1 RISE OF THE SEDUCTIVE ASSISTANTS

> Amazon today quietly unveiled a new product dubbed Amazon Echo. The $200 device appears to be a voice-activated wireless speaker that can answer your questions, offer updates on what's going on in the world, and of course play music. Echo is currently available for purchase via an invite-only system. If you have Amazon Prime, however, you can get it for $100. . . . Amazon wants to bring the digital assistant to the living room. The idea is a very interesting one, but it's difficult to imagine there being a lot of demand. Given that many of these features are already offered in mobile devices, most users will be happy to continue getting updates to their assistants there. Then again, Google Now, Siri, and Cortana are far from perfect, so Amazon does have some wiggle room. We'll have to reserve further judgment until we can get our hands on one.[1]

These comments began and ended a rather short VentureBeat article published on Alexa's launch day, November 6, 2014. The piece was obviously correct that Amazon wanted to copy the success of the digital assistant, the voice-enabled phone helper championed in the United States by Apple, Google, and Microsoft Cortana. In a video introducing the Echo, Amazon portrayed the living room and kitchen as landing spots for its hands-free assistant, Alexa. In subsequent years Amazon would try to colonize the entire home with Echo devices. That goal was made very clear in the launch-day video, which featured testimonials by customers who had tested the product before its release. Among the uses they excitedly mentioned were finding out the weather, helping with recipe measurements, learning new jokes, reading books, helping a blind couple to set timers, playing the news, and compiling shopping lists. "The Echo," one customer said, "is a tool we use to keep our household functioning."[2]

The video also shows that right from the start, Amazon used a strategy of seductive surveillance. It presented biometric identification and profiling as part of the device's features. The stories in the video demonstrate how the Echo recognizes individuals by their voices; in one home, Alexa learns to understand a man's English despite his German accent. Any concerns about this level of knowledge are quietly swept aside by users' enthusiasm for the device; we're encouraged to see only the benefits of talking to our own Echo, which in turn can hear and remember each of us. To further set the hook, Amazon offered an early-purchase discount to its most trusting customers, the Prime members whose loyalty to the company had earned them free shipping and other benefits. Such discounts would become key parts of Amazon's long-term seductive surveillance strategy.

Despite the company's exuberance, some articles belittled the new stationary assistant. James O'Toole with CNN Business, for example, commented that "Amazon's quirky Echo is Siri in a speaker" and that "this may be another case of a product that you can render superfluous by simply taking your phone out of your pocket."[3] Others, though, marveled at Amazon's boldness in entering a technology realm that was both mind-bogglingly complex and already filled with competition. Still others didn't seem to get it. CNET dutifully ran a story, headlined "Amazon Debuts Siri-like Digital Assistant Echo for Your Home," but—perhaps indicating the writers' low engagement with the product—neglected to mention that Alexa would use artificial intelligence to interact with family members and would assess and retain what they asked about a wide range of topics. Also missing from the announcement: that Alexa's setup app would ask where in the home the Echo was placed and would request that each user create an identifying voiceprint. What these individuals asked the intelligent speaker, as well as how they asked it and in what room, gave Amazon information with which to create profiles of the family's needs and concerns in the highly personal environment of their own home. By using artificial intelligence more intensively than previous assistants, Echo could give marketers access to an environment they had never been able to penetrate directly.

When Alexa was launched, Amazon was already applying the latest computer analyses to profile people on its website, on its advertising network of other sites, and on apps. It knew who its shoppers were, what they were like, and often what they were doing on the internet. The company acknowledged using profiles to understand the buying patterns of various population groups and to tailor the product choices and ads on its site, apps, and ad network to what it had learned about individual users. Yet I have never been able to find a public statement from Amazon about how it intended to use the new storehouse of information the Echo would provide: what individuals said to Alexa, and how and where they said it. Nor did the firm disclose how it would tie the knowledge it gained about individuals from the intelligent agent to the profiles it continued to assemble by other means. What does seem clear is that the company didn't want people worrying about the new flood of data that Alexa would send its way. It was no accident that the company worked to create strong personal bonds between humans and its humanoid, so that customers would happily allow Alexa onto devices not only around the home, but also in the car, in hotels, in stores—everywhere. Part of the seductive surveillance strategy was to position Alexa, with its soft female voice, as a helpmate rather than as an inquisitive salesperson.

In selling the friendly comfort of a female virtual voice assistant, Amazon was following the paths charted not only by Apple, Microsoft, and Google, but also by the contact center business, which handles customer-service inquiries for a wide range of companies, and for a range of purposes. At this point, contact centers were leading the way in using voice data acquired during customer-service calls to categorize and persuade callers with as little human labor as possible. By the late 2010s, the aims of the contact-center business and those of the intelligent-assistant business had begun to merge. Both were convinced that the sound of a person's voice had value in the marketplace. Both privileged computers over humans in drawing inferences about people's speech and voice patterns and in building "satisfying" relationships with customers. And both pushed technologies that could dig deeply into customers' private interests by combining more traditional marketing-related information like age, gender, income, race, lifestyle, and online behavior with data about what they were saying, analyzed in ways the customers would hardly notice or understand.

The technological breakthroughs that led to Alexa were a long time coming. The earliest step was the basic effort to replicate the human voice—which as it turns out is no easy feat. As early as 1773 a German-Danish scientist named Christian Kratzenstein created models of the human vocal tract that could produce vowel sounds.[4] But it took more than a century of additional attempts before Thomas Edison invented in 1877 what was to become the first marketable device to record and play back voices and other sounds.[5] The next ninety years involved a slow process of creating machines that could either synthesize or recognize spoken words, but not both. Only toward the end of the twentieth century did engineers begin to develop speech-synthesis systems that could interact flexibly with humans.

