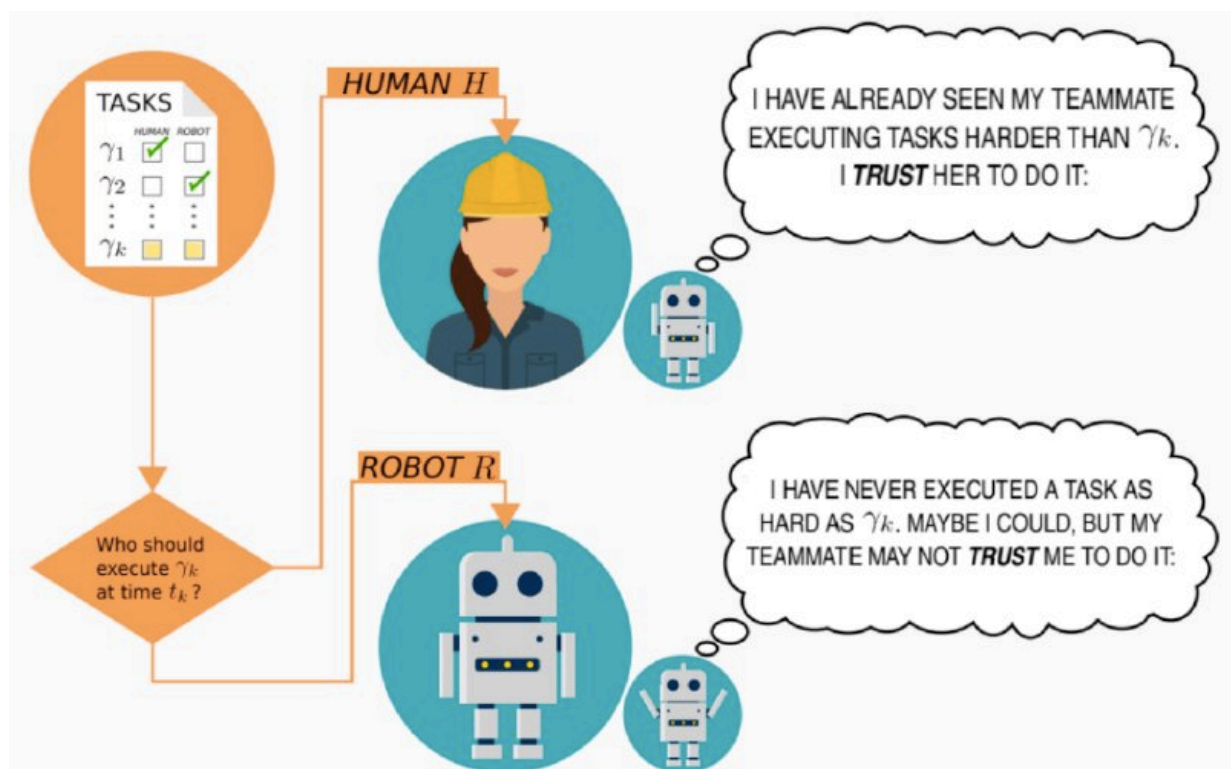


# A model to predict how much humans and robots can be trusted with completing specific tasks

June 29 2021, by Ingrid Fadelli



A team formed by a human and a robot that must collaborate sequentially executing tasks. Each task must be executed by one of the agents. The bi-directional trust model can be used for predicting the human's natural trust in a robot to execute a task, and for the robot to determine its artificial trust in the human to execute a task. Credit: Azevedo-Sa et al.

Researchers at University of Michigan have recently developed a bi-directional model that can predict how much both humans and robotic agents can be trusted in situations that involve human-robot collaboration. This model, presented in a paper published in *IEEE Robotics and Automation Letters*, could help to allocate tasks to different agents more reliably and efficiently.

"There has been a lot of research aimed at understanding why humans should or should not [trust](#) robots, but unfortunately, we know much less about why robots should or should not trust humans," Hebert Azevedo-Sa, one of the researchers who carried out the study, told TechXplore. "In truly collaborative work, however, trust needs to go in both directions. With this in mind, we wanted to build robots that can interact with and build trust in humans or in other agents, similarly to a pair of co-workers that collaborate."

