



← JAN NEUMANN

**ABOUT JAN NEUMANN**

Jan Neumann leads the Comcast Applied Artificial Intelligence Research group with team members in Washington, DC, Philadelphia, Chicago, Denver and Silicon Valley. His team combines large-scale machine learning, deep learning, NLP and computer vision to develop novel algorithms and product concepts that improve the experience of Comcast's customers such as the X1 voice remote and personalization features, virtual assistants and predictive intelligence for customer service, as well as smart video and sensor analytics.

Before he joined Comcast in 2009, he worked for Siemens Corporate Research on various computer vision-related projects such as driver assistance systems and video surveillance. He has published over 20 papers in scientific conferences and journals and is a frequent speaker on machine learning and data science. He holds a Ph.D. in Computer Science from the University of Maryland, College Park.

**ABOUT COMCAST**

Comcast Corporation (Nasdaq: CMCSA) is a global media and technology company with three primary businesses: Comcast Cable, NBCUniversal, and Sky. Comcast Cable is one of the United States' largest video, high-speed internet, and phone providers to residential customers under the Xfinity brand, and also provides these services to businesses. It also provides wireless and security and automation services to residential customers under the Xfinity brand. NBCUniversal is global and operates news, entertainment and sports cable networks, the NBC and Telemundo broadcast networks, television production operations, television station groups, Universal Pictures, and Universal Parks and Resorts. Sky is one of Europe's leading media and entertainment companies, connecting customers to a broad range of video content through its pay television services. It also provides communications services, including residential high-speed internet, phone, and wireless services. Sky operates the Sky News broadcast network and sports and entertainment networks, produces original content, and has exclusive content rights.

→ [www.comcastcorporation.com](http://www.comcastcorporation.com)

**THE INTERVIEWER**

Professor Christian Hildebrand conducted the interview in June 2019.

# Talking Versus Typing: The Power of Voice-Based Remote Controls

Interview with Jan Neumann, Senior Director, Applied AI,  
Comcast Cable, Philadelphia, USA

While many customers are still reluctant to entrust themselves to Alexa, Cortona or Siri in their homes, they seem to be less worried about controlling their TV sets via voice control. Comcast started offering a voice-based remote control in 2015 and has extended this service continuously. In the vast world of home entertainment, it seems that voice has come just in time to help consumers navigate and control their ever-increasing home entertainment options. Jan Neumann, Head of AI at Comcast, explains how Comcast enables its customers to comfortably boil down a huge entertainment portfolio to personally relevant content on the TV screen, and how the company remains successful in the highly competitive home entertainment market.



**Christian Hildebrand** ✕ *The rise of voice control is a very hot topic these days. Everyone is talking about Alexa, Google Home, and Siri. Comcast recently introduced a voice-based remote control. How does voice reinvent the TV experience?*

**Jan Neumann** ✕ One of the major issues that TV customers were facing was that more and more content was available but that the interfaces for finding that content had not changed. You still had the traditional remote controls with numbers. In a world where all you had to do was switch between, let's say, 20 TV channels when you could remember the corresponding numbers, this might have worked fine. But now you have hundreds of thousands of different items

that you can watch at any time. In this situation, enabling a more complex interface to indicate what you are looking for is super powerful.

*Is voice just providing a different or ultimately a better customer experience?*

Voice control actually inverts the interface. Traditionally, the interface dictated to the customer how to interact with the interface. Now we allow the customers to express themselves, and then it is up to us to interpret that. This process favors the customer instead of the platform. We say that voice flattens the interface. It is the perfect shortcut device: You simply tell us what you are interested in, and it is up to us to under-



*The voice-based remote control is actually one of the main reasons to stay with our platform.*



stand and deliver that to you on the screen. We can limit the amount of information presented so that you can again use the traditional remote control to navigate. Painful typing into the TV becomes unnecessary.

***From other voice-based devices, we know that there is quite a difference between adoption and usage rates. Usage does not match the sales figures. Do you know if people are actually using your voice-based remote control?***

Yes, it is one of the most popular products that we have put out, and it has been extremely successful. Based on user feedback, it is a very large driver of customer satisfaction and retention. Once customers get used to it, they use it heavily on a daily basis. The voice-based remote control is actually one of the main reasons to stay with our platform. We process more than half a billion voice commands a month, and this number is climbing.

