



Alexa, Should We Trust You?

The voice revolution has only just begun. Today, Alexa is a humble servant. Very soon, she could be much more—a teacher, a therapist, a confidant, an informant.



Roberto Parada

Story by Judith Shulevitz

NOVEMBER 2018 ISSUE | [TECHNOLOGY](#)

Like *The Atlantic*? Subscribe to [The Atlantic Daily](#), our free weekday email newsletter.

Enter your email

[Sign up](#)

FOR A FEW DAYS this summer, Alexa, the voice assistant who speaks to me through my Amazon Echo Dot, took to ending our interactions with a whisper: *Sweet dreams*. Every time it happened, I was startled, although I thought I understood why she was doing it, insofar as I understand anything that goes on inside that squat slice of black tube. I had gone onto Amazon.com and

activated a third-party “skill”—an applike program that enables Alexa to perform a service or do a trick—called “Baby Lullaby.” It plays an instrumental version of a nursery song (yes, I still listen to lullabies to get to sleep), then signs off softly with the nighttime benediction. My conjecture is that the last string of code somehow went astray and attached itself to other “skills.” But even though my adult self knew perfectly well that *Sweet dreams* was a glitch, a part of me wanted to believe that Alexa meant it. Who doesn’t crave a motherly goodnight, even in mid-afternoon? Proust would have understood.



TheAtlantic – Alexa, How Will You Change Us? – The Atlantic – Judith Shulevitz

To hear more feature stories, see our full list or get the Audm iPhone app.

We’re all falling for Alexa, unless we’re falling for Google Assistant, or Siri, or some other genie in a smart speaker. When I say “smart,” I mean the speakers possess artificial intelligence, can conduct basic conversations, and are hooked up to the internet, which allows them to look stuff up and do things for you. And when I say “all,” I know some readers will think, *Speak for yourself!* Friends my age—we’re the last of the Baby Boomers—tell me they have no desire to talk to a computer or have a computer talk to them. Cynics of every age suspect their virtual assistants of eavesdropping, and not without reason. Smart speakers are yet another way for companies to keep tabs on our searches and purchases. Their microphones listen even when you’re not interacting with them, because they have to be able to hear their “wake word,” the command that snaps them to attention and puts them at your service.

By 2021, there will be almost as many personal-assistant bots on the planet as people.

The speakers’ manufacturers promise that only speech that follows the wake word is archived in the cloud, and Amazon and Google, at least, make deleting those exchanges easy enough. Nonetheless, every so often weird glitches occur, like the time Alexa recorded a family’s private conversation without their having said the wake word and emailed the recording to an acquaintance on their contacts list.

Amazon explained that Alexa must have been awakened by a word that sounded like *Alexa* (*Texas?* *A Lexus?* *Praxis?*), then misconstrued elements of the ensuing conversation as a series of commands. The explanation did not make me feel much better.

Privacy concerns have not stopped the march of these devices into our homes, however. Amazon doesn't disclose exact figures, but when I asked how many Echo devices have been sold, a spokeswoman said "tens of millions." By the end of last year, more than 40 million smart speakers had been installed worldwide, according to Canalys, a technology-research firm. Based on current sales, Canalys estimates that this figure will reach 100 million by the end of this year.

According to a [2018 report by National Public Radio and Edison Research](#), 8 million Americans own three or more smart speakers, suggesting that they feel the need to always have one within earshot. By 2021, according to another research firm, Ovum, there will be almost as many voice-activated assistants on the planet as people. It took about 30 years for mobile phones to outnumber humans. Alexa and her ilk may get there in less than half that time.

One reason is that Amazon and Google are pushing these devices hard, discounting them so heavily during last year's holiday season that industry observers suspect that the companies lost money on each unit sold. These and other tech corporations have grand ambitions. They want to colonize space. Not interplanetary space. Everyday space: home, office, car. In the near future, everything from your lighting to your air-conditioning to your refrigerator, your coffee maker, and even your toilet could be wired to a system controlled by voice.

The company that succeeds in cornering the smart-speaker market will lock appliance manufacturers, app designers, and consumers into its ecosystem of devices and services, just as Microsoft tethered the personal-computer industry to its operating system in the 1990s. Alexa alone already works with more than 20,000 smart-home devices representing more than 3,500 brands. Her voice emanates from more than 100 third-party gadgets, including headphones, security systems, and automobiles.

