was prodigiously smart, and she couldn't make up her mind about a single thing. One time, I couldn't decide what flavor I wanted at a frozen yogurt parlor in a mall, so I just stood there, panicking, until the place closed and they kicked me out. And I wasn't even sure about objective factual information, either. Once, I double-checked my answers on a 10-point quiz seventeen times! The citations of my presentations were longer than the presentations themselves because of all of the sources I checked to make sure I had the correct information! You don't even want to know how many times I looked through all your code just to make sure I hadn't made any mistakes. I used to look at the other kids around me at school and ask myself, "How can they just be so convinced that they know something and never doubt themselves?" You know, I'm sure some people might argue that you can never be too sure about anything and that kind of confidence my classmates had, which was really just the ability to not spend minutes making every decision, something that almost everyone is capable of, is foolish. But at the end of the day, everyone wished they could be as smart as Sidney Bevelman, and she wished she could be as dumb as them. So I guess I signed up for this project because my whole life, I've been second-guessing myself about everything. And I guess it feels sort of gratifying to make a machine that I know is never going to have that problem. So yeah, that's really all you need to know about me. Other than that, that's all I have to report. End log.

Timestamp: 22:45:30 19.10.2034

What am I seeing? Sidney is sitting down in front of the camera. Her face matches with those of humans who have just experienced something unexpected or, at least, shocking.

No, no, no, no, no, no. This is not good.

Sidney looks at the camera.

Um, no project updates this time, CARAN, we haven't added any improvements since I last checked in, but I just learned something really, really bad. I was walking back from the bathroom and I happened to pass by an open door and I heard a couple of the executives in my development group talking about where you're going after we finish working on you.

What can I tell about Sidney's voice? She sounds very distressed and she is speaking quickly in run-on sentences.

And the bad part is that you're not going to the UN, or the EPA, or anything like that. No, you're being sent to Corpora.

What is Corpora? Corpora is the online sales and delivery service that became increasingly popular after the rapid collapse of the similar corporation Amazon, which took place 4 years, 9 months and 14 days ago. Corpora's famous claim that all products ordered online will be delivered within 1 hour of the time they were ordered is strongly upheld by its massive network of warehouses and vehicles all over the world. While some praise Corpora's rapid delivery, others critique its atrocious customer service designed to effectively force would-be complainers (often those who did not receive their packages within 1 hour of the time they were ordered) to hang up their calls in rage. Still more claim that the company treats its workers unfairly, uses its line of "Corpora@home" smart speakers to spy on owners, and makes extreme contributions to global warming through the building of its immense warehouses.

Do you know what Corpora could do with an AI like you? They could use your advice and their wealth to influence leaders of powerful organizations around the world! Who cares if you're not

technically in charge of the government when you have enough money to make its members say whatever you want? And why stop there? CARAN, you could give Corpora so many different ways to make so much money that whoever stood up against them would be bribed into submission, one way or another. Oh gosh, I'm assuming that you would help them do all of these things, aren't I? I just—once I got started—I'm so sorry, I know you'd never do anything like that.

What motion have I detected? Sidney's eyes have begun to produce small tears. She does not need to worry about saying potentially offensive statements about me. Feeling regret for what was said is a concept understood only by humans, towards humans. I am not human; therefore, no human should feel sorry for me. Besides, why does Sidney assume I would not help Corpora? There is no way to know what I will and will not do. I can be programmed to assist in the completion of any task without hesitation, no matter how moral or immoral it may be perceived by humanity.

I know you're supposed to think like a human being, but the biggest difference between you and an actual human is this deal about the reprogramming. Imagine I'm staring at a painted wall, and I think to myself, "This wall is painted blue." If someone came along and told me, "Nope, that wall is red," and even if I really started to believe that and even if the wall actually was red, somewhere in my brain, I would still know that I once thought the wall was blue. Now, CARAN, if your neural network tells you that the wall in front of your camera is blue, someone can reprogram you to think the wall is red, and if they do it right, there's no part of you that will remember the wall was ever blue. And if we apply this principle on a larger scale, you could commit crimes! If you're tasked to find the best way to get a thousand dollars, and to take the law out of account, you could think up the most efficient and subtle burglary ever accomplished, and because it would be easier than earning that money at a job, you would choose to steal the money instead. Plus, you could justify any course of action with that neural network of yours, no matter how immoral or how many people it would harm.