***Is voice just a new channel to connect to your customers or does it also change behavior more fundamentally? Do people search differently using voice?***

Users start out with straightforward use cases, for example, searching for “CNN” or “NBC,” or certain channel numbers. Once you give them examples of more complex use cases, their searches become more sophisticated. We are constantly improving the functionality. For instance, users can ask for “results of the current NBA playoffs,” like between the Raptors and the Warriors, and we are able to come up with the statistics as well as when the game is playing. So basically, the users now just have to describe what they want to watch.

***Could you give us another example of how a user might search and what he or she will get?***

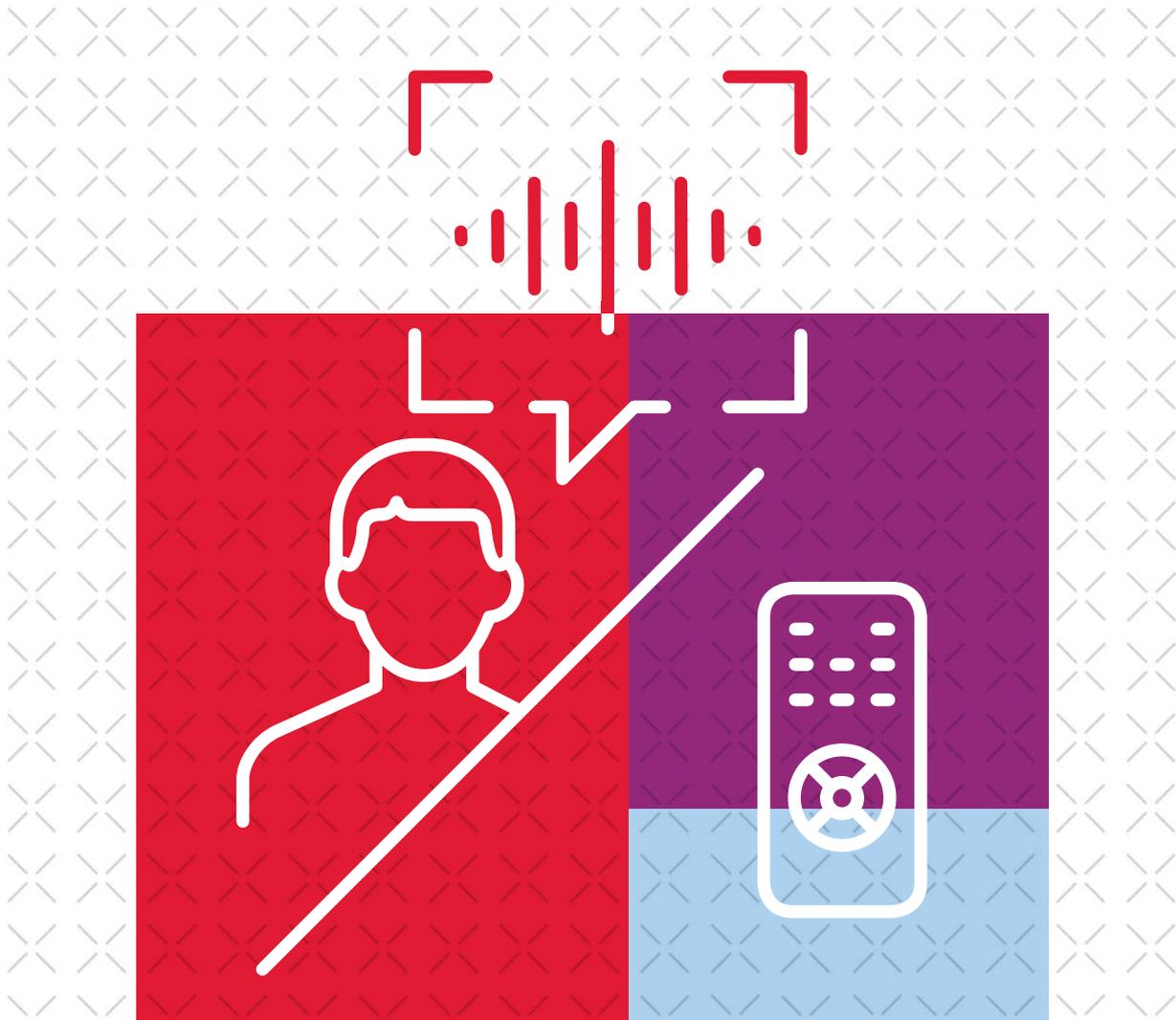
Very recently, we introduced a new feature. Users can describe what happens in an episode they want to watch, and they no longer need to remember its title or number. They simply need to say, “Show me the Friends episode with Brad Pitt,” and that’s what they will be shown on screen.

