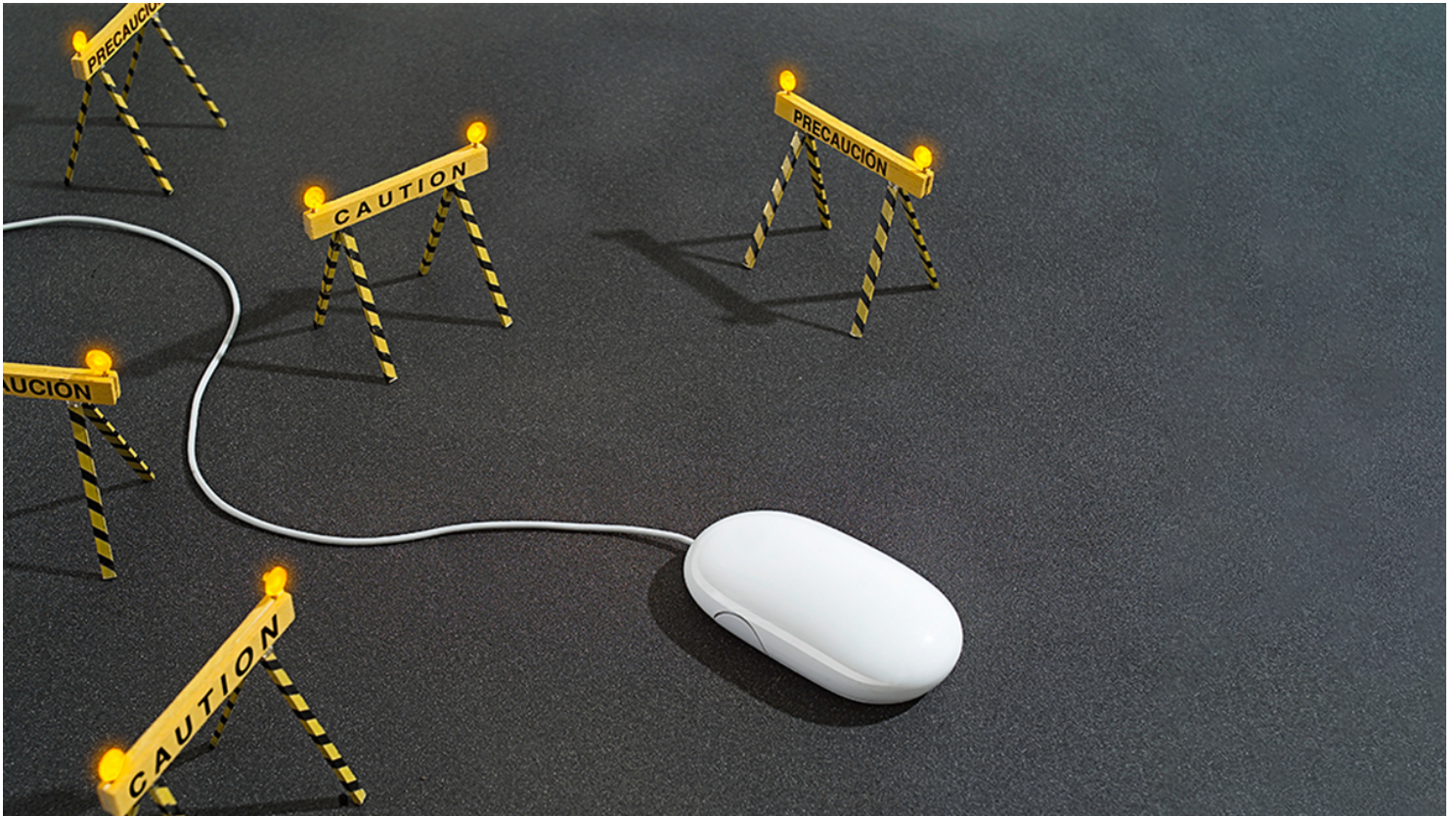# Can Algorithms Help Us Decide Who to Trust?

by David De Cremer, Jack McGuire, Yorck Hesselbarth, and Ke Michael Mai

JUNE 04, 2019

C. J. BURTON/GETTY IMAGES

The use of artificial intelligence (AI) and algorithms is increasing within organizations to manage business processes, hire employees, and automate routine organizational decision making. This comes as no surprise, since the application of simple linear algorithms have been shown to outperform human judgment in the accuracy of many administrative tasks. A 2017 Accenture survey also revealed that 85% of executives want to invest more extensively in AI-related technologies over the next three years.