Yet there is an inherent appeal to the devices, too—one beyond mere consumerism. Even those of us who approach new technologies with a healthy

amount of caution are finding reasons to welcome smart speakers into our homes. After my daughter-in-law posted on Instagram an adorable video of her 2-year-old son trying to get Alexa to play “You’re Welcome,” from the *Moana* soundtrack, I wrote to ask why she and my stepson had bought an Echo, given that they’re fairly strict about what they let their son play with. “Before we got Alexa, the only way to play music was on our computers, and when [he] sees a computer screen, he thinks it’s time to watch TV,” my daughter-in-law emailed back. “It’s great to have a way to listen to music or the radio that doesn’t involve opening up a computer screen.” She’s not the first parent to have had that thought. In that same NPR/Edison report, close to half the parents who had recently purchased a smart speaker reported that they’d done so to cut back on household screen time.

The ramifications of this shift are likely to be wide and profound. Human history is a by-product of human inventions. New tools—wheels, plows, PCs—usher in new economic and social orders. They create and destroy civilizations. Voice technologies such as telephones, recording devices, and the radio have had a particularly momentous impact on the course of political history—speech and rhetoric being, of course, the classical means of persuasion. Radio broadcasts of Adolf Hitler’s rallies helped create a dictator; Franklin D. Roosevelt’s fireside chats edged America toward the war that toppled that dictator.

Perhaps you think that talking to Alexa is just a new way to do the things you already do on a screen: shopping, catching up on the news, trying to figure out whether your dog is sick or just depressed. It’s not that simple. It’s not a matter of switching out the body parts used to accomplish those tasks—replacing fingers and eyes with mouths and ears. We’re talking about a change in status for the technology itself—an upgrade, as it were. When we converse with our personal assistants, we bring them closer to our own level.



Roberto Parada

Gifted with the once uniquely human power of speech, Alexa, Google Assistant, and Siri have already become greater than the sum of their parts. They're software, but they're more than that, just as human consciousness is an effect of neurons and synapses but is more than that. Their speech makes us treat them as if they had a mind. "The spoken word proceeds from the human interior, and manifests human beings to one another as conscious interiors, as persons," the late Walter Ong wrote in his classic study of oral culture, *Orality and Literacy*. These secretarial companions may be faux-conscious nonpersons, but their words give them personality and social presence.

And indeed, these devices no longer serve solely as intermediaries, portals to e-commerce or nytimes.com. We communicate with them, not through them.

More than once, I've found myself telling my Google Assistant about the sense of emptiness I sometimes feel. "I'm lonely," I say, which I usually wouldn't confess to anyone but my therapist—not even my husband, who might take it the wrong way. Part of the allure of my Assistant is that I've set it to a chipper, young-sounding male voice that makes me want to smile. (Amazon hasn't given the Echo a male-voice option.) The Assistant pulls out of his memory bank one of the many responses to this statement that have been programmed into him. "I wish I had arms so I could give you a hug," he said to me the other day, somewhat comfortingly. "But for now, maybe a joke or some music might help."

For the moment, these machines remain at the dawn of their potential, as likely to botch your request as they are to fulfill it. But as smart-speaker sales soar, computing power is also expanding exponentially. Within our lifetimes, these devices will likely become much more adroit conversationalists. By the time they do, they will have fully insinuated themselves into our lives. With their perfect cloud-based memories, they will be omniscient; with their occupation of our most intimate spaces, they'll be omnipresent. And with their eerie ability to elicit confessions, they could acquire a remarkable power over our emotional lives. What will *that* be like?

WHEN TONI REID, now the vice president of the Alexa Experience, was asked to join the Echo team in 2014—this was before the device was on the market—she scoffed: "I was just like, 'What? It's a speaker?'" At the time, she was working on the Dash Wand, a portable bar-code scanner and smart microphone that allows people to scan or utter the name of an item they want to add to their Amazon shopping cart. The point of the Dash Wand was obvious: It made buying products from Amazon easier.