When humans collaborate on a given [task](#), they typically observe those they are collaborating with and try to better understand what tasks they can and cannot complete efficiently. By getting to know one another and learning what others are best or worst at, they establish a rapport of some kind.

"This where trust comes into play: you build trust in your co-worker to do some tasks, but not other tasks," Azevedo-Sa explained. "That also happens with your co-worker, who builds trust in you for some tasks but not others."

As part of their study, Azevedo-Sa and his colleagues tried to replicate the process through which humans learn what tasks others they are collaborating with can or cannot be trusted on using a [computational model](#). The [model](#) they created can represent both a [human](#) and a robot's trust, thus it can make predictions both about how much humans trust it and about how much it can trust a human to complete a given task.

"One of the big differences between trusting a human versus a robot is that humans can have the ability to perform a task well but lack the integrity and benevolence to perform the task," Azevedo-Sa explained. "For example, a human co-worker could be capable of performing a task well, but not show up for work when they were supposed to (i.e., lacking integrity) or simply not care about doing a good job (i.e., lacking benevolence). A robot should thus incorporate this into its estimation of trust in the human, while a human only needs to consider whether the robot has the ability to perform the task well."

The model developed by the researchers provides a general representation of an agent's capabilities, which can include information about its abilities, integrity and other similar factors, while also considering the requirements of the task that the agents is meant to execute. This representation of an agent's capabilities is then compared to the requirements of the task it is meant to complete.

If an agent is deemed more than capable of executing a given task, the model considers the agent as highly worthy of trust. On the other hand, if a task is particularly challenging and an agent does not seem to be capable enough or have the qualities necessary to complete it, the model's trust in the agent becomes low.

"These agent capability representations can also change over time, depending on how well the agent executes the tasks assigned to it," Azevedo-Sa said. "These representations of agents and tasks in terms of capabilities and requirements is advantageous, because it explicitly captures how hard a task is and allows that difficulty to be matched with the capabilities of different agents."

In contrast with other previously developed models to predict how much agents can be trusted, the model introduced by this team of researchers is applicable to both humans and robots. In addition, when they

evaluated their model, Azevedo-Sa and his colleagues found that it predicted trust far more reliably than other existing models.

"Previous approaches tried to predict trust transfer by assessing how similar tasks were, based on their verbal description," Azevedo-Sa said. "Those approaches represented a big first step forward for trust models, but they had some issues. For example: the tasks 'pick up a pencil' and 'pick up a whale' have very similar descriptions, but they are in fact very different."

Essentially, if a robot efficiently picked up or grasped a pencil, previously developed approaches for predicting trust would automatically assume that the same robot could be trusted to pick up a far bigger item (e.g., a whale). By representing tasks in terms of their requirements, on the other hand, the model devised by Azevedo-Sa and his colleagues could avoid this mistake, differentiating between different objects that a robot is meant to pick up.

In the future, the new bi-directional model could be used to enhance human-robot collaboration in a variety of settings. For instance, it could help to allocate tasks more reliably among teams comprised of human and robotic agents.

"We would like to eventually apply our model to solve the task allocation problem we mentioned before," Azevedo-Sa said. "In other words, if a [robot](#) and a human are working together executing a set of tasks, who should be assigned each task? In the future, these agents can probably negotiate which tasks should be assigned to each of them, but their opinions will fundamentally depend on their trust in each other. We thus want to investigate how we can build upon our trust model to allocate tasks among humans and robots."

**More information:** A unified bi-directional model for natural and

artificial trust in human-robot collaboration. *IEEE Robotics and Automation Letters* (RA-L). [DOI: 10.1109/LRA.2021.3088082](https://doi.org/10.1109/LRA.2021.3088082).

© 2021 Science X Network

Citation: A model to predict how much humans and robots can be trusted with completing specific tasks (2021, June 29) retrieved 24 April 2024 from <https://techxplore.com/news/2021-06-humans-robots-specific-tasks.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.