New research suggests robots appear more persuasive when pretending to be human

November 12 2019

Talal Rahwan, associate professor of Computer Science at NYU Abu Dhabi. Credit: NYU Abu Dhabi

Recent technological breakthroughs in artificial intelligence have made it possible for machines, or bots, to pass as humans. A team of

researchers led by Talal Rahwan, associate professor of Computer Science at NYU Abu Dhabi, conducted an experiment to study how people interact with bots whom they believe to be human, and how such interactions are affected once bots reveal their identity. The researchers found that bots are more efficient than humans at certain human-machine interactions, but only if they are allowed to hide their non-human nature.

In their paper titled "Behavioral Evidence for a Transparency-Efficiency Tradeoff in Human-Machine Cooperation" published in *Nature Machine Intelligence*, the researchers presented their experiment in which participants were asked to play a cooperation game with either a [human] associate or a bot associate. This game, called the Iterated Prisoner's Dilemma, was designed to capture situations in which each of the interacting parties can either act selfishly in an attempt to exploit the other, or act cooperatively in an attempt to attain a mutually beneficial outcome.

Crucially, the researchers gave some participants incorrect information about the identity of their associate. Some participants who interacted with a human were told they were interacting with a bot, and vice versa. Through this experiment, researchers were able to determine whether people are prejudiced against social partners they believe to be bots, and assess the degree to which such prejudice, if it exists, affects the efficiency of bots that are transparent about their non-human nature.

The results showed that bots posing as humans were more efficient at persuading the partner to cooperate in the game. However, as soon as their true nature was revealed, cooperation rates dropped and the bots' superiority was negated.

"Although there is [broad consensus] that machines should be transparent about how they make decisions, it is less clear whether they should be

transparent about who they are," said Rahwan.

"Consider, for example, Google Duplex, an automated voice assistant capable of generating human-like speech to make phone calls and book appointments on behalf of its user. Google Duplex's speech is so realistic that the person on the other side of the phone may not even realize that they are talking to a bot. Is it ethical to develop such a system? Should we prohibit bots from passing as humans, and force them to be transparent about who they are? If the answer is 'Yes', then our findings highlight the need to set standards for the efficiency cost that we are willing to pay in return for such transparency."

Provided by New York University