Despite this forecast, the reality is that, at least in some cases, humans display strong feelings of aversion to the use of autonomous algorithms. For example, surveys reveal that 73% of Americans report that they are afraid to ride in a self-driving vehicle. Human doctors are also preferred over algorithms in the medical context, despite evidence that algorithms might sometimes deliver more accurate diagnoses. Such aversion creates work situations where the implementation of AI leads to a sub-optimal, inefficient, and biased use of algorithms. So, if AI is to become an important management tool in our organizations, algorithms need to be used as trusted advisors to human decision-makers. They should also help promote trust within the company.

This conclusion has not gone unnoticed within the business community. Companies like Trust Science have put algorithms on the market that can help compile trustworthiness profiles of individuals and organizations. But does AI really possess such a "social" skill? This is an important question to ask because trust requires socially sensitive skills that are perceived to be uniquely human. In fact, the unique ability to understand human emotions and desires is a prerequisite for judging individual's trustworthiness and is hard to resemble artificially. So can algorithms providing advice in this area of human interaction be accepted by human decision-makers?

To examine this question, we ran an experimental study via the online system Prolific Academic (ProA), which has earned a reputation of providing a high level of data quality. We analyzed the responses of a total of 136 participants (68 females and 66 males, average work experience of 8.73 years), who participated in an online simulation of an organization consisting of three hierarchical levels: employees, middle management, and top management. Participants were allocated to an employee position and were informed that they would be partnered up with another unknown person to work on a task. Good performance would increase their work budget.

Participants then received information about this individual, which was a judgment score indicating the individual's trustworthiness. The individual's trustworthiness level could be determined in one of two ways. The first way involved an assessment conducted by an algorithm. When explaining this option, participants were shown the website of a company selling algorithms developed to provide advice about the trustworthiness of individuals and explained

that this algorithm would be used in the experiment. A second way to assess the trustworthiness level of the individual involved a 15-minute conversation between the leader of the study and the unknown person.

Our results suggest that people think of humans and algorithms as good at providing different types of information, including about who to trust. Humans are seen as a better source of intuition, better at social skills, and better at taking another person's perspective. But algorithms can provide information about who to trust in cases where that information is less intuitive and more factual. In other words, participants considered humans to possess more appropriate skills needed to take the perspective of other humans in social interactions than AI does. In further support of this idea, the algorithm was regarded as a more rational and less intuitive approach in evaluating individual's trustworthiness. Our participants, however, also indicated that if an algorithm could deliver information about the trustworthiness of a human, this would not make them feel more uncertain about the reliability, authenticity, and accuracy of the information provided compared to a human delivering this information.

So, although humans were judged to possess the necessary social skills to assess someone's trustworthiness, they did not feel that the use of an algorithm would reveal less reliable social trust information. This finding was further endorsed by the observation that, when participants were asked to indicate which assessment method they preferred to use, most participants opted to use AI (61%) rather than the judgments of the human (39%).

Participants were then randomly allocated to a condition in which either the algorithm or the human experimenter would determine the trustworthiness of their new colleague. We also manipulated the level of trustworthiness communicated to the participants by having half of the participants receive information that their new work partner was high in trustworthiness, whereas the other half received information that their partner was low in trustworthiness. Participants were then paired with their new colleague and required to play a trust game. The participants received a work budget of 10 units (each unit was worth 2 pence) and had to decide how many units they wanted to transfer to their new colleague. The number of units sent was tripled. Their colleague then decided how much to give back. The number of units given back was not tripled. So, by enlarging the work budget, both parties can earn more; but this could only be achieved if the participant was willing to trust their new colleague in the first place.