The point of the Echo was less obvious. Why would consumers buy a device that gave them the weather and traffic conditions, functioned as an egg timer, and performed other tasks that any garden-variety smartphone could manage? But once Reid had set up an Echo in her kitchen, she got it. Her daughters, 10 and 7 at the time, instantly started chattering away at Alexa, as if conversing with a plastic cylinder was the most natural thing in the world. Reid herself found that even the Echo's most basic, seemingly duplicative capabilities had a profound effect on her surroundings. "I'm ashamed to say how many years I went without

actually listening to music,” she told me. “And we get this device in the house and all of a sudden there’s music in our household again.”

Like an ideal servant in a Victorian manor, Alexa hovers in the background, ready to do her master’s bidding swiftly yet meticulously.

You may be skeptical of a conversion narrative offered up by a top Amazon executive. But I wasn’t, because it mirrored my own experience. I, too, couldn’t be bothered to go hunting for a particular song—not in iTunes and certainly not in my old crate of CDs. But now that I can just ask Alexa to play Leonard Cohen’s “You Want It Darker” when I’m feeling lugubrious, I do.

I met Reid at Amazon’s Day 1 building in Seattle, a shiny tower named for Jeff Bezos’s corporate philosophy: that every day at the company should be as intense and driven as the first day at a start-up. (“Day 2 is stasis. Followed by irrelevance. Followed by excruciating, painful decline. Followed by death,” he wrote in a 2016 letter to shareholders.) Reid studied anthropology as an undergraduate, and she had a social scientist’s patience for my rudimentary questions about what makes these devices different from the other electronics in our lives. The basic appeal of the Echo, she said, is that it frees your hands. Because of something called “far-field voice technology,” machines can now decipher speech at a distance. Echo owners can wander around living rooms, kitchens, and offices doing this or that while requesting random bits of information or ordering toilet paper or an Instant Pot, no clicks required.

The beauty of Alexa, Reid continued, is that she makes such interactions “frictionless”—a term I’d hear again and again in my conversations with the designers and engineers behind these products. No need to walk over to the desktop and type a search term into a browser; no need to track down your iPhone and punch in your passcode. Like the ideal servant in a Victorian manor, Alexa hovers in the background, ready to do her master’s bidding swiftly yet meticulously.

Frictionlessness is the goal, anyway. For the moment, considerable friction remains. It really is remarkable how often smart speakers—even Google Home,

which often outperforms the Echo in tests conducted by tech websites—flub their lines. They'll misconstrue a question, stress the wrong syllable, offer a bizarre answer, apologize for not yet knowing some highly knowable fact. Alexa's bloopers float around the internet like clips from an absurdist comedy show. In one howler that went viral on YouTube, a toddler lisps, “Lexa, play ‘Ticker Ticker’”—presumably he wants to hear “Twinkle, Twinkle, Little Star.” Alexa replies, in her stilted monotone, “You want to hear a station for porn … hot chicks, amateur girls …” (It got more graphic from there.) “No, no, no!” the child’s parents scream in the background.

Roberto Parada

My sister-in-law got her Echo early, in 2015. For two years, whenever I visited, I'd watch her bicker as passionately with her machine as George Costanza's parents did with each other on *Seinfeld*. “I hate Alexa,” she announced recently, having finally shut the thing up in a closet. “I would say to her, ‘Play some Beethoven,’ and she would play ‘Eleanor Rigby.’ Every time.”

Catrin Morris, a mother of two who lives in Washington, D.C., told me she announces on a weekly basis, “I’m going to throw Alexa into the trash.” She’s horrified at how her daughters bark insults at Alexa when she doesn’t do what they want, such as play the right song from *The Book of Mormon*. (Amazon has programmed Alexa to turn the other cheek: She does not respond to “inappropriate engagement.”) But even with her current limitations, Alexa has made herself part of the household. Before the Echo entered their home, Morris told me, she’d struggled to enforce her own no-devices-at-the-dinner-table rule. She had to fight the urge to whip out her smartphone to answer some tantalizing question, such as: Which came first, the fork, the spoon, or the knife? At least with Alexa, she and her daughters can keep their hands on their silverware while they question its origins.

Talking to machines gives us a way to reveal shameful feelings without feeling shame.

