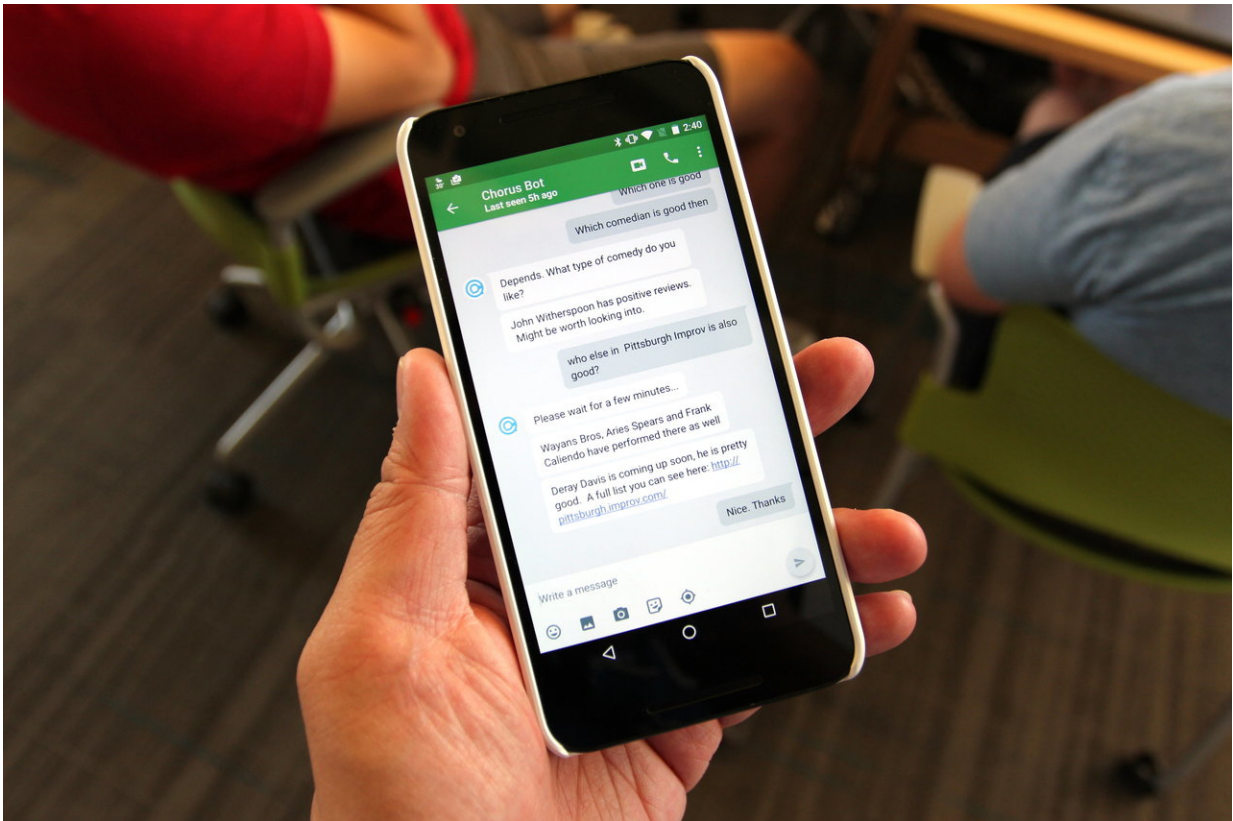# Crowd workers, AI make conversational agents smarter

February 7 2018

Evorus is a chatbot that uses a combination of artificial intelligence and crowd workers to answer questions on a broad range of topics. Credit: Carnegie Mellon University

Conversational agents such as Siri, Alexa and Cortana are great at giving

you the weather, but are flummoxed when asked for unusual information, or follow-up questions. By adding humans to the loop, Carnegie Mellon University researchers have created a conversational agent that is tough to stump.

The chatbot system, called Evorus, is not the first to use human brainpower to answer a broad range of questions. What sets it apart, says Jeff Bigham, associate professor in the Human-Computer Interaction Institute, is that humans are simultaneously training the system's artificial intelligence, making it gradually less dependent on people.

Like an earlier CMU agent called Chorus, Evorus recruits crowd workers on demand from Amazon Mechanical Turk to answer questions from users, with the crowd workers voting on the best answer. Evorus also keeps track of questions asked and answered and, over time, begins to suggest these answers for subsequent questions. The researchers also have developed a process by which the AI can help to approve a message with less crowd worker involvement.

"Companies have put a lot of effort into teaching people how to talk to conversational agents, given the devices' limited command of speech and topics," Bigham said. "Now, we're letting people speak more freely and it's the agent that must learn to accommodate them."

The system isn't in its final form, but it is available for download and use by anyone willing to be part of the research effort: http://talkingtothecrowd.org/.

A research paper on Evorus, already available online, will be presented by Bigham's research team later this year at CHI 2018, the Conference on Human Factors in Computing Systems in Montreal.

Totally automated conversational agents can do well answering simple,

common questions and commands and can converse in depth when the subject is relatively narrow, such as advising on bus schedules. Systems with people in the loop can answer a wide variety of questions, Bigham said, but with the exception of concierge or travel services for which users are willing to pay—agents that depend on humans are too expensive to be scaled up for wide use. A session on Chorus costs an average of $2.48.

"With Evorus, we've hit a sweet spot in the collaboration between the machine and the crowd," Bigham said. The hope is that as the system grows, the AI is able to handle an increasing percentage of questions, while the number of crowd workers necessary to respond to "long tail" questions will remain relatively constant.

Keeping humans in the loop also reduces the risk that malicious users will manipulate the conversational agent inappropriately, as occurred when Microsoft briefly deployed its Tay chatbot in 2016, said Ting-Hao Huang, a Ph.D. student in the Language Technologies Institute (LTI). Huang developed Evorus with Bigham and Joseph Chee Chang, also a Ph.D. student in LTI.

During Evorus' five-month deployment with 80 users and 181 conversations, automated responses to questions were chosen 12 percent of the time, crowd voting was reduced by almost 14 percent and the cost of crowd work for each reply to a user's message dropped by 33 percent.

Evorus is a text chatbot, but is deployed via Google Hangouts, which can accommodate voice input, as well as access from computers, phones and smartwatches. To enhance its scalability, Evorus uses a software architecture that can accept automated question-answering components developed by third parties.