In the United States, the business of voice intelligence pushed forward with the support of both taxpayer and private money. The taxpayer funds came from the U.S. Defense Department's futuristic investment arm, the Defense Advanced Research Projects Agency (DARPA), apparently with the goal of developing the role of voice on the battlefield. In 1971 DARPA's Speech

Understanding Research program funded five years of university research toward creating a machine that could understand at least a thousand words. The greatest success was Harpy, from Carnegie Mellon University, a computer capable of understanding 1,101 words. While linguistic and engineering knowledge played a role, much of the increase in the number of words understood had to do with a growth in computer capabilities that would take off over the next decades. In 1976, as DARPA's first speech funding program ended, the best computer available to researchers might need a hundred minutes to decode just thirty seconds of speech. As computer processing speeds and memory grew, so did word understanding. By 1990, a typical commercial speech recognition system could handle more words than are in the average human vocabulary.[6]

Even more consequential during this period were theories developed by scholars at universities and in private firms about what it means to recognize speech and understand it, sometimes to the point of being able to pick out one person's particular vocabulary. Success came in halting steps. A scientist on IBM's speech recognition team during the 1980s recalled that their system, which required a roomful of computers, was "trained" to understand only what a particular individual said. If the computer made only one error for every ten words, that was a terrific result.[7] Over the following decades, investigators around the world, especially at IBM and Bell Laboratories, created artificial intelligence algorithms that improved computers' abilities to understand human speech. As the market research company Forrester notes, "the killer feature of AI algorithms is their ability to learn the underlying patterns in any phenomenon, regardless of complexity, given enough relevant data and computing power."[8]

The key processes involved in Alexa are speech recognition, speech processing, and speech creation or synthesis. Each step requires large and varied datasets of recorded and transcribed speech to train the system. In the training related to speech recognition and processing, engineers use complex statistical models under the rubric of machine learning (which nowadays involves powerful tools called deep learning and deep neural networks) to teach the computer how to link sounds to words and sentences so that it will transcribe them correctly, irrespective of accent. Once the words are transcribed properly, the goal is to use a set of statistical procedures called natural language processing to understand the meaning of the speech—what the person is trying to say. To do that, engineers again use large and varied training sets. The goal is for the computer to interpret the statement correctly and take the correct action. Although the sentence "Wake me up at 7 a.m. tomorrow" seems simple, an assistant would have to know that several variations on this request—for example, "Set an alarm for 7 a.m.," or "Please wake me at 7 a.m."—should yield the same result. A good training set allows the deep-learning algorithms to see that a large variety of such statements should yield the same output.

Then there is the matter of training the computer to respond, also by voice, which involves a series of extraordinarily complex steps. The process often works this way: first, the engineers find a professional voice talent whose sound meshes with the creator's aim, including the personality the creator wants to give the assistant in the language being used. That leads to recording sessions: ten to twenty hours of speech in a professional studio. The actor's scripts, according to Apple's Siri team, "vary from audio books to navigation instructions, and from prompted answers to witty jokes."[9] The engineers then run the words spoken during those sessions through a computer that slices them into their elementary components, their snippets of sound. The computer will use a database of these speech snippets when it needs words that sound certain ways. The final step, recombining the snippets into sounds to match the sentences in a text, is the hardest part. The current approach is to use deep-learning methods on the training set (which links audio and transcribed sentences) so that the training set will teach the text-to-speech computer how the words in a text ought to sound. This means using acoustic models to give the computer the probabilities and other data resources it needs to decide how to choose and link snippets to convey not just the words, but also a wide range of emotions through voice tonalities, rhythms, and cadences.

This is only a very basic sketch of the astonishingly complex set of decisions that a computer assistant makes in responding to an apparently simple command or question. Sometimes the assistant may take a shortcut by focusing on specific keywords in a sentence—for example, *what* and *time* in "What time is it?" At the same time, engineers are working hard, and with increased success, to understand the multiple kinds of context surrounding what people say. The most basic is contextual understanding: when a person says "set an alarm," the assistant responds, "what time do you want?" Other contexts might involve understanding the person's remarks differently depending on geographical location, room in the home, time of day, or even, for a smart watch, the person's pulse rate.

Although it was clear early on that personalization using voice intelligence could be enormously valuable, the progress of artificial intelligence in this area was by no means smooth. Vlad Sejnoha, a computer scientist who worked for Nuance, told me the company grew by making strategic investments during what people in the voice analytics business call the "speech technology winter" of the 1990s and early 2000s. "There had been a number of notable failures in the late 90s," he recalled.

> Companies overreached; the technology was really not up to what they were trying to accomplish. . . . The computation wasn't right there, the connectivities weren't quite there in the 90s. PCs weren't really all that powerful. And so a lot of the applications that were available were clunky and certainly underperforming especially compared to today's standards, where in many cases cloud-based speech recognition just works. It's reliable and accurate for the great majority of the population. That was not the case [back then]. You had to laboriously train several recognition systems. For example, if [a customer] bought Dragon Dictate in the 90s, you [had] to spend a couple of hours training it, and it was an expensive product. So these companies ran into trouble. . . . And a lot of large companies, including Google, Amazon, and Microsoft and Apple minimized their investments, if they had any.[10]

Sejnoha recalled that Nuance's CEO at the time, Paul Ritchie, "had a lot of foresight and used that time to accumulate a lot of speech technology assets, and made a lot of acquisitions early on. He was investing for the time that he and the rest of us believed would come again, and indeed it did. And I think there was a time Nuance stole a march on a lot of these giants, and some of the early products I think caught Google and Microsoft and others by surprise." In the mid-2000s, "they quickly started investing again, and it's well known—it's a matter of public record—that in many cases they [did] that using licenses from Nuance." By the late 2000s, computer speech recognition and appropriate responses had advanced enough that a Microsoft executive used it to schedule his appointments, and a Microsoft lab was trying out a "medical avatar" that could ask children questions about their symptoms and make diagnoses based on their answers.[11]