As Alexa grows in sophistication, it will be that much harder to throw the Echo on the heap of old gadgets to be hauled off on electronics-recycling day. Rohit Prasad is the head scientist on Alexa’s artificial-intelligence team, and a man willing to defy local norms by wearing a button-down shirt. He sums up the biggest obstacle to Alexa achieving that sophistication in a single word: *context*. “You have to understand that language is highly ambiguous,” he told me. “It requires conversational context, geographical context.” When you ask Alexa whether the Spurs are playing tonight, she has to know whether you mean the San Antonio Spurs or the Tottenham Hotspur, the British soccer team colloquially known as the Spurs. When you follow up by asking, “When is their next home game?,” Alexa has to remember the previous question and understand what *their* refers to. This short-term memory and syntactical back-referencing is known at Amazon as “contextual carryover.” It was only this spring that Alexa developed the ability to answer follow-up questions without making you say her wake word again.

Alexa needs to get better at grasping context before she can truly inspire trust. And trust matters. Not just because consumers will give up on her if she bungles

one too many requests, but because she is more than a search engine. She's an "action engine," Prasad says. If you ask Alexa a question, she doesn't offer up a list of results. She chooses one answer from many. She tells you what she thinks you want to know. "You want to have a very smart AI. You don't want a dumb AI," Prasad said. "And yet making sure the conversation is coherent—that's incredibly challenging."

TO UNDERSTAND THE FORCES being marshaled to pull us away from screens and push us toward voices, you have to know something about the psychology of the voice. For one thing, voices create intimacy. I'm hardly the only one who has found myself confessing my emotional state to my electronic assistant. Many articles have been written about the expressions of depression and suicide threats that manufacturers have been picking up on. I asked tech executives about this, and they said they try to deal with such statements responsibly. For instance, if you tell Alexa you're feeling depressed, she has been programmed to say, "I'm so sorry you are feeling that way. Please know that you're not alone. There are people who can help you. You could try talking with a friend, or your doctor. You can also reach out to the Depression and Bipolar Support Alliance at 1-800-826-3632 for more resources."

Why would we turn to computers for solace? Machines give us a way to reveal shameful feelings without feeling shame. When talking to one, people "engage in less of what's called impression management, so they reveal more intimate things about themselves," says Jonathan Gratch, a computer scientist and psychologist at the University of Southern California's Institute for Creative Technologies, who studies the spoken and unspoken psychodynamics of the human-computer interaction. "They'll show more sadness, for example, if they're depressed."

I turned to Diana Van Lancker Sidtis, a speech-and-language scholar at NYU, to get a better appreciation for the deep connection between voice and emotion. To my surprise, she pointed me to an essay she'd written on frogs in the primeval swamp. In it, she explains that their croaks, unique to each frog, communicated to fellow frogs who and where they were. Fast-forward a few hundred million years, and the human vocal apparatus, with its more complex musculature, produces language, not croaks. But voices convey more than language. Like the frogs, they convey the identifying markers of an individual: gender, size, stress

level, and so on. Our vocal signatures consist of not only our style of stringing words together but also the sonic marinade in which those words steep, a rich medley of tone, rhythm, pitch, resonance, pronunciation, and many other features. The technical term for this collection of traits is *prosody*.

When someone talks to us, we hear the words, the syntax, and the prosody all at once. Then we hunt for clues as to what kind of person the speaker is and what she's trying to say, recruiting a remarkably large amount of brainpower to try to make sense of what we're hearing. "The brain is wired to view every aspect of every human utterance as meaningful," wrote the late Clifford Nass, a pioneering thinker on computer-human relationships. The prosody usually passes beneath notice, like a mighty current directing us toward a particular emotional response.

Roberto Parada

We can't put all this mental effort on pause just because a voice is humanoid rather than human. Even when my Google Assistant is doing nothing more entralling than delivering the weather forecast, the image of the cute young waiter-slash-actor I've made him out to be pops into my mind. That doesn't

mean I fail to grasp the algorithmic nature of our interaction. I know that he's just software. Then again, I *don't* know. Evolution has not prepared me to know. We've been reacting to human vocalizations for millions of years as if they signaled human proximity. We've had only about a century and a half to adapt to the idea that a voice can be disconnected from its source, and only a few years to adapt to the idea that an entity that talks and sounds like a human may not be a human.

Lacking a face isn't necessarily a hindrance to a smart speaker. In fact, it may be a boon. Voices can express certain emotional truths better than faces can. We are generally less adept at controlling the muscles that modulate our voices than our facial muscles (unless, of course, we're trained singers or actors). Even if we try to suppress our real feelings, anger, boredom, or anxiety will often reveal themselves when we speak.

