DM22-0418

Script: Trust and AI Systems
SME(s): *Carol Smith and Dustin Updyke*
Interviewer/Facilitator: *Dustin Updyke*
Interview Conducted: *Wednesday, March 4 at 3 p.m. ET*

<Canned Intro>

Dustin Updyke: Welcome to the SEI Podcast Series. My name is Dustin Updyke, and I am a senior cybersecurity engineer in the SEI's CERT Division.

The concept of trust is at the forefront of AI system concerns. In 2021, the National Institute of Standards and Technology, whose research often helps shape government policy, released an approach on how organizations can identify and manage bias in AI. NIST has stated that "Alongside research toward building trustworthy systems, understanding user trust in AI will be necessary in order to achieve the benefits and minimize the risks of this new technology."

Joining me today to talk about trust in AI systems s Carol Smith, a senior research scientist in Human Machine Interaction in the SEI's AI Division.
Welcome, Carol.

Carol: *Thank you.*

1. **Dustin**: Let's start by telling our audience about ourselves, what brought each of us to the SEI, and the work that we do here. Carol, you can start.
   Carol/Dustin:

2. **Dustin:** First I think we should clarify what we mean when we use the term "trust" in the context of artificial intelligence and autonomy. I also think for our audience we should define the scope of the discussion: Are we talking about human trust in AI systems, how such systems would trust one another, how to build trustworthy systems, or all or any of the above?
   Carol/Dustin:

3. **Dustin:** There have been many reported examples of bias and mistakes in AI systems such as facial recognition technologies to automated predictive policing, and a relevant question that seems to underlie these discussions is often, "should we actually trust these AI systems?"

4. **Dustin**: One complexity that I am exploring in my graduate work is that we, as humans, have historically thought about or used about computers as tools. As a result, it was easier to trust them because of their repeated success at

mechanical tasks. AI upends this long-held assumption because humans are now partnering with machines on an increasing number of tasks. How should our perception of trust shift in response to this new reality?
Carol/Dustin:

5. **Dustin:**

   Carol/Dustin:

6. **Dustin:** What are the roadblocks we face in achieving increased human trust in AI systems and what are possible pathways forward for addressing them?
   Carol/Dustin:

7. **Dustin:** One aspect of our work that we like to highlight in our podcasts is transition. If I am working with AI systems (and so many of us are) how can I ensure that it will behave in the intended manner? What resources are available to me in terms of documentation and guidance?
   Carol/Dustin:

8. **Dustin:** Sometimes it seems there are innumerable trust issues within computing in general, and AI likely exacerbates that. Some of these are rooted in technology

and others are social. Maybe because it's a mix of things, real progress seems uniquely daunting. Are there fundamentals or recent progress that you are optimistic about in this space?
Carol/Dustin:

Dustin: Thank you for talking with us today. We will include links in the transcript to resources mentioned during this podcast.

Finally, a reminder to our audience that our podcasts are available on Soundcloud, Stitcher, Apple Podcasts, and Google Podcasts as well as the SEI's YouTube Channel. If you like what you see and hear today, give us a thumbs up.

Thanks again for joining us.

<Canned Outro>