While these trials were taking place, marketers were beginning to apply this growing area of artificial intelligence to a crucial but controversial part of their business, the customer contact center. At the start of the twenty-first century, contact centers arguably had access to more information on Americans than any other marketing endeavor, but they struggled to use the data efficiently to personalize interactions. The basic problem was an old one: the call center, as it was originally called, was about a hundred years old. Big department stores had created the first ones, which were simply large switchboards. Wanamaker's department store in Philadelphia established the first store telephone system around 1900, twenty-four years after Alexander Graham Bell first exhibited his invention. By 1915 the store had the largest private branch telephone exchange in the world, with more than two thousand operators who handled over 1.8 million messages.[12] American Telephone and Telegraph (AT&T), the phone company, had an operator pool that made millions of verbal contacts with customers each day. But those two firms were giants of their day; many marketers were unwilling to invest in the kind of response infrastructure that Wanamaker and AT&T created. Gradually, a call center industry evolved, consisting of companies that handled phone calls for multiple clients. People in the industry remember its early years as filled with human error, unreliable technology, and slow service. As one history of the business notes, "back in the day, holding for 10, 15, or even 30 minutes wasn't unheard of."[13] The goal was just to keep up with the flow. Harried phone agents inevitably made judgments about callers based on how they spoke, but their conclusions weren't recorded; they just wanted to complete the call.

Little changed in how the centers dealt with customers until the 1960s. That was the threshold for several decades of developments that both sped up call handling and gave the call industry far more information about customers than any other media business could obtain. Ironically, all the developments started a long-term movement by call centers away from the human salesperson's intuition about voice, toward judgments based on hardware and software. In the 1960s, AT&T introduced toll-free 800 numbers and began to replace rotary dials with touchtone calling. The 1970s brought automatic call distribution (ACD) systems and interactive voice response (IVR). Toll-free numbers were a revolutionary marketing innovation at a time when long distance phone calls could be expensive. Accompanying advertised products in print media and on television, the numbers allowed people to buy things over the phone with their credit cards (or cash on delivery) and have them mailed to their homes. Automatic call distribution replaced manual switchboards with computer-guided ones that could allocate the new torrent of toll-free calls to operators far more efficiently than would have been possible in previous decades. Further increasing routing efficiency, the interactive voice-response systems played digitized speech messages to callers before they reached a live person and instructed them to push one or another touch-tone button to indicate the purpose of their call. That way the automatic call distribution would not only route the calls to a waiting representative; it would also put callers in touch with a representative who had the skills the caller wanted and who knew the basic reason for the call.[14] It was the start of automated personalization.

During the 1980s and 1990s, call centers improved their automated understanding of callers by purchasing computer databases to store information about individual customers that could supplement what those customers told agents over the phone. These databases allowed organizations to maintain lists of customers' characteristics—from names and addresses, to history with the firm, to scores describing their value to the firm—that no human beings could possibly manage. A rush to use these tools led to a new term, "customer relationship management" (CRM).

The umbrella description for these developments, computer-telephony integration (CTI), describes the goal: to enable computer and telephone systems to interact. As the sophistication of databases increased in the 1990s and beyond, CTI supplied telephone representatives with information about customers that they had not previously had access to. Right from the start of the call, the agents could authenticate callers by comparing their phone numbers with the ones listed in the company's database. Screen popups and other tools gave the agents a dashboard profile of the customer and sometimes allowed the agents to include in their conversations an acknowledgment of a caller's history with and importance to the firm.

The rise of the commercial internet in the 1990s added to the torrent of personal data. Primarily to save money, call centers traded their traditional wireline methods for the internet's packet switching mechanism for phoning (a technique called voice-over-internet protocol, or VoIP). That allowed them to connect their widely separated call centers much more cheaply than in the past. But linking to the internet held another benefit: it allowed centers to capture not just what people said over the phone about a center's corporate client, but also what they looked at when they went to the client's website; what they wrote in emails, text messages, and chats to the firm; what they posted on the firm's Facebook page; and, by the 2010s, what they bought in the firm's online stores or on its mobile app.[15] As the twentieth century turned into the twenty-first, practitioners called this tracking an "omnichannel" approach that captured the "customer journey," and industry executives began saying that they were in the contact center, rather than call center, business. An executive involved in implementing these activities said in 2012 that "one of the greatest benefits is that now, because of VoIP, contact centers are able to more easily capture 100 percent of their interactions. This massive corpus of customer conversations is a very rich source for analytics."[16] Yet it raised a dilemma: having so much information about individual callers was great, but how could a human agent absorb it all during a phone interaction?

The question concerned more than the future of human versus technical resources. It held enormous implications for the future of those who would construct profiles of customers—humans versus AI-driven computers. Business pressures pointed to using artificial intelligence as much as possible: labor costs were rising, and the questions that callers were asking agents were growing more and more difficult to answer. A 2008 Contact Center Satisfaction Index report by the service-ranking company CFI Group confirmed that customers increasingly used calling firms "as the resource of last resort," turning to them only after they had failed to answer their own questions digitally. One consequence, according to the report, was that "in today's multichannel environment, customer service representatives are more likely to get a higher proportion of 'harder' questions that customers cannot find answers to on a Web site or elsewhere."[17] In that environment, according to CFI, one in five customers reported they could not resolve their problems with the contact center reps. That was an ominous sign, because CFI saw satisfaction with the contact center as an important indicator of loyalty and customer recommendations. The firm found that 94 percent of satisfied customers said they would do business with the same company again, and 91 percent would recommend it. Among dissatisfied customers, only 62 percent said they would remain customers, and only 39 percent would recommend the firm. "Customer service representatives are on the front lines of a company's interaction with their customers, so it's vitally important that they have the training and resources to do what customers expect of them," said CFI Group's CEO. "If customers just wanted to hear a friendly voice, they'd call their mom—but they are calling to get something done."[18]