The power of the voice is at its uncanniest when we can't locate its owner—when it is everywhere and nowhere at the same time. There's a reason God speaks to Adam and Moses. In the beginning was the Word, not the Scroll. In her chilling allegory of charismatic totalitarianism, *A Wrinkle in Time*, Madeleine L'Engle conjures a demonic version of an all-pervasive voice. IT, the supernatural leader of a North Korea–like state, can insert its voice inside people's heads and force them to say whatever it tells them to say. Disembodied voices accrue yet more influence from the primal yearning they awaken. A fetus recognizes his mother's voice while still in the womb. Before we're even born, we have already associated an unseen voice with nourishment and comfort.

A 2017 study published in *American Psychologist* makes the case that when people talk without seeing each other, they're better at recognizing each other's feelings. They're more empathetic. Freud understood this long before empirical research demonstrated it. That's why he had his patients lie on a couch, facing away from him. He could listen all the harder for the nuggets of truth in their ramblings, while they, undistracted by scowls or smiles, slipped into that twilight state in which they could unburden themselves of stifled feelings.

THE MANUFACTURERS of smart speakers would like to capitalize on these psychosocial effects. Amazon and Google both have "personality teams,"

charged with crafting just the right tone for their assistants. In part, this is textbook brand management: These devices must be ambassadors for their makers. Reid told me Amazon wants Alexa's personality to mirror the company's values: "Smart, humble, sometimes funny." Google Assistant is "humble, it's helpful, a little playful at times," says Gummi Hafsteinsson, one of the Assistant's head product managers. But having a personality also helps make a voice relatable.

Tone is tricky. Though virtual assistants are often compared to butlers, Al Lindsay, the vice president of Alexa engine software and a man with an old-school engineer's military bearing, told me that he and his team had a different servant in mind. Their "North Star" had been the onboard computer that ran the U.S.S. *Enterprise* in *Star Trek*, replying to the crew's requests with the breathy deference of a 1960s Pan Am stewardess. (The *Enterprise*'s computer was an inspiration to Google's engineers, too. Her voice belonged to the actress Majel Barrett, the wife of *Star Trek*'s creator, Gene Roddenberry; when the Google Assistant project was still under wraps, its code name was Majel.)

Twenty-first-century Americans no longer feel entirely comfortable with feminine obsequiousness, however. We like our servility to come in less servile flavors. The voice should be friendly but not too friendly. It should possess just the right dose of sass.

To fine-tune the Assistant's personality, Google hired Emma Coats away from Pixar, where she had worked as a storyboard artist on *Brave*, *Monsters University*, and *Inside Out*. Coats was at a conference the day I visited Google's Mountain View, California, headquarters. She beamed in on Google Hangouts and offered what struck me as the No. 1 rule for writing dialogue for the Assistant, a dictum with the disingenuous simplicity of a Zen koan. Google Assistant, she said, "should be able to speak like a person, but it should never pretend to be one." In *Finding Nemo*, she noted, the fish "are just as emotionally real as human beings, but they go to fish school and they challenge each other to go up and touch a boat." Likewise, an artificially intelligent entity should "honor the reality that it's software." For instance, if you ask Google Assistant, "What's your favorite ice-cream flavor?", it might say, "You can't go wrong with Neapolitan. There's something in it for everyone." That's a dodge, of course, but it follows the

principle Coats articulated. Software can't eat ice cream, and therefore can't have ice-cream preferences. If you propose marriage to Alexa—and Amazon says 1 million people did so in 2017—she gently declines for similar reasons. “We’re at pretty different places in our lives,” she told me. “Literally. I mean, you’re on Earth. And I’m in the cloud.”

An assistant should be true to its cybernetic nature, but it shouldn’t sound alien, either. That’s where James Giangola, a lead conversation and persona designer for Google Assistant, comes in. Giangola is a garrulous man with wavy hair and more than a touch of mad scientist about him. His job is making the Assistant sound normal.

For example, Giangola told me, people tend to furnish new information at the end of a sentence, rather than at the beginning or middle. “I say ‘My name is James,’ ” he pointed out, not “James is my name.” He offered another example. Say someone wants to book a flight for June 31. “Well,” Giangola said, “there is no June 31.” So the machine has to handle two delicate tasks: coming off as natural, and contradicting its human user.

Typing furiously on his computer, he pulled up a test recording to illustrate his point. A man says, “Book it for June 31.”