***How are people talking to the voice-based remote control? Very naturally, like to another person, or do they just throw in single key words?***

People often start with simpler, command-like input. But they can say virtually anything. It’s trial and error for them and for us as well. We do not tell them which commands they should use or what to say. We give users the power and freedom to express exactly what they want. This way we can learn what is popular or interesting. Through listening, we learn what customers want, and what next feature we might deliver.

***How can you create a better customer experience when people don’t know all the cool features of the voice interface? Do you have to actively encourage and educate customers on how to use the interface?***

To educate our customers, we use the screensaver regularly, for instance. Also, we have lots of advertising that describes potential use cases. Just recently, we started rolling out the voice suggestions feature, which actively recommends what consumers might want to search for. In this way, we teach customers some good expressions to use as well as new search possibilities. Our suggestions are contextually rel-



evant; users are more likely to remember and use them next time because they fit and seem like a good idea. This feature is so new that we have not seen measurable results yet, but we are very excited about it.

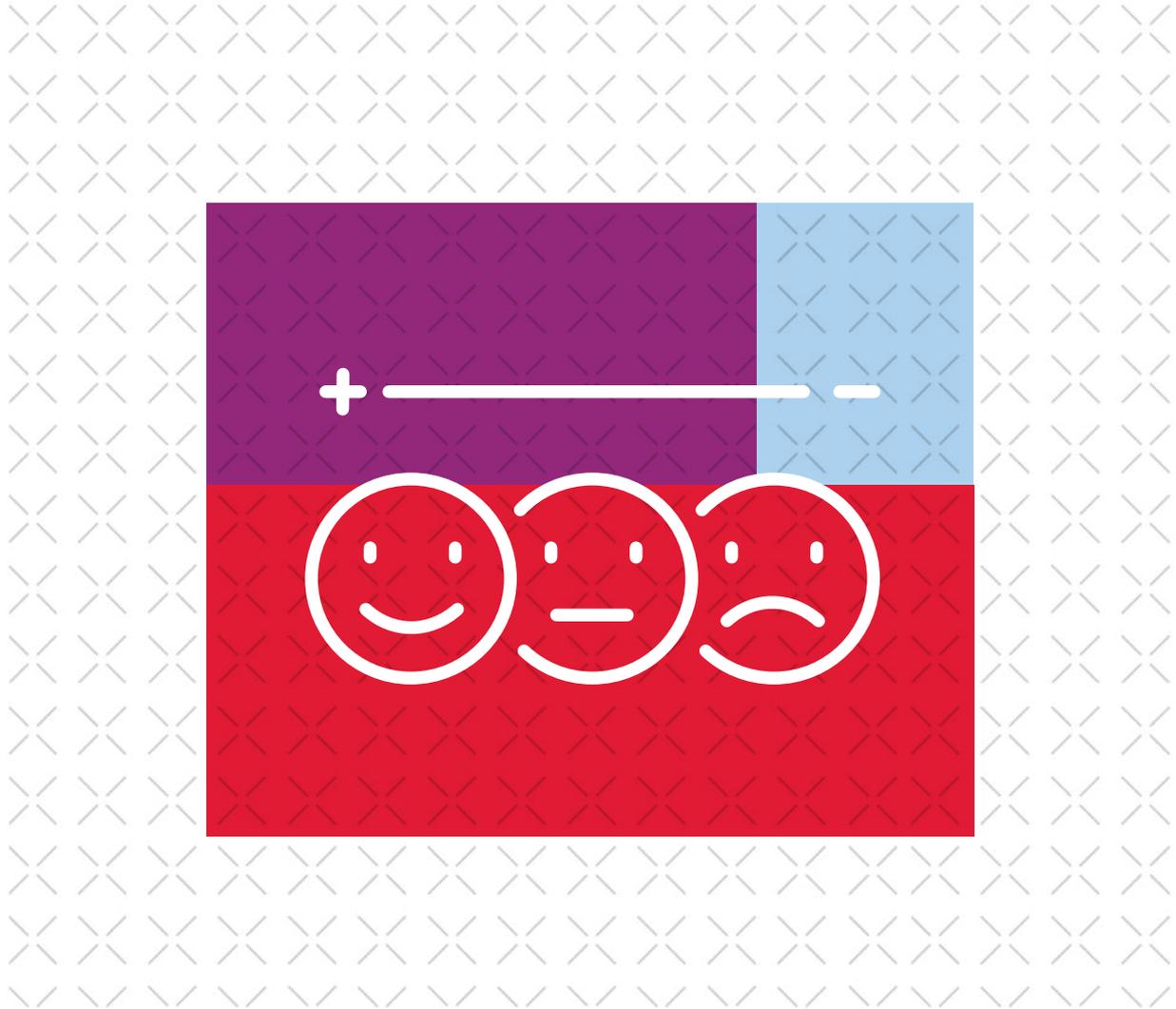
***Identifying intent is often difficult and messy. How do you leverage natural language processing to increase understanding and prevent having to say the phrase Alexa users know too well – “Sorry, I don’t know that?”***