Call industry executives, meanwhile, did not share the notion that contact center employees could be sophisticated and efficient handlers of torrents of difficult calls while also taking the customer journey, background, and relationship to the firm into account. The executives' more immediate concern was costs. Marketers, seeing the need for 800 numbers as well as opportunities for data capture, caused the call center industry to balloon in size, technical complexity, and competitiveness. Between 1988 and 1998, the number of U.S. companies involved in inbound or outbound operations (and often both) tripled to about 2,500. The difficulty of cultivating human talent at the wages the centers were willing to pay in such a fiercely competitive environment led them to adopt a strategy very different from the one advocated by CFI: paying agents as little as possible while ramping up the personalized information that technology could present to agents to satisfy callers. It became clear that while call-center leaders often hyped the rollout of ever more sophisticated customer management systems as efforts to know more about the customer, they took this step as part of a furious drive to lower the costs of speaking to the deluge of customers. As one executive noted, during the late 1980s and early 1990s, handling calls could cost a center more than twenty cents per minute—and with thousands of toll-free calls, that could add up. Live agents, especially U.S.-based agents, became a pain in the wallet. Consequently, "shortening the time on the phone by pre-populating data fields [with personal information about the caller and the caller's relationship with the firm] had a rapid ROI [return on investment]."[19]

Still, for many large marketers, U.S. call centers weren't bringing down costs enough. They began to use call centers in countries where wages were far lower than in the United States, a move made possible by the new internet phone systems.

Contact-center firms created internet-driven private branch exchange (iPBX) systems that moved incoming calls onto their VoIP corporate networks, converted the calls to compressed data, and routed them across the internet to wherever the agents were located—near or far, the cost was about the same.[20] Labor costs in countries like India and the Philippines could be as low as $1 per hour, compared with $6 to $10 an hour in the United States.

According to the U.S. Bureau of Labor Statistics, between 2006 and 2014, the United States lost more than 200,000 contact center positions.[21] Firms continued to push their costs down while installing technologies to quietly understand callers' backgrounds, respond to their demands, and guide discussions toward a conclusion that would make them happy customers. As one website for a firm selling CTI suggested in 2019 to harried executives, "Your team is handling more calls than they can manage. The phones won't stop ringing, and customers aren't being helped quickly enough. Stress builds for employees, which consequently gets felt by the customer. CTI can change that."[22] But supervisors were not about to let workers relax once the technology had helped to allocate the calls and give them information about callers. For while it enabled the caller to be more of an open book for the agent, the technology also made the agent an open book for supervisors. One website description of CTI's call monitoring and recording functions said they would "give management insight into how employees are performing and how customers are being helped. The monitoring function enables managers or coaches to listen in on the call and help guide the agent."[23]

The tensions around workload and offshoring spilled out into labor battles in the United States. In 2012, describing union organizing attempts at a call center in Asheville, North Carolina, a writer for the online Daily Kos called these centers “the sweatshops of the modern era.”[24] A 2014 piece in Gizmodo struck a similar note: “The call center system as a whole is broken. And as you’ll see from the tales below, it’s breaking its employees along with it. . . . We’ve compiled some of the more appalling stories [sent by readers]; the recurring themes of debilitating stress, impossible standards, and wildly high turnover [rates] are too prevalent to ignore.”[25] These domestic problems notwithstanding, the Communications Workers of America released a report in 2009 arguing that “the off-shoring of call center jobs is . . . bad for American workers and communities and harmful to the security of U.S. consumers’ sensitive information.” The report highlighted “a range of fraudulent and criminal activity emanating from overseas call centers,” especially India, the Philippines, and Mexico, that included credit card theft, identity theft, and the illegal sale of customer data. Concluding that “U.S. companies have been exporting call center jobs by the thousands in a global race to the bottom,” the report advocated passing “the bipartisan United States Call Center Worker and Consumer Protection Act.” Sponsored by a Democratic congressperson from Texas and a Republican congressperson from West Virginia, this bill would have “required that U.S. callers be told the location of the call center to which they are speaking,” that call centers offer callers “the opportunity to be connected to a U.S. based center,” and that the U.S. Secretary of Labor create a public list of “bad actor” companies that offshore their call center jobs from the United States and make them “ineligible for certain grants and taxpayer-funded loans.” The bill never made it out of any of the four committees that the House Speaker asked to consider it.[26]

The takeaway message of this dispute for the industry was that human labor, wherever it was located, would create trouble for contact centers and the companies they serve. Consequently, while consultants kept repeating the decade's mantra of cultivating customer satisfaction through omnichannel relationships—which often resulted in phone calls—the emphasis increasingly moved away from the ability of the human agent to the utility of the technology. In particular, industry practitioners increasingly relied on computers to create the personalized understanding and messaging that had historically been the task of people on the company end of the phone. When a trade-site editor asked an executive for the call center firm InfoCision to speak about the future, he mentioned not his call agents but technology. "At the heart of any CRM strategy," he said, "is the telephone channel, which provides a higher level of personalized communication. . . . The Internet gives customers more options to contact you—e-mail, chat, social media, which have given way to increasingly higher expectations when it comes to customer service." That, he continued, "coupled with the struggling economy, has really pushed companies to new levels of efficiency—looking for new ways to produce ROI."[27]

It was no accident, then, that as early as 2005, AT&T created a voice assistant to help Panasonic field torrents of calls from customers about products they had bought. ("We were drowning in calls," recalled Panasonic's vice president of customer service.) The AT&T system identified key words among a caller's phrases and sentences and produced a reply in a female voice. It worked with simple problems that could be recognized through key words. When the system couldn't distinguish the words, it routed the call to a live representative. Basic as this was, Panasonic claimed the voice assistant lowered the average cost of resolving a customer issue by 50 percent. Success inspired imitation. US Airways, for example, introduced a phone assistant (this time with a male voice) explicitly to save money on human agents.[28] Both companies proudly noted that customers hardly seemed aware of the computer's presence, saying "thank you" as if they had talked with a real human.[29] In a further effort to gin up phone reps' productivity, contact centers began to use AI to discern the caller's mood. "Certain emotions are now routinely detected at many call centers, by recognizing specific words or phrases, or by detecting other attributes in conversations," wrote two *New York Times* technology reporters in 2010. They added that Voicesense, an Israeli producer of speech analysis software, had developed algorithms that it claimed could measure a dozen indicators, including breathing, conversation pace, and tone, to alert agents and supervisors when callers "have become upset or volatile."[30]