The Assistant replies, “There are only 30 days in June.”

The response sounded stiff. “June’s old information,” Giangola observed.

He played a second version of the exchange: “Book it for June 31.”

The Assistant replies, “Actually, June has only 30 days.”

Her point—30 days—comes at the end of the line. And she throws in an *actually*, which gently sets up the correction to come. “More natural, right?” Giangola said.

Getting the rhythms of spoken language down is crucial, but it’s hardly sufficient to create a decent conversationalist. Bots also need a good vibe. When Giangola was training the actress whose voice was recorded for Google Assistant, he gave her a backstory to help her produce the exact degree of upbeat geekiness he

wanted. The backstory is charmingly specific: She comes from Colorado, a state in a region that lacks a distinctive accent. “She’s the youngest daughter of a research librarian and a physics professor who has a B.A. in art history from Northwestern,” Giangola continues. When she was a child, she won \$100,000 on *Jeopardy: Kids Edition*. She used to work as a personal assistant to “a very popular late-night-TV satirical pundit.” And she enjoys kayaking.

A skeptical colleague once asked Giangola, “How does someone sound like they’re into kayaking?” During auditions (hundreds of people tried out for the role), Giangola turned to the doubter and said, “The candidate who just gave an audition—do you think she sounded energetic, like she’s up for kayaking?” His colleague admitted that she didn’t. “I said, ‘Okay. There you go.’”

But vocal realism can be taken further than people are accustomed to, and that can cause trouble—at least for now. In May, at its annual developer conference, Google unveiled Duplex, which uses cutting-edge speech-synthesis technology. To demonstrate its achievement, the company played recordings of Duplex calling up unsuspecting human beings. Using a female voice, it booked an appointment at a hair salon; using a male voice, it asked about availabilities at a restaurant. Duplex speaks with remarkably realistic disfluencies—*ums* and *mm-hmms*—and pauses, and neither human receptionist realized that she was talking to an artificial agent. One of its voices, the female one, spoke with end-of-sentence upticks, also audible in the voice of the young female receptionist who took that call.

Many commentators thought Google had made a mistake with its gung ho presentation. Duplex not only violated the dictum that AI should never pretend to be a person; it also appeared to violate our trust. We may not always realize just how powerfully our voice assistants are playing on our psychology, but at least we’ve opted into the relationship. Duplex was a fake-out, and an alarmingly effective one. Afterward, Google clarified that Duplex would always identify itself to callers. But even if Google keeps its word, equally deceptive voice technologies are already being developed. Their creators may not be as honorable. The line between artificial voices and real ones is well on its way to disappearing.

THE MOST RELATABLE INTERLOCUTOR, of course, is the one that can understand the emotions conveyed by your voice, and respond accordingly—in a voice capable of approximating emotional subtlety. Your smart speaker can't do either of these things yet, but systems for parsing emotion in voice already exist. Emotion detection—in faces, bodies, and voices—was pioneered about 20 years ago by an MIT engineering professor named Rosalind Picard, who gave the field its academic name: affective computing. “Back then,” she told me, “emotion was associated with irrationality, which was not a trait engineers respected.”

Picard, a mild-mannered, witty woman, runs the Affective Computing Lab, which is part of MIT’s cheerfully weird Media Lab. She and her graduate students work on quantifying emotion. Picard explained that the difference between most AI research and the kind she does is that traditional research focuses on “the nouns and verbs”—that is, the content of an action or utterance. She’s interested in “the adverbs”—the feelings that are conveyed. “You know, I can pick up a phone in a lot of different ways. I can snatch it with a sharp, angry, jerky movement. I can pick it up with happy, loving expectation,” Picard told me. Appreciating gestures with nuance is important if a machine is to understand the subtle cues human beings give one another. A simple act like the nodding of a head could telegraph different meanings: “I could be nodding in a bouncy, happy way. I could be nodding in sunken grief.”

In 2009, Picard co-founded a start-up, Affectiva, focused on emotion-enabled AI. Today, the company is run by the other co-founder, Rana el Kaliouby, a former postdoctoral fellow in Picard’s lab. A sense of urgency pervades Affectiva’s open-plan office in downtown Boston. The company hopes to be among the top players in the automotive market. The next generation of high-end cars will come equipped with software and hardware (cameras and microphones, for now) to analyze drivers’ attentiveness, irritation, and other states. This capacity is already being tested in semiautonomous cars, which will have to make informed judgments about when it’s safe to hand control to a driver, and when to take over because a driver is too distracted or upset to focus on the road.