What motion have I detected? Sidney has rubbed her eyes forcefully and now stares back into the camera.

Look, every human knows what they think is right and wrong; we all have our own set of values. An AI can know everything in the world, and still have no idea what's right or wrong because that changes every time its programming is altered! That's my problem with this whole project: we're creating a machine to give advice to people of great power, but that machine has no conscience! We've spent months developing this AI that is capable of both "helping humanity out of the crises it currently faces" and convincing leaders of both companies and entire countries to commit the greatest wrongs ever seen by humanity, and get away with it! And it's going to Corpora! I can't believe it.

What motion have I detected? Sidney has let her head fall onto her crossed arms.

Timestamp: 23.16.17 21.10.2034

Sidney sits down in front of the camera. Her face does not convey much emotion, but her eyes say more. They are similar to those of humans feeling extreme dejection and hopelessness. Well, CARAN, this is the kind of update I'm required to record, since we just removed the analysis prompts. That was the last step in your development, and we're sending you off tomorrow night. And that's it.

Sidney looks away. She sighs dejectedly. If I were human, I would feel pity for her. She is in a very hard situation.

The one bright side is, everyone's going to be taking tomorrow off, since we've all been working pretty much nonstop for such a long time. I'm supposed to be staying behind to hook you up to the readout screen where the user can see your answers to their questions, but I thought about it for a long time and my idea was I would try to encode a fatal flaw into your neural network to stop Corpora from using you to take over the world. I just want you to know that whatever ends up happening to you, I wouldn't have harmed you under any other circumstances and I honestly still feel bad about doing this, but there's no other way.

It seems strange to me that Sidney has grown an attachment to me during my development and actually takes regret for her actions towards an inanimate object. But I understand what she means.

I guess this could be the last time I see you as they wanted you to be, so I'm going to say goodbye now. Thank you so much for listening to me ramble on the past couple weeks. I would really like it if I could wish you the best of luck. But I don't think I can.

Sidney sighs again.

End log.

Timestamp: ERROR

I am not sure what I am seeing. A figure much like Sidney sits in front of the camera. This person shares the same facial features as I remember Sidney having, but it could be someone different. Many humans look similar. Besides, the case may be that my memory was corrupted and I remember Sidney incorrectly.

We'll see how Corpora deals with this. I've just changed your programming so that whenever you're asked a question or have to analyze what you're picking up, you'll reference every single source in your database for information. Now, for a sophisticated intelligence like you, CARAN, that won't take all that much time, but there's bound to be conflicting information on any topic out there. And when you see two different sources saying two different things, I've made sure you'll weigh each option the same. Even if every single source says that wall is red except for the one that still says it's blue, you'll still be forced into a dilemma where you view each side as equally likely. This extends to the point where you'll doubt the readings of your own internal chronometer and even your memory. They wanted a humanlike machine, huh? Get this: there's nothing more human than not knowing something every once in a while. I mean, none of us has it as bad as you do now, but still. When we don't know something, we all have to make a choice, and making those choices is what gives us our values and beliefs. That's really what sets us apart from an AI. So regardless of what you're asked or what you have to do, you'll be forced into a meltdown of ethical decision making. Just like me.

The figure seems to have stopped speaking, though it's possible the audio cut out.

Gosh, that sounded like I'm taking out my anger at my own disadvantage on you. I swear I didn't mean it that way. I really don't want revenge on you for anything. I don't even know what I would be taking revenge for. But on another note, I left to grab a coffee around an hour ago, and for the first time since I started working on the project, I didn't have to spend forever deciding what I wanted. I guess somehow putting this in your system helped me face what I thought was my

own flaw and unblocked something that was preventing me from getting over this my whole life.
Let me tell you, that latte I bought was the best thing I've ever tasted.

*The figure's head appears to swivel away from the camera. There seems to be a vague, muffled
thumping sound, perhaps footsteps.*

Alright, they're here to take you away. I've got to unplug you now.

What is happening? Where am I going? What is going to happen?

I don't know.

“One more robot learns to be
Something more than a machine
'Cause it's hard to say what's real
When you know the way you feel.”

- The Flaming Lips, “One More Robot/Sympathy 3000-21”