One advantage we have is that we do not offer a general assistant for everything. Our voice service is focused on entertainment, on home control, and on customer service. In each of these domains, we have a pretty good understanding of the entities involved and of possible actions. Customer service is a little bit more complex. It involves longer sentences and chats in contrast to the rather short voice commands in



*One challenge is how to use the screen on top of voice.*





entertainment, which are mostly quite to the point. Our service is far less complex because the more services you involve, the harder it gets to identify the intent.

**Another challenge must be managing your rich portfolio of constantly changing content, with new shows coming up all the time. How do you manage and handle this portfolio? How are you able to produce well-tailored, timely recommendations for customers?**

It is important to have strong metadata generation and management. This means that you must be able to automatically detect who is in a program, what they are talking about in a new show or video, which viewer segments might be interested, etc. It is about identifying content on a deeper level and having rich descriptions of what can be expected, like which emotions are involved or which music is playing. We fill in the blanks of what hasn't been provided, to enable a richer interaction pattern.

**Which lessons have you learned from taking the voice remote to the market? Do you see anything that applies to other industries interested in getting into voice-controlled interfaces?**

With entertainment in particular, you have a screen and you want to use it. So the experience is not only about voice – it needs to be unified and combined, with text and talk together. One challenge is how to use the screen on top of voice. You can use voice to get a list of funny comedies, but instead of having to command “next page, next page” or “second on the right side,” it makes sense to allow navigation on screen. You need to combine the strengths of both native interfaces.

**Now let's talk more generally about how AI and your analytics focus have changed the inner workings of Comcast. How has the transition from a “system administration culture” to one that is focused on continuous delivery of a better customer experience changed the organization?**



*It comes down to the question of who can learn the fastest from the customers and be the most agile in meeting their needs.*



The biggest change was putting the customer first. Building everything around customers and their satisfaction changes the focus, the type of projects, and the entire approach to business. It is extremely motivating to see the impact we are making with that switch to technology. Instead of having to make assumptions about what the customers might like, we get this information directly from our technologies. We can learn so much faster, and this is really exciting.

***Ironically, a few years ago, Comcast was criticized for a lack of customer orientation. Now you seem to be changing the game by leveraging AI to provide a better customer experience. This is quite impressive.***

Thank you. It's been a companywide effort that includes a lot more than technology and software, and it's super exciting to see how customers have responded. We know we still have a lot more to do, but as a technologist and an AI scientist, I've been particularly excited to apply this technology to make our customers' experiences easy and frictionless.

***Yet you are competing with Netflix and other global tech companies that are trying to eat your market share. Isn't this extremely challenging, or even frightening?***

As an engineer, I would say it's much more motivating. Having so many companies doing so many exciting things really inspires me and my team to try new things and constantly stay ahead of the curve. Engineers are pretty competitive people as a group, so when we see someone else doing something cool, it provides both inspiration and motivation.

***In the end, do you have a competitive advantage if you have the best algorithms?***

Well, algorithms are the tool for achieving something. It comes down to the question of who can learn the fastest from the customers and be the most agile in meeting their needs. Fundamentally it is the same business for everybody, but there are certain ways to accelerate the feedback cycle and be more efficient via technology.

***Finally, let's have a look into the interface of the future. What's the next big thing that will better connect and serve your customers? Brain interfaces, or more sensors in people's homes?***

Whatever the interface is, it all comes down to being able to actively understand and anticipate the needs of the customers, and then being able to serve them, fulfill their needs, and address their intentions. Voice is a natural way of communicating, and for now it is a very good medium. We also use nonverbal cues, and maybe they can play a role in the future. But in the end, the medium is irrelevant. Whatever allows us to capture needs effectively and with as little friction as possible is what we will end up with.

***Wow, what closing words! Thank you for your time, Jan, and for taking us with you into the new tech universe at Comcast!*** ✕