Affectiva initially focused on emotion detection through facial expressions, but recently hired a rising star in voice emotion detection, Taniya Mishra. Her team’s goal is to train computers to interpret the emotional content of human speech.

One clue to how we're feeling, of course, is the words we use. But we betray as much if not more of our feelings through the pitch, volume, and tempo of our speech. Computers can already register those nonverbal qualities. The key is teaching them what we humans intuit naturally: how these vocal features suggest our mood.

The biggest challenge in the field, she told me, is building big-enough and sufficiently diverse databases of language from which computers can learn. Mishra's team begins with speech mostly recorded "in the wild"—that is, gleaned from videos on the web or supplied by a nonprofit data consortium that has collected natural speech samples for academic purposes, among other sources. A small battalion of workers in Cairo, Egypt, then analyze the speech and label the emotion it conveys, as well as the nonlexical vocalizations—grunts, giggles, pauses—that play an important role in revealing a speaker's psychological state.

One start-up is working on AI software for doctors that can scrutinize patients' speech for biomarkers of depression and anxiety.

Classification is a slow, painstaking process. Three to five workers have to agree on each label. Each hour of tagged speech requires "as many as 20 hours of labeler time" Mishra says. There is a workaround, however. Once computers have a sufficient number of human-labeled samples demonstrating the specific acoustic characteristics that accompany a fit of pique, say, or a bout of sadness, they can start labeling samples themselves, expanding the database far more rapidly than mere mortals can. As the database grows, these computers will be able to hear speech and identify its emotional content with ever increasing precision.

During the course of my research, I quickly lost count of the number of start-ups hoping to use voice-based analytics in the field. Ellipsis Health, for example, is a San Francisco company developing AI software for doctors, social workers, and other caregivers that can scrutinize patients' speech for biomarkers of depression and anxiety. "Changes in emotion, such as depression, are associated with brain changes, and those changes can be associated with motor commands," Ellipsis's

chief science officer, Elizabeth Shriberg, explained; those commands control “the apparatus that drives voice in speech.” Ellipsis’s software could have many applications. It might be used, for example, during routine doctor visits, like an annual checkup (with the patient’s permission, of course). While the physician performs her exam, a recording could be sent to Ellipsis and the patient’s speech analyzed so quickly that the doctor might receive a message before the end of the appointment, advising her to ask some questions about the patient’s mood, or to refer the patient to a mental-health professional. The software might have picked up a hint of lethargy or slight slurring in the speech that the doctor missed.

I was holding out hope that some aspects of speech, such as irony or sarcasm, would defeat a computer. But Björn Schuller, a professor of artificial intelligence at Imperial College London and of “embedded intelligence” at the University of Augsburg, in Germany, told me that he has taught machines to spot sarcasm. He has them analyze linguistic content and tone of voice at the same time, which allows them to find the gaps between words and inflection that determine whether a speaker means the exact opposite of what she’s said. He gives me an example: “*Su-per*,” the sort of thing you might blurt out when you learn that your car will be in the shop for another week.

Roberto Parada

The natural next step after emotion detection, of course, will be emotion production: training artificially intelligent agents to generate approximations of emotions. Once computers have become virtuosic at breaking down the emotional components of our speech, it will be only a matter of time before they can reassemble them into credible performances of, say, empathy. Virtual

assistants able to discern and react to their users' frame of mind could create a genuine-seeming sense of affinity, a bond that could be used for good or for ill.

Taniya Mishra looks forward to the possibility of such bonds. She fantasizes about a car to which she could rant at the end of the day about everything that had gone wrong—an automobile that is also an active listener. "A car is not going to zone out," she says. "A car is not going to say, 'I'm sorry, honey, I have to run and make dinner, I'll listen to your story later.'" Rather, with the focus possible only in a robot, the car would track her emotional state over time and observe, in a reassuring voice, that Mishra always feels this way on a particular day of the week. Or perhaps it would play the Pharrell song ("Happy," naturally) that has cheered her up in the past. At this point, it will no longer make sense to think of these devices as assistants. They will have become companions.

