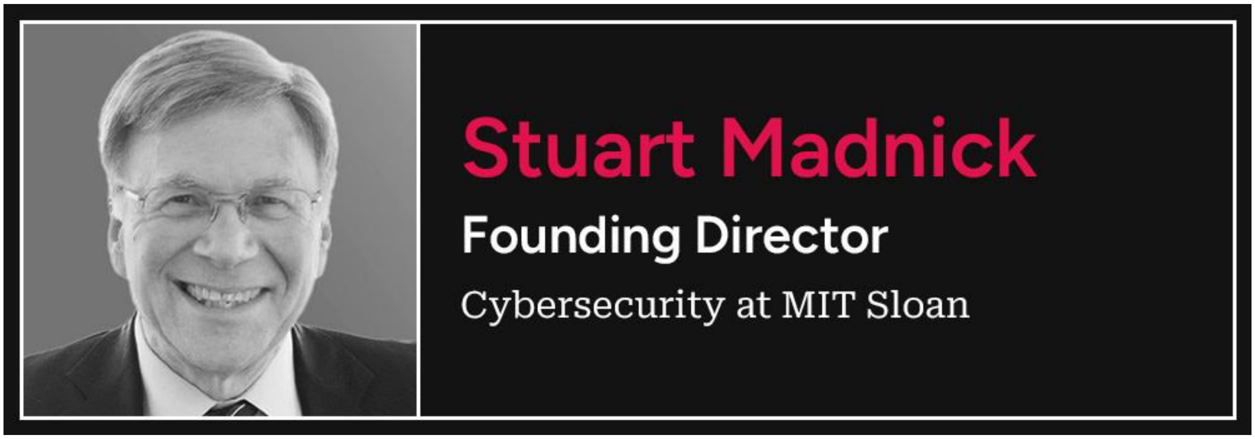
From [AI and Cybersecurity: The New “Arms Race” | MIT Horizon]()

AI and Cybersecurity: The New “Arms Race”

A conversation with MIT Sloan’s Stuart Madnick

10 min

- AI, Cybersecurity, Generative AI



Stuart Madnick is the founding director of the Cybersecurity at MIT Sloan research group. He is a longtime expert in cybersecurity, coauthoring his first book on the subject in 1979. Madnick sees AI’s growing use in cybersecurity as a new kind of arms race—one cyberattackers are currently winning. MIT Horizon spoke with Madnick in January 2025 about the impact of AI on cybersecurity and building resilience to new cyber threats. Our conversation has been edited for length and clarity.

What do you see as the biggest future trends in cybersecurity?

When you talk about innovation in IT, it’s hard to say that without using the letters A and I very quickly. There are lots of things I can say about AI and lots of things I can say about cybersecurity, but I'm going to look at the intersection of those two things. I divide it into two bins: the good and the bad.

On the good side: One of the big things AI is being talked about is supporting or replacing knowledge workers. For example, for doctors AI could be a doctor assistant or possibly a doctor replacement. Doctors deal with sick people. Cybersecurity workers deal with sick computers, so there’s the same kind of logic. A lot of the things that human cybersecurity experts do are things that an AI system could do possibly as well, maybe even better. This is important because there has been increasing concern about the shortage of cybersecurity personnel.

Has there been a lot of progress in that? I wouldn’t say a lot so far. But the idea of supplementing the cybersecurity workforce with AI robots, if you will, aiding them or replacing them, is obviously one of the positive results one can see in cybersecurity. One of the things experts don't do a great job at, and that AI could do even better, is vulnerability detection—realizing that a piece of software has a flaw in it that has escaped attention. AI is being used both to write and evaluate code. It's quite possible that AI systems could find vulnerabilities in software that humans haven't yet found so they can be repaired.

One area of using AI to improve cybersecurity with the most research is called anomaly detection. That is, how do you identify that something suspicious is going on? It takes time to check each and every anomaly. The problem is not that AI can't do anomaly detection, but can you get it focused enough and high enough quality that it isn't coming up with too many anomalies so you’re not spreading your resources too widely? Some anomaly detection systems today are setting off thousands of alarms every day. It is impossible to investigate every case. That is something where there is hope for AI. I'm not as optimistic as some are, but obviously a lot of research is going on to improve AI ability to more accurately detect such problems.

“ I often refer to this as an arms race, because both the attacker and defender are always seeking an advantage, much like the arms race in warfare. Furthermore, often a particular weapon can be used for offense or for defense. ”

—Stuart Madnick

Those are examples where AI is being looked at as a very valuable weapon. I often refer to this as an arms race, because both the attacker and defender are always seeking an advantage, much like the arms race in warfare. Furthermore, often a particular weapon can be used for offense or for defense. For example, the vulnerability detection I just mentioned is a great defensive weapon. It is also a great offensive weapon. From everything I'm seeing, the bad guys are much more aggressive in using it to try to detect vulnerabilities than the good guys are in trying to detect and fix them in time.

On the bad side: You probably are familiar with spam. Almost everyone gets a mailbox full of it. Are you familiar with the term spear phishing?

