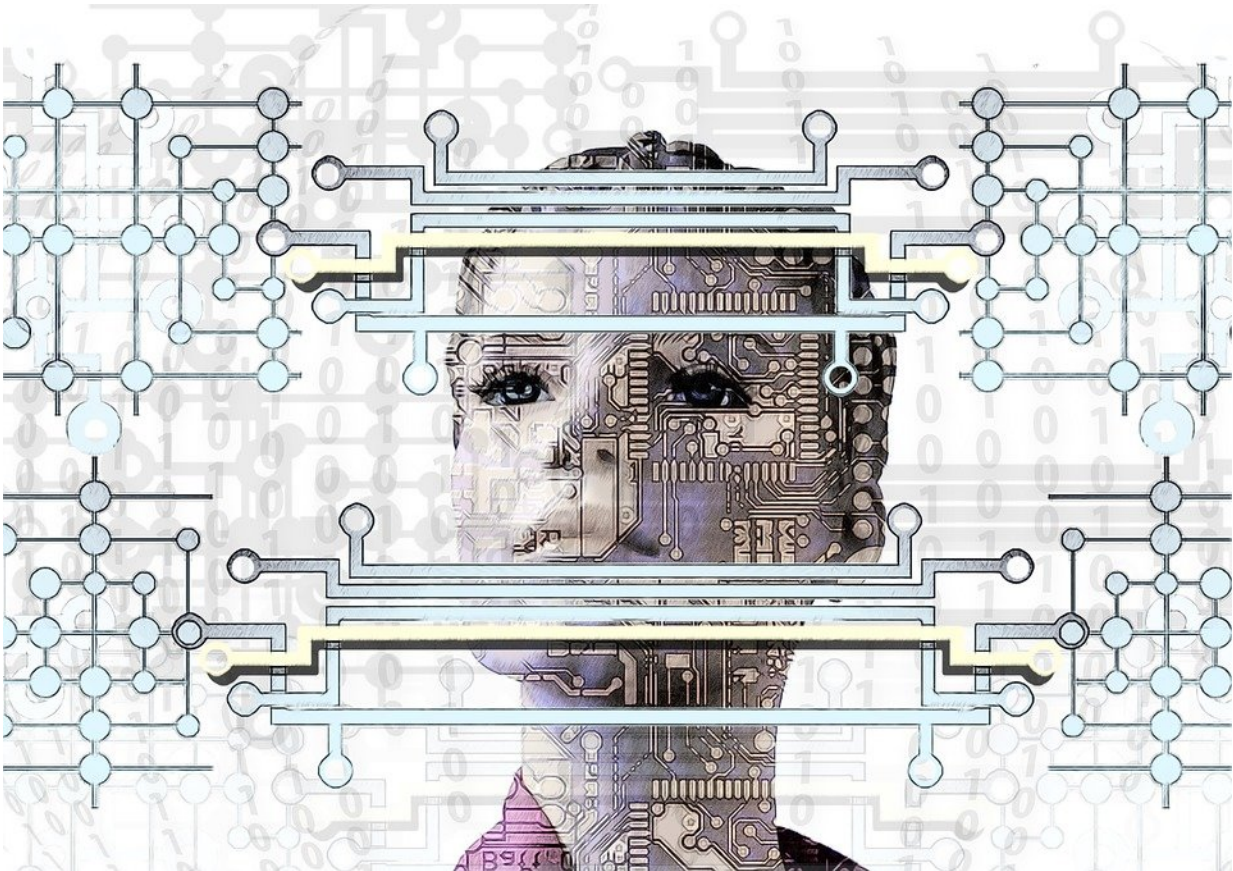


After Uber, Tesla incidents, can artificial intelligence be trusted?

April 10 2018, by Andrew Careaga



Credit: Missouri University of Science and Technology

Given the choice of riding in an Uber driven by a human or a self-driving version, which would you choose?

Considering last month's fatal crash of a self-driving Uber that took the life of a woman in Tempe, Arizona, and the recent death of a test-driver of a semi-autonomous vehicle being developed by Tesla, peoples' [trust](#) in the technology behind autonomous vehicles may also have taken a hit. The reliability of self-driving cars and other forms of [artificial intelligence](#) is one of several factors that affect humans' trust in AI, machine learning and other technological advances, write two Missouri University of Science and Technology researchers in a recent journal article.