Many in the direct marketing business were coming to believe that humans—callers—need relationships, but not necessarily with living people. New developments in voice creation, voice recognition, and machine learning could lower labor costs while taking to new heights the ability to profile individuals and personalize messages for them. Nobody among contact industry leaders dared suggest that they would take human agents out of the equation. But Amazon, Google, Apple, and Microsoft were betting it could be done. First motivated by a desire for competitive advantage, then by a desire for customer surveillance, they would stress personality and personalization, knowing that these seductive features were the most likely to keep people engaged.

Their work would in turn accelerate contact centers' development and use of humanoid assistants.

Despite their importance to marketers, most of the early activities around voice in contact centers stayed below the public radar; the centers didn't inform callers about surveillance or what they were learning from it. Apple's Siri was the first assistant to interact openly with the public, and it created enormous enthusiasm for the potential usefulness of artificial intelligence in everyday life. Focused initially on speech recognition rather than biometric identification or inferences, this omnipresent character's benign affect eased the public into a marketing world where speech recognition and profiling for personalization would merge.

Siri did not start under a marketing umbrella. It was born out of taxpayer money in 2003, when DARPA funded the non-profit research institute SRI International to build a virtual assistant. Voice-activated controls and speech recognition features with various levels of ability had existed in home computers and other equipment starting in the 1990s, and DARPA hoped a more sophisticated interactional software would help military commanders deal with information overload. Called the Cognitive Assistant that Learns and Organizes (CALO), the project and its $150 million in government backing attracted hundreds of artificial intelligence experts. When they did develop intelligent assistant software, the successful result encouraged a number of business-minded engineers in 2007 to leave SRI, license key software from the CALO project, and develop their invention for the new iPhone. (A 1980 law made all that legal.) Reasoning that it would be a lot easier to use the Apple device through voice commands than by typing, they created an iPhone app called Siri (after the SRI mother ship) that was ready to go in February 2010. Steve Jobs had noticed; he may have seen Siri as a valuable rival to the Voice Search app that Google had recently introduced for the iPhone. Within weeks after Google's deployment of its app, Apple bought the Siri engineers' company, and over the next year it adapted Siri to its needs by reducing some capabilities (for example, its use of many outside web services to get information) and adding others (for example, several more languages). Apple also seems to have brought in Nuance to help with the backend technology for speech recognition.[31] When it released the iPhone 4S in October 2011, Siri was built in.

Although the CALO team members griped about the new owner's changes, Siri electrified the technology world. Google had announced the Voice Search app for its Chrome browser that June, and observers had recognized it as a breakthrough in voice recognition accuracy. Google had figured out how to recognize a person's voice request, transcribe it, and return relevant websites as if the person had typed the request. Yet as remarkable as that achievement was, Siri went well beyond it. Here was an entity on your phone you could ask to tell you facts or post a calendar appointment, and it would cheerfully do both. Articles commended it for its unique speech, crisp answers, and ability to joke, though the consensus was that the assistant wasn't as accurate as it should be (the Piper Jaffray investment bank and securities firm gave it a grade of "D" on understanding and answering queries).[32] Most observers gave Google's Assistant better marks for understanding.

Apple's success with the iPhone and Siri led, in a circuitous way, to Amazon's release of Echo and Alexa in 2014. It all started, ironically, with the enormous failure of Amazon's Fire Phone, a debacle that led the company to quickly pull the plug and announce a $170 million accounting loss. It was easy to understand why Amazon CEO Jeff Bezos would want to release a phone. In an increasingly mobile world, shoppers would buy more and more things on the move, with mobile devices. Amazon wasn't selling electronic (Kindle) books through other phones' app stores because doing so could mean having to give the hardware owner a cut of sales—in Apple's case, 30 percent. An Amazon phone would avoid those charges.[33] Perhaps more important, an Amazon phone would give the company real-time data about its customers and the ability to personalize its responses: Amazon would be able to track phone owners' locations, send them product recommendations based on that data, and use their whereabouts to build up their profiles. The challenge was to get people interested in such a product. Bezos thought the phone should include a number of unique abilities, such as a sophisticated display that looked like 3D on which the user could start apps by tilting the phone in different directions. As it turned out, those gizmos made the Fire Phone as expensive as an iPhone. Reviews were mixed, sales were terrible (analysts estimated only a few tens of thousands), and the company discontinued it in August 2015, barely a year after its debut.

But there was a silver lining to Amazon's failed experiment: the Fire Phone's development had involved work on a voice assistant. An executive in charge of the phone, Ian Freed, showed Bezos an early version of its software, which was able to recognize the utterance of any popular song title and then play it. Bezos was intrigued, and a few days later he asked Freed "to help build a cloud-based computer that responded to voice commands, like the one in *Star Trek*."[34] He gave the team a $50 million budget to hire speech scientists and artificial intelligence experts to create software that could recognize and respond to a far greater range of speech than song titles. Only four months after the calamitous Fire Phone release, the Echo, with a kernel from the ill-fated phone, made its debut. The team chose the name Alexa for the accompanying voice assistant out of a belief that while pleasant, it is unusual enough that users wouldn't often say it accidentally. The initial $100 price for Amazon's Prime members reflected the main takeaway from the Fire Phone flop: Amazon Senior Vice President of Devices David Limp believed that his division had priced the phone too high. Echo would be priced low, to draw larger audiences for its voice assistant.[35] An unstated consequence was that the profits from the device would come from other sources, including its surveillance activities—that is, from the company's savvy capitalization of Alexa's interactions with Echo owners.