As expected, participants transferred more units when the new colleague was high, rather than low, in trustworthiness. Surprisingly, the source of the trustworthiness information — algorithm or human — did not influence the trusting behavior of the participants. This observation further underscores the idea that people do not consider the rational assessment method of AI to be less suited than humans to evaluate an individual's trustworthiness.

What are the implications of these findings for organizations?

First, many team projects are often temporary in nature and require bringing employees together who do not yet know each other. In such settings, trust needs to be built quickly. People do so by screening distinctive features like someone's background, the job label, and other information that is easy to scan. This type of trust building involves no emotions and leads to what is called "swift trust". Our findings show that to facilitate this type of "cognitive" trustworthiness information, AI presents a reliable and legitimate assessment tool. Moreover, employing AI for this kind of swift assessment also reduces the time needed for employees to get to know each other's background.

Second, social skills like perspective-taking, intuition, and social sensitivity are prerequisites to determine someone's trustworthiness and are considered to be uniquely human. Our findings nevertheless indicate that when it comes down to starting a work relationship with a colleague, algorithms seem to be accepted as being equally reliable.

It is, however, important to stress that we examined the initial phase of a relationship where people are still strangers to each other. Prior research has indicated that in that case, people rely more on cognitive-based information that is gathered in rational and thus less emotional ways. So, AI as a source of trustworthiness information seems well-suited to be employed in those initial stages of a work relationship. When relationships are developing, however, people become more reliant on emotion-driven information, which AI is not able to provide yet. We therefore argue that the predictive impact of human judgments will increase (and AI assessments decrease) when it comes down to fostering trusting work relationships over the long term.

Third, implementing the use of AI as advisors promoting trusting work cultures implies that organizations will need to train humans supervising those cultures. Supervisors will have to learn to develop a sense of awareness about when it is effective and when not to delegate assessments

of the work climate to an algorithm. Finally, supervisors will also have to learn how to communicate the trustworthiness information provided by AI to their teams in ways that will not be ignored.

David De Cremer is the Provost chair and professor in management and organizations at NUS Business School, National University of Singapore, a visiting fellow at the Hoover Institution, Stanford University, and a member of the Justice Collaboratory at Yale Law School. Before moving to NUS, he was the KPMG endowed professor in management studies at Judge Business School, University of Cambridge. He is the founder of the Leading Artificial intelligence & Digital management (LEAD) platform focusing on the role of humanity in the new technology era. He has published over more than 250 articles and book chapters and is the author of the book *Pro-active Leadership: How to overcome procrastination and be a bold decision-maker.*

Jack McGuire is the Experimental Lab Manager of the Cambridge Experimental and Behavioural Economics Group (CEBEG) and Research Assistant in the Department of Organisational Leadership and Decision-Making at Judge Business School, University of Cambridge. He has obtained an MA (SocSci) from the University of Glasgow, an MSc from University College London, and has also held research positions in the University of Tubingen and the University of Hong Kong. His research interests are primarily in behavioural ethics, decision making, and organizational behaviour.

Yorck Hesselbarth is a PhD student in Leadership, Decision Making and Artificial Intelligence, currently at ESCP Europe in Berlin.

Ke Michael Mai is an assistant professor at the department of management and organization, NUS Business School, National University of Singapore.

This article is about COLLABORATION

⊕ Follow This Topic

Related Topics: Technology

Comments

Leave a Comment

Post Comment

2 COMMENTS

S. Brovkin   8 days ago

… or WHOM to trust rather.
Can't trust anybody these days, even the HBR editors.

↪ Reply   0 👍 0 👎

∨ Join The Conversation

POSTING GUIDELINES
We hope the conversations that take place on HBR.org will be energetic, constructive, and thought-provoking. To comment, readers must sign in or register. And to ensure the quality of the discussion, our moderating team will review all comments and may edit them for clarity, length, and relevance. Comments that are overly promotional, mean-spirited, or off-topic may be deleted per the moderators' judgment. All postings become the property of Harvard Business Publishing.