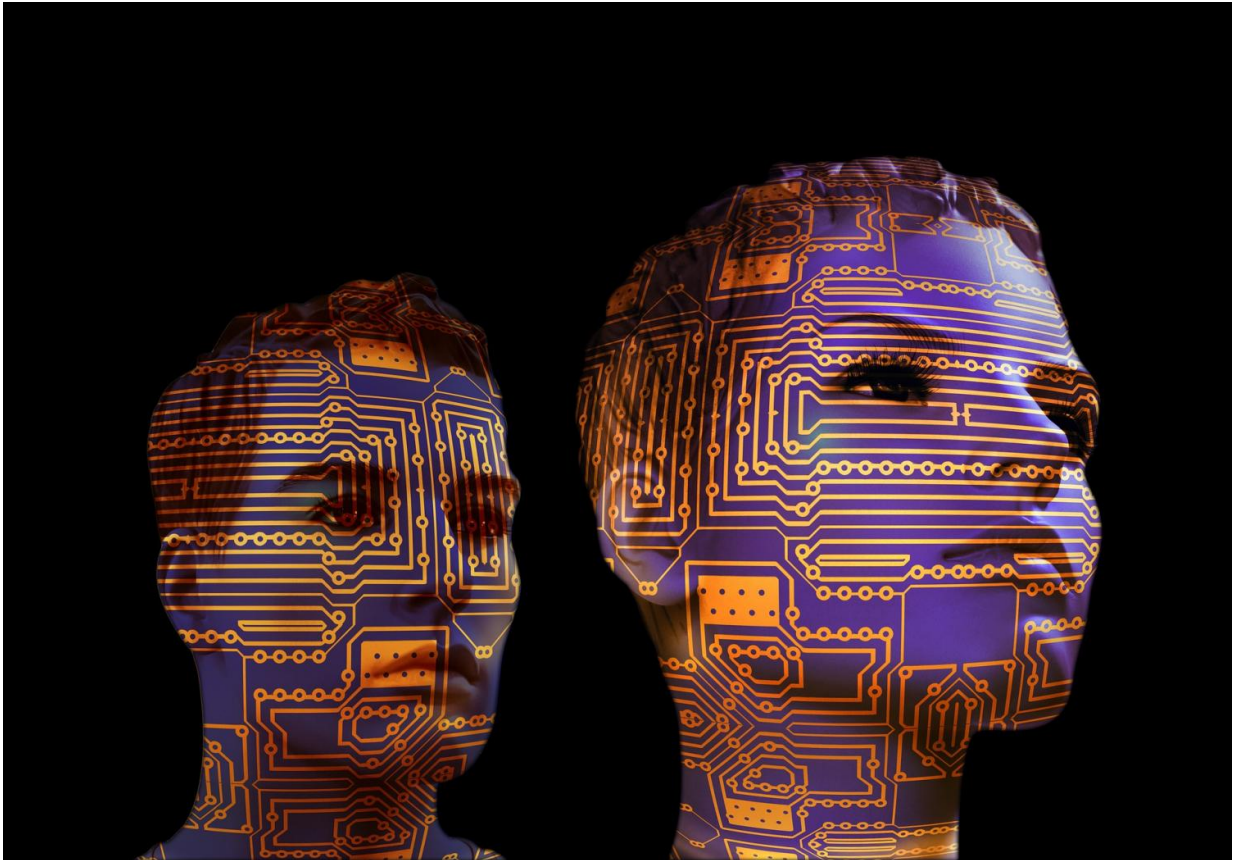


# How to hold AI accountable for its actions

June 1 2017, by Bob Yirka

---



Credit: CC0 Public Domain

(Tech Xplore)—A trio of academics has written and published a Focus piece in the journal *Science Robotics* on the need to hold artificial intelligence systems accountable for their actions. Sandra Wachter, Brent Mittelstadt and Luciano Floridi, all with the Oxford Internet Institute,

suggest that there are three main difficulties standing in the way of such action, but they should not be enough to discourage implementation of strategies meant to protect from actions taken by machines.

Science fiction movies offer a clear view of what robots gone rogue might look like, but less has been demonstrated regarding the underlying technology—[artificial intelligence](#). The term is thrown around loosely, but it covers a lot of ground. When you apply for a credit card, for example, an AI system makes the [decision](#), or when you sit idly in the back seat of a self-driving car, an AI system ferries you to your destination. AI will be responsible for making decisions such as whether to run over a pedestrian or smash into one of the cars on either side to avoid accidents. As humans, we want accountability—if a Google car crashes sometime soon, we will all likely blame Google. But what about in the more distant future, when there are two such vehicles from different companies crashing? How could an insurance agent be expected to decipher the technical jargon likely to arise during a trial if humans have been harmed? These are the sorts of issues the authors consider.

In their article, the authors suggest there are three main difficulties standing in the way of making sure machines are held to the same standards as the humans that create them. The first is the diverse nature of such machines. They note a distinction needs to be made between robots and AI systems; the former is hardware and driven by software, the latter is a system capable of making decisions based on what it has learned. But they also point out that AI systems are so different from one another it seems almost impossible to come up with a system of accountability that applies to all of them.

Another problem is transparency. When encountering an AI system, wouldn't it be nice to know how it worked? How it made its decisions? We all know that software is involved, that routines running algorithms

are grinding away and that data is stored and retrieved—that somehow the cloud is likely involved. But how do we know what a given system is actually doing? One that might be making a decision that could very seriously impact our life? That, the authors note, is going to be difficult.

The third problem is construction. AI comes in many forms, as an unseen system behind the robot voice that answers when you call for support, for example, or as a robot that might be watching and judging your behavior, or the simple but annoying spell-check on your phone that minces your meaning.

Despite these difficulties, the researchers suggest that solutions are possible. They highlight how different countries are addressing the problem, noting attempts at dealing with such issues such as passing legislation. We may not have the answers just yet, but they argue that it should not stop us from finding them and then taking action.

**More information:** Transparent, explainable, and accountable AI for robotics, *Science Robotics* 31 May 2017: Vol. 2, Issue 6, ean6080 , [DOI: 10.1126/scirobotics.aan6080](https://doi.org/10.1126/scirobotics.aan6080) , <http://robotics.sciencemag.org/content/2/6/ean6080>

## Abstract

To create fair and accountable AI and robotics, we need precise regulation and better methods to certify, explain, and audit inscrutable systems.

© 2017 Tech Xplore

Citation: How to hold AI accountable for its actions (2017, June 1) retrieved 2 May 2024 from <https://techxplore.com/news/2017-06-ai-accountable-actions.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.