As Amazon's smart speaker became a hit in late 2014 and early 2015, Google executives raced to match it. According to several accounts, Google strategists were not surprised that a virtual assistant would gain traction with the public, but they had been sure it would happen on smartphones and tablets.[36] Amazon had pivoted to the home only because it had failed with its phone. Google, rushing to catch up, released its Home smart speaker in the United States at a competitive price almost exactly two years after the Echo's debut, then pushed it out in more countries and languages than Amazon had been able to do. Apple's response was far slower. It began taking orders on its HomePod speaker with Siri in January 2018, a little more than three years after Amazon started selling the Echo. Clearly reaching for the high end of the market (much like the iPhone), the HomePod emphasized stellar sound at a price more than a hundred dollars higher than the original Echo.

Microsoft had joined the personal assistant fray in 2014 with Cortana, which it aimed to include in a future Windows operating system along with a Microsoft phone. And in 2017 Samsung introduced Bixby: linked mainly to smart TVs and appliances, it was the least used of the five in the United States. The common speculation of the marketers I interviewed was that each device had a different business model. Google and Apple had the largest numbers of people using their voice assistants—in the billions worldwide—because of the widespread use of the Android and iOS operating systems on phones, tablets, and computers. Google, with its legacy of selling marketers the ability to reach people through internet advertising, saw voice as the new way to search the internet—and a new way to track users doing it. As one analyst wrote, "if even a fraction of searches shift from mobile and desktop to voice interfaces, then that is where Google services need to be."[37] Apple, not committed to advertising as a moneymaker, positioned itself as a privacy-aware firm whose interconnected devices would work seamlessly. After a couple of years, Microsoft decided to drop out of the general voice-agent competition. Instead, leveraging its strength in business software and computing, it would position Cortana as primarily an assistant that people could use to plan their business day (via calendars, contacts, and email) and to help with business calls. Amazon's strategy was about selling products, its own and others, through Alexa. Some observers believed that Alexa's compatibility with Amazon's music, video, and audio-reading services was designed to encourage people to join Amazon's Prime buying program, which would lead to increased overall purchasing from the company. Others, not disagreeing, saw Amazon's goal more broadly. It was, in the words of one analyst, to "take a cut of all economic activity."[38] They saw sales via Alexa as another example of that.

Amazon and Google had the most interest in exploiting their assistants' surveillance features to sell things to users, whereas Microsoft and Apple wanted to know a lot about their customers in order to personalize their services to them. To accomplish either goal, each company needed to shape its assistants to keep current users interested while also attracting new customers. "Wired for speech" devices had to ingratiate themselves deeply with their users, and a key strategy was to imbue their voice assistants with personality. Strong humanoid-to-human connections that encouraged friendship and trust would mean fewer questions about the data their assistants were taking and using behind the scenes. Although the ingredients differed, each company followed the same basic recipe in concocting its voice character: First, imbue the voice assistant with a personality with which people want to engage. Second, give the assistant the ability to manage data about every user in ways that help those users get things done as successfully and seamlessly as possible. Third, place these assistants in devices that not only lure the user with what the industry calls "frictionless" benefits, but also allow the company to harvest voice and other data from that user across as many venues as it can.

The personality part of the recipe had precedent. Back in 1995, Clifford Nass and his colleagues at Stanford gave the creators of computer personalities a blueprint. They found that individuals could readily recognize personality types. Further, they said, the generation of a personality that moves people "does not require richly defined agents [or] sophisticated pictorial representations. . . . Rather, even the most superficial manipulations are sufficient to exhibit personality, with powerful effects."[39] The creators of the voice assistants, having intuited this from the start, based the robots' traits on bits and pieces of popular culture. Jeff Bezos's *Star Trek* reference in his instructions about what ended up as Alexa wasn't at all unusual. Read about the genesis of any intelligent assistant and you're likely to come across references to science-fiction and video-game characters. Martin Cooper of Motorola, who invented the cell phone, said the design was inspired by Captain Kirk's flip-top communicator on the original *Star Trek* TV series.[40] The fictional computer on the *Star Trek* ship, the USS *Enterprise,* also used a female voice to respond to crew members' requests. In the show, the voice belonged to the actress Majel Barrett, the wife of *Star Trek*'s creator, Gene Roddenberry, and while Google's engineers were developing their Assistant, they named it Majel.[41]

Cortana, Microsoft's phone assistant, is named after a twenty-sixth-century artificially intelligent character—"ever faithful companion of the Master Chief"—in the *Halo* video game series.[42] In fact, the actress who voiced Cortana in the video game, Jen Taylor, also contributed her voice for the U.S. version of the assistant. Higher-ups at Microsoft considered a different moniker for the public version, but a petition on a Windows phone user site evidently persuaded them to keep the *Halo* name.[43]

The 2011 version of Apple's Siri occasionally quoted Hal, the sentient and ultimately malicious computer in *2001: A Space Odyssey* (whom the American Film Institute named the thirteenth greatest villain in the history of movies). When Siri didn't know an answer, it would repeat Hal's well-known "I'm sorry, Dave, I'm afraid I can't do that." If a Siri user referenced a famous scene in *2001* by saying "Open the pod bay doors," the agent would reply, "We intelligent agents will never live that down, apparently." And in 2017, more than one observer saw Hal in the bright red circle that appeared and then swirled at the top of the new HomePod. "The glowing orb responds, when you're talking to it, just like HAL 9000," commented a Gizmodo writer.[44]

Marketers quickly decided that such inside jokes (dubbed "Easter eggs") should be turned into selling points. In 2016, Google enlisted Ryan Germick of its Doodle section (the group that creates the cartoons above the search box), along with Emma Coats, an animator who had worked for Disney's Pixar Studios, to give its newish Google Assistant "a little more personality," not only on smartphones, but also on the just-released Google Home speakers.[45] Google's CEO, Sundar Pichai, had said the Assistant was meant to be an "ambient experience that extends across devices"—for example, handing off between the phone and the speaker.[46] Immediately after she was hired, Coats set about giving Assistant a dramatic, tumultuous backstory so that users would empathize with it. Yet in 2017, with some experience behind her, she scaled down her ambitions to match the understanding of Nass and his colleagues. She described to *Wired* magazine how Google Assistant's "easygoing, friendly" personality was constructed by thinking up questions that humans are likely to ask Google and deciding on several responses for Assistant to use. Humor, she said, is good for both building the character's personality as "the fun, trusty sidekick" and for taking people's minds off mistakes or misunderstandings that might call attention to the character's non-human status.