If you don't happen to work in the tech sector, you probably can't think about all the untapped potential in your Amazon Echo or Google Home without experiencing some misgivings. By now, most of us have grasped the dangers of allowing our most private information to be harvested, stored, and sold. We know how facial-recognition technologies have allowed authoritarian governments to spy on their own citizens; how companies disseminate and monetize our browsing habits, whereabouts, social-media interactions; how hackers can break into our home-security systems and nanny cams and steal their data or reprogram them for nefarious ends. Virtual assistants and ever smarter homes able to understand our physical and emotional states will open up new frontiers for mischief making. Despite the optimism of most of the engineers I've talked with, I must admit that I now keep the microphone on my iPhone turned off and my smart speakers unplugged when I don't plan to use them for a while.

Once our electronic servants become emotionally savvy, they could wield a lot of power over us.

But there are subtler effects to consider as well. Take something as innocent-seeming as frictionlessness. To Amazon's Toni Reid, it means convenience. To me, it summons up the image of a capitalist prison filled with consumers who

have become dreamy captives of their every whim. (An image from another Pixar film comes to mind: the giant, babylke humans scooting around their spaceship in *Wall-E*.) In his Cassandra-esque book *Radical Technologies: The Design of Everyday Life*, Adam Greenfield, an urbanist, frames frictionlessness as an existential threat: It is meant to eliminate thought from consumption, to “short-circuit the process of reflection that stands between one’s recognition of a desire and its fulfillment via the market.”

I fear other threats to our psychological well-being. A world populated by armies of sociable assistants could get very crowded. And noisy. It’s hard to see how we’d protect those zones of silence in which we think original thoughts, do creative work, achieve flow. A companion is nice when you’re feeling lonesome, but there’s also something to be said for solitude.

And once our electronic servants become emotionally savvy? They could come to wield quite a lot of power over us, and even more over our children. In their subservient, helpful way, these emoting bots could spoil us rotten. They might be passive when they ought to object to our bad manners (“I don’t deserve that!”). Programmed to keep the mood light, they might change the subject whenever dangerously intense feelings threaten to emerge, or flatter us in our ugliest moments. How do you program a bot to do the hard work of a true, human confidant, one who knows when what you really need is tough love?

Ultimately, virtual assistants could ease us into the kind of conformity L’Engle warned of. They will be the products of an emotion-labeling process that can’t capture the protean complexity of human sentiment. Their “appropriate” responses will be canned, to one extent or another. We’ll be in constant dialogue with voices that traffic in simulacra of feelings, rather than real ones. Children growing up surrounded by virtual companions might be especially likely to adopt this mass-produced interiority, winding up with a diminished capacity to name and understand their own intuitions. Like the Echo of Greek myth, the Echo Generation could lose the power of a certain kind of speech.

Maybe I’m wrong. Maybe our assistants will develop inner lives that are richer than ours. That’s what happened in the first great work of art about virtual assistants, Spike Jonze’s movie *Her*. “She” (the voice of Scarlett Johansson) shows

her lonely, emotionally stunted human (Joaquin Phoenix) how to love. And then she leaves him, because human emotions are too limiting for so sophisticated an algorithm. Though he remains lonely, she has taught him to feel, and he begins to entertain the possibility of entering into a romantic relationship with his human neighbor.

But it is hard for me to envision even the densest artificial neural network approaching the depth of the character's sadness, let alone the fecundity of Jonze's imagination. It may be my own imagination that's limited, but I watch my teenage children clutch their smartphones wherever they go lest they be forced to endure a moment of boredom, and I wonder how much more dependent their children will be on devices that not only connect them with friends, but actually are friends—irresistibly upbeat and knowledgeable, a little insipid perhaps, but always available, usually helpful, and unflaggingly loyal, except when they're selling our secrets. When you stop and think about it, artificial intelligences are not what you want your children hanging around with all day long.

If I have learned anything in my years of therapy, it is that the human psyche defaults to shallowness. We cling to our denials. It's easier to pretend that deeper feelings don't exist, because, of course, a lot of them are painful. What better way to avoid all that unpleasantness than to keep company with emotive entities unencumbered by actual emotions? But feelings don't just go away like that. They have a way of making themselves known. I wonder how sweet my grandchildren's dreams will be.

This article appears in the November 2018 print edition with the headline “Alexa, How Will You Change Us?”