

AI and Trust

Bruce Schneier

Belfer Center for Science and International Affairs

Harvard Kennedy School (Nov. 2023)

Extensive comments on this essay are found at:

<https://www.schneier.com/blog/archives/2023/12/ai-and-trust.html>

Copyright 2023. All rights reserved.

Reprinted with permission.





Adobe Stock/wei

ANALYSIS & OPINIONS – *Belfer Center for Science and International Affairs, Harvard Kennedy School*

AI and Trust

Author: **Bruce Schneier** | Nov. 27, 2023

*This essay was originally presented at an **AI Cyber Lunch** on September 20, 2023, at Harvard Kennedy School.*

I trusted a lot today. I trusted my phone to wake me on time. I trusted Uber to arrange a taxi for me, and the driver to get me to the airport safely. I trusted thousands of other drivers on the road not to ram my car on the way. At the airport, I trusted ticket agents and maintenance engineers and everyone else who keeps airlines operating. And the pilot of the plane I flew. And thousands of other people at the airport and on the plane, any of which could have attacked me. And all the people that prepared and served my breakfast, and the entire food supply chain—any of them could have poisoned me. When I landed here, I trusted thousands more people: at the airport, on the road, in this building, in this room. And that was all before 10:30 this morning.

Trust is essential to society. Humans as a species are trusting. We are all sitting here, mostly strangers, confident that nobody will attack us. If we were a roomful of chimpanzees, this would be impossible. We trust many thousands of times a day. Society can't function without it. And that we don't even think about it is a measure of how well it all works.

In this talk, I am going to make several arguments. One, that there are two different kinds of trust—interpersonal trust and social trust—and that we regularly confuse them. Two, that the confusion will increase with artificial intelligence. We will make a fundamental category error. We will think of AIs as friends when they're really just services. Three, that the corporations controlling AI systems will take advantage of our confusion to take advantage of us. They will not be trustworthy. And four, that it is the role of government to create trust in society. And therefore, it is their role to create an environment for trustworthy AI. And that means regulation. Not regulating AI, but regulating the organizations that control and use AI.

Okay, so let's back up and take that all a lot slower. Trust is a complicated concept, and the word is overloaded with many meanings. There's personal and intimate trust. When we say that we trust a friend, it is less about their specific actions and more about them as a person. It's a general reliance that they will behave in a trustworthy manner. We trust their intentions, and know that those intentions will inform their actions. Let's call this "interpersonal trust."