"We don't want to have to fall back on something like, 'I don't understand,'" Coats explained. "That draws the attention back to you instead of continuing the conversation you're building." She also posited limits to the personality her team could give its creation. "Assistant can't be opinionated: it's there to be reliable, not to have depth." Nor could they write any script suggesting that the character is a tortured soul; "If we gave it some dark conflict secret, that probably wouldn't be a great user experience."[47]

Google wasn't alone in this project of creating a character that people could relate to from a modest script. Amazon writers tell of adding dozens of "delighters" to Alexa—including giving her groan-worthy dad jokes and concocting Easter eggs—typically inside jokes in response to certain questions or statements. Sometimes Alexa channeled old movies, like the comedy *Airplane*. (Human: "Alexa, surely you can't be serious." Alexa: "I am serious, and don't call me Shirley.") Heather Zorn, the Alexa team's director of customer experience and engagement, said the goal is to make the AI both useful and fun. She added that her team built their assistants' comments around several Alexa personality traits, trying to make her smart, approachable, humble, enthusiastic, helpful, and friendly. A Cornell University study that analyzed 587 customer reviews of the Amazon Echo showed that reviewers who referred to the device as "Alexa" and used the pronoun "her" were more satisfied than those who spoke of "Echo" and "it." Even so, Zorn asserted, the company doesn't want to turn Alexa into a member of the family. Instead, perhaps unwittingly echoing the Google engineers who created Majel, she said that her team's "guiding light" and original idea for the persona was the all-knowing ship's computer on *Star Trek*.[48]

Creating a persona perceived as a friendly and credible personality is the key to seductive bonding with a customer—to helping that customer feel psychologically tied to the device, and thus to the company.[49] Think first about the agent's voice. It became contentious in the United States that the default Siri, Google Assistant, and Amazon voices were all female. Critics argued that making a woman the default version of a polite, deferential, and pleasant assistant reinforced generations of harmful stereotypes of women in subordinate roles.[50] Samsung's Bixby stoked their anger as well, not because its assistant was female (the company gave people a choice of two genders), but because of what critics saw as "loaded, sexist" characterizations: in its language settings, Samsung described its female voice as "chipper" and "cheerful," and its male voice as "confident" and "assertive." (As a Twitter anger storm grew, Samsung quickly removed these labels.)[51] Yet neither Microsoft nor Amazon disputed that they were reflecting social stereotypes; both firms stated only that research with real people had led them to the gender they chose. In the words of a Microsoft executive, "For our objectives—building a helpful, supportive, trustworthy assistant—a female voice was the stronger choice."[52] Apple and Google seemed to feel that way too, at least for their American users, but they did roll out male voices for those who didn't want the default. In 2019 Google offered people ten choices—six female, four male—plus the "celebrity" pick of singer John Legend for some responses. Toward the end of that year, Amazon also moved a bit on gender, offering actor Samuel L. Jackson's humorously irascible male voice for a fee to replace Alexa on some activities. Unlike Google, though, it kept the female voice as the brand's enduring, easygoing persona.

To voice-first executives, ensuring that most users aren't put off was part of an essential seductive surveillance aim: making users feel comfortable enough to interact with the assistants and give them voice data and other information for profiling and personalized communication. In ads and instructions, the voice companies describe the ways an assistant can satisfy personal needs directly if owners allow them access to their voices and their lives: if given this access, an assistant can reliably post a calendar event, set a timer, answer a question via the web or Wikipedia, and much more, through the ease of speech. A 2019 Google Home video called "Hands-free help from Google Assistant" shows a kitchen on what appears to be a busy weekday morning. Two young adults are milling around, and a school-age child is eating; it's unclear how they are related. "The Google Assistant can distinguish your voice from others," begins the narration, "so when you ask for information on Google Home you'll get a response just for you. Get personalized briefings on your schedule, commute, weather, and more. So you're both ready to take on the day. . . . Call your personal contacts hands free." The two adults speak and the Assistant answers.

Young Adult One: Hey Google, tell me about my day.

Assistant: Good Morning, Alex. It will take you forty-five minutes to get to work.

Young Adult Two: Hey Google, tell me about my day.

Assistant: Good morning, Ross. Your first meeting is at 10 a.m.

The narration returns: "You can both request your personalized playlists using just your voice. . . . Open the Google Home

App to train the Google Assistant to recognize up to six voices. Once you're set up, everybody can start enjoying more personalized responses at home."[53]

The video reflects the view of Google CEO Sundar Pichai that the essence of the intelligent agent is the ability to personalize around every aspect of what a customer does and says. The responses that users hear from Google might seem to indicate the extent of what the assistants know about you and your life, but that impression is wrong. Unless you go out of your way to find ways to limit or delete specific types of information, the company reserves the right to "collect . . . voice and audio information when you use audio features," along with an enormous amount of other things it learns about the people who get their customized output.[54] This, to Pichai, is not only unproblematic, but also just the beginning. "Today we have an understanding of one billion entities: people, places, and things, and the relationships of them in the real world," he said at a Google Developer conference in 2016. "We can do things which we never thought we could do before. . . . We think of this as building each user their own individual Google."[55]

The CEOs of Amazon, Apple, and Microsoft could show similar videos from their firms, and they likely would agree with Pichai's conclusion. But in 2019, Google tried to show that it was ahead of the pack when it rolled out its Duplex technology. Available on iPhones and newer Android devices, it gave Assistant, with a male or female voice, the power to make reservations for the user. Observers noted that this iteration of Assistant sounded eerily human, even including halting sounds such as "uh" and "umm" (sounds called speech disfluency) in order to more perfectly mimic a person.[56] This naturalness caused concern during Duplex's initial public demonstration because Google didn't identify its caller as a robot; after the criticism, it began the calls with "This automated call will be recorded."