"Trust is the cornerstone of humanity's relationship with artificial intelligence," write Dr. Keng Siau, professor and chair of business and information technology at Missouri S&T, and Weiyu Wang, a Missouri S&T graduate student in information science and technology. "Like any type of trust, trust in AI takes time to build, seconds to break and forever to repair once it is broken."

The Uber and Tesla incidents point to the need to rethink the way AI applications such as autonomous driving systems are developed, and for designers and manufacturers of these systems to take certain steps to build greater trust in their products, Siau says.

Despite these recent incidents, Siau sees a strong future for AI, but one fraught with trust issues that must be resolved.

'A dynamic process'

"Trust building is a dynamic process, involving movement from initial trust to continuous trust development," Siau and Wang write in "Building Trust in Artificial Intelligence, Machine Learning, and Robotics," published in the February 2018 issue of Cutter Business Technology Journal.

In their article, Siau and Wang examine prevailing concepts of trust in general and in the context of AI applications and human-computer interaction. They discuss the three types of characteristics that determine trust in this area – human, environment and technology – and outline ways to engender trust in AI applications.

Siau and Wang point to five areas that can help build initial trust in artificial intelligence systems:

- **Representation.** The more "human" a technology is, the more likely humans are to trust it. "That is why humanoid robots are so popular," Siau says, adding that it is easier to "establish an emotional connection" with a robot that looks and acts more like a human or a robotic dog that acts more like a canine. Perhaps first-generation autonomous vehicles should have a humanoid "chauffeur" behind the wheel to help ease concerns.
- **Image or perception.** Science fiction books and movies have given AI a bad image, Siau says. People tend to think of AI in dystopian terms, as colored by Terminator or Bladerunner movies or Isaac Asimov and Philip K. Dick novels. "This image and perception will affect people's initial trust in AI," Siau and Wang write.
- **Reviews from other users.** People tend to rely on online product reviews, and "a positive review leads to greater initial trust."
- **Transparency and "explainability."** When a technology's inner workings are hidden in a "black box," that opacity can hinder trust. "To trust AI applications, we need to understand how they are programmed and what function will be performed in certain conditions," Siau says.
- **Trialability.** The ability to test a new AI application before being asked to adapt it leads to greater acceptance, Siau says.

How to maintain trust in AI

Beyond developing initial trust, however, creators of AI also must work to maintain that trust. Siau and Wang suggest seven ways of "developing continuous trust" beyond the initial phases of product development:

- Usability and reliability. AI "should be designed to operate easily and intuitively," Siau and Wang write. "There should be no unexpected downtime or crashes."
- Collaboration and communication. AI developers want to create systems that perform autonomously, without human involvement. Developers must focus on creating AI applications that smoothly and easily collaborate and communicate with humans.
- Sociability and bonding. Building social activities into AI applications is one way to strengthen trust. A robotic dog that can recognize its owner and show affection is one example, Siau and Wang write.
- Security and privacy protection. AI applications rely on large data sets, so ensuring privacy and security will be crucial to establishing trust in the applications.
- Interpretability. Just as transparency is instrumental in building initial trust, interpretability – or the ability for a machine to explain its conclusions or actions – will help sustain trust.
- Job placement. As concerns about AI replacing humans on the job continue to grow, policies must be put in place to provide retraining and education to those affected by this trend.
- Goal congruence. "Since artificial intelligence has the potential to demonstrate and even surpass human intelligence, it is understandable that people treat it as a threat," Siau and Wang write. "Making sure that AI's goals are congruent with human goals is a precursor in maintaining continuous trust." Policies to govern how AI should be used will be important as technology advances, the authors add.

"The AI age is going to be unsettling, transformative and revolutionary,"

Siau writes in another recent article ("How Will Technology Shape Learning?" published in the March 2018 issue of the *Global Analyst*). But in this unsettling environment, higher education can play a significant role.

"Higher education must rise to the challenge to prepare students for the AI revolution and enable students to successfully surf in the AI age," Siau writes.

Provided by Missouri University of Science and Technology

Citation: After Uber, Tesla incidents, can artificial intelligence be trusted? (2018, April 10) retrieved 27 April 2024 from

<https://phys.org/news/2018-04-uber-tesla-incidents-artificial-intelligence.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.