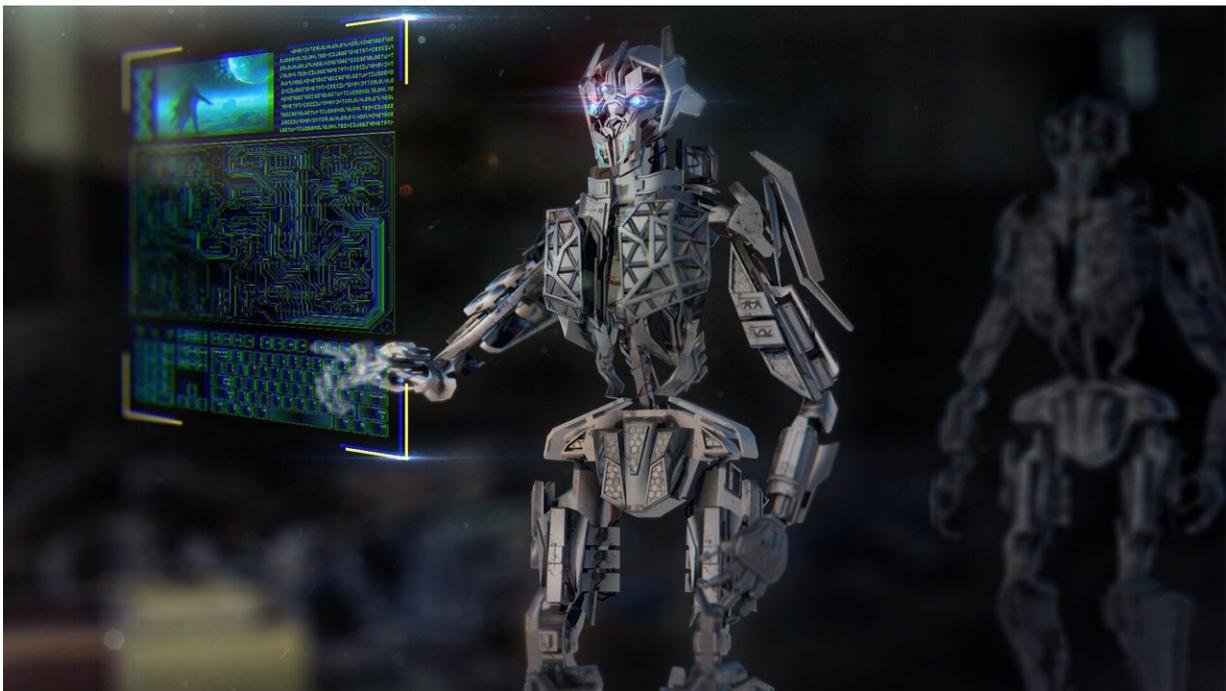# People who distrust fellow humans show greater trust in artificial intelligence

September 22 2022, by Jonathan F. McVerry

Credit: Pixabay/CC0 Public Domain

A person's distrust in humans predicts they will have more trust in artificial intelligence's ability to moderate content online, according to a recently published study. The findings, the researchers say, have practical implications for both designers and users of AI tools in social media.

"We found a systematic pattern of individuals who have less trust in other humans showing greater trust in AI's classification," said S. Shyam Sundar, the James P. Jimirro Professor of Media Effects at Penn State. "Based on our analysis, this seems to be due to the users invoking the idea that machines are accurate, objective and free from ideological bias."

The study, published in the journal of *New Media & Society* also found that "power users" who are experienced users of information technology, had the opposite tendency. They trusted the AI moderators less because they believe that machines lack the ability to detect nuances of human language.

The study found that individual differences such as distrust of others and power usage predict whether users will invoke positive or negative characteristics of machines when faced with an AI-based system for content moderation, which will ultimately influence their trust toward the system. The researchers suggest that personalizing interfaces based on individual differences can positively alter user experience. The type of content moderation in the study involves monitoring social media posts for problematic content like hate speech and suicidal ideation.

"One of the reasons why some may be hesitant to trust content moderation technology is that we are used to freely expressing our opinions online. We feel like content moderation may take that away from us," said Maria D. Molina, an assistant professor of communication arts and sciences at Michigan State University, and the first author of this paper. "This study may offer a solution to that problem by suggesting that for people who hold negative stereotypes of AI for content moderation, it is important to reinforce human involvement when making a determination. On the other hand, for people with positive stereotypes of machines, we may reinforce the strength of the machine by highlighting elements like the accuracy of AI."

The study also found users with conservative political ideology were more likely to trust AI-powered moderation. Molina and coauthor Sundar, who also co-directs Penn State's Media Effects Research Laboratory, said this may stem from a distrust in mainstream media and social media companies.

The researchers recruited 676 participants from the United States. The participants were told they were helping test a content moderating system that was in development. They were given definitions of hate speech and suicidal ideation, followed by one of four different social media posts. The posts were either flagged for fitting those definitions or not flagged. The participants were also told if the decision to flag the post or not was made by AI, a human or a combination of both.

The demonstration was followed by a questionnaire that asked the participants about their individual differences. Differences included their tendency to distrust others, political ideology, experience with technology and trust in AI.

"We are bombarded with so much problematic content, from misinformation to hate speech," Molina said. "But, at the end of the day, it's about how we can help users calibrate their trust toward AI due to the actual attributes of the technology, rather than being swayed by those individual differences."

Molina and Sundar say their results may help shape future acceptance of AI. By creating systems customized to the user, designers could alleviate skepticism and distrust, and build appropriate reliance in AI.

"A major practical implication of the study is to figure out communication and design strategies for helping users calibrate their trust in automated systems," said Sundar, who is also director of Penn State's Center for Socially Responsible Artificial Intelligence. "Certain

groups of people who tend to have too much faith in AI technology should be alerted to its limitations and those who do not believe in its ability to moderate content should be fully informed about the extent of human involvement in the process."