



A Moderator ChatBot for Civic Discourse

By Katharine Miller

February 7, 2020

HAI seed grant helps scale-up Stanford-designed deliberative discourse strategy to an online platform

“Thank you all for joining today,” the chatbot says in a calm, friendly voice. “This discussion is an opportunity for everyone to learn.”

Thus begins a discussion about the pros and cons of the electoral college on the Stanford Online Deliberation platform. Like a Google hangout or Zoom meeting, everyone in the chatroom can see the person who is speaking and join the conversation by pressing the “request to speak” button. But there’s a difference: The conversation follows a protocol for civic discussion created by Stanford’s [Center for Deliberative Democracy](#), and relies on an automated moderator designed in collaboration with a team from Stanford’s Management Science and Engineering department.

The goal is to encourage civic discourse around important policy questions.

“Social media and the internet are making people more divided and polarized,” said Sukolsak Sakshuwong, a PhD candidate in Stanford’s Management Science and Engineering program who is the lead developer of the project. “But I think there are ways we can use technology for good, to bring people together to learn from each other.”

About 25 years ago, James Fishkin, director of Stanford’s Center for Deliberative Democracy and professor of communication and of political science at Stanford, first tested a unique approach to civic discourse. Called Deliberative Polling®, the method provides what Fishkin calls “good conditions” for face-to-face small group deliberation. Over the years, it has enabled high-quality conversations around many diverse policy issues all over the world. Typically, participants are polled before and after their deliberations to determine whether their opinions have shifted. Often, they have.

But the standard model of Deliberative Polling® doesn’t scale well to larger populations. “The goal, in our dreams, is to have hundreds of thousands of groups all deliberating at the same time,” said Alice Siu, associate director of the Center for Deliberative Democracy. That’s where technology and artificial intelligence can make a difference.

So Fishkin and Siu have collaborated with Ashish

Goel, professor of Management Science and Engineering at Stanford, to bring Deliberative Polling® into the digital age. They realized almost immediately that an automated moderator would be necessary. “It would be too hard to train thousands of human moderators if you want to scale to an online platform,” Siu said.

The moderator chatbot Goel’s students designed strives to mimic a human moderator, but it also does things a human moderator cannot. And some of the moderator’s skills are boosted with artificial intelligence, work that was funded with a.

“A human moderator would call on someone to speak, would engage with people who talk too much or not enough, would ask for views on certain issues, would keep time, keep the agenda,” Siu said. “All of those things are built into the platform.”

For example, on the platform, participants put their names in a queue when they want to speak, with each opportunity to speak limited to 45 seconds. This helps maintain civility and order during the deliberations. The platform also uses a timed agenda with prompts to ask if people want to keep talking about an item or move on.

There are, of course, human tasks that a chatbot cannot do well, Goel notes. A human moderator can read body language, knows which topics are going to be sensitive, and can sense the undercurrents in the room. But a chatbot can also do things that a human moderator cannot. Because the platform has the

ability to engage with everyone simultaneously, for example, it can nudge people to engage if they aren't, and poll them with various questions.

For certain tasks, the team is experimenting with the use of artificial intelligence to improve the moderator chatbot's skills. For example, the platform uses a tool developed by Google to transcribe what participants are saying in real time. Then, using natural language processing (via the Perspective API by Jigsaw and Google) it evaluates that text for offensive comments to output a toxicity score.

“We can use that to understand which people are on the brink of being too toxic,” said Lodewijk Gelauff, a PhD student in Management Science and Engineering at Stanford who is the student in charge of experiment design and deployments. If a certain threshold toxicity score is detected, the bot polls those who are not speaking to confirm whether the speaker is saying something offensive. If the majority says yes, the person will cut off from speaking for a few minutes.

“In a real conversation, you can't constantly pull people aside to ask if they find a particular person or comment offensive,” Goel said. “But online, because people are all on separate screens, we can give a targeted prompt to each individual.”

The team is also experimenting with using artificial intelligence for agenda management, sending participants a prompt if the conversation seems to have gone off track. “The tools for agenda

management aren't perfect yet," Goel notes. But the team is continuing to work on it, and is also experimenting with AI for detecting tonal variations (is this person angry, for example).

The Stanford Online Deliberation Platform has so far been used in classroom settings and is being more thoroughly tested in a controlled experiment. And in the Spring of 2020, it will go live to a random sample of 200 people in Japan. They will discuss climate change policy in 12-15 person groups on the online forum. After that, there will be a similar test in Korea. "We want to test the platform as much as possible," Siu said.

The prospect of scaling the project excites Gelauff. "It's really bringing people closer together—having people with different opinions have a constructive conversation with one another and walk away with a better understanding of the opposing arguments," he said.

The Stanford Online Deliberation platform also fulfills a key mission for HAI as well as for Goel himself. "Our mission is to take societal decision making into the digital age in a positive way," Goel said. If you look at industry, he said, "it's not clear that anyone else is doing this."

Contact

Faculty Director
Fei-Fei Li and John Etchemendy
General Contact

Connect