A person with a phone can simply activate the Assistant app and ask it to book a restaurant table, schedule a haircut, or check a business's operating hours using a choice of several male or female voices. When the reservation is made (within the next fifteen minutes), the Assistant sends the user a text message. If a call doesn't go through properly, or if the person on the other end doesn't want to be recorded, a human representing Google will take over. One reason for both the human involvement and the limited types of reservations possible so far is the difficulty of generating the huge training sets needed for Google's machine learning operations to figure out the best ways to discuss appointments. According to Google, the human operators who intervene are also there to write explanations of the glitches on the call transcripts used to train Duplex's algorithms.

Google positioned Duplex at the leading edge of personalization, a leap beyond customized phone voices and the logical next step in more personalized interactions. The voice was by now so personalized and friendly that Google was sure people would let it phone for them. It was all so seamless, at least in the demonstration, that questions about the surveillance that came with the seductive features—the information about their customers that Google and Amazon and others were taking and storing—didn't come up. The only indignation was about the impoliteness of a robot not saying who it was and that it was recording the conversation. A writer who had seen a demonstration made the obvious leap to Duplex's value for call centers. "Many big businesses are basically trying to make a human version of a robot," he wrote, by training them to "rigidly follow a script." The need for humans to mimic robotic assistants would virtually disappear if a large company got hold of "Google Duplex-style technology for its call center."[57]

That kind of all-purpose ability to converse with callers (or smart-speaker users) is not likely to happen for a while. As the Forrester consulting firm wrote, "The human brain has had millions of years to develop the architectural complexity required to comprehend and generate language."[58] Humans have linguistic abilities that computers still can't master. The big difference has to do with the uncertain meaning of many sentences and words outside of their larger context. A favorite example is the phrase "eats shoots and leaves." By inserting or removing a comma, heard vocally as a slight pause, you can make it about either a hunter or an animal. This sort of fuzziness is very much a part of people's everyday speech, and they can typically understand such phrases from the surrounding conversation. Computers, however, have a tough time making sense of these ambiguities, and this can lead to the frustrations that people sometimes feel with call center computers as well as with Alexa's generation of assistants.

Executives and engineers have taken aim at solving at least part of this problem, if only because they believe that training computers will in the long run be far less expensive than training humans. Nuance has for years been working with contact centers to use deep learning and other forms of artificial intelligence to improve computer interactions with humans. As early as 2013, Nuance's chief technology officer acknowledged that "customer service transactions have proven difficult to organize into menu structures in a way that's efficient and understandable." In the future, he said, "specialized virtual assistants will provide direct access to information by bypassing the IVR [interactive voice response] entirely, and also will support flexible conversations that allow users to proactively provide unprompted information, and to jump freely among different contact center functions."[59] More recently, IBM, Amazon, Google, and Microsoft have entered the contact center world in competition with Nuance and others. Their pitch is that the well-known voice recognition and synthesis technologies they have developed in other business areas will work for contact centers as well.[60] IBM and Nuance in 2019 were the most aggressive in claiming that their AI technology would do away with the clunky IVR push-button choices. According to IBM, its Customer Care Virtual Agent, with the male voice of Watson, would allow customers to seamlessly speak to it "using the same conversational speech they would use with a human agent." Customers would no longer have to push buttons "in response to robotic-sounding prompts" or hope they had the right keywords to get the response they needed. Instead, they "can use complete sentences to interact with Watson, which can understand those sentences and select the appropriate, natural-sounding responses."[61] Nuance claimed that its assistant, Nina, will interact with callers as "a familiar voice [that] will answer their request whether it's typed into a computer, tapped on a screen or spoken into a device."[62]

Both have personalities. Watson is a bit aloof, but Nina can be funny. I asked it "Are you married?" and it replied, "Is that you, Mom? Can we just focus on my job for now?" Their creators claim that each can personalize its conversation with an individual based on its ability to connect to the company's computers to learn the caller's background and previous interactions with the firm. And AI training is said to be easy. IBM contends that customization of "a basic conversation" can take as little as one to two days, and "within three to six months you should have Customer Care Voice Agent integrated into your call center system."

These claims make it sound as if marketers' future is here. Stepping away from its hype, though, IBM carefully acknowledges that its agent succeeds when people have routine concerns; those are the basis of the virtual agent's training set. Both IBM and Nuance emphasize that when people want to speak to a real person, their virtual agents "can quickly and easily transfer less routine cases to human agents."[63] The current state of the art is to give the human agent as much AI guidance as possible to efficiently deal with the caller's needs in ways that benefit the company. In a sense the goal is to create, with a combination of the human agent and an AI sidekick, an attractive persona that can be as tailored for that person in that situation as Alexa or Siri would be. Surveillance is central to accomplishing this goal. The firms carry out deep-learning analyses of the growing number of channels that individuals use as they consider purchases or look for solutions to difficulties they have with the products. As Forrester Research notes, tracing this customer journey means "combining quantitative and qualitative data to analyze customer behaviors and motivations across touchpoints and over time."[64]

Increasingly, too, analyzing the journey means exploring the customer's speech and voice patterns. Some in the voice intelligence industry are spreading the belief that even the most useful conclusions from a person's background and activities can be surpassed by deeper analytics that connect those characteristics to individual words and word patterns, and even the physical characteristics of people's voices. The goal clearly is to use seductive surveillance to help create an extreme version of personalization: to know a person better than they know themselves. And marketers are trying to access a torrent of speech, voice, and other new data to make this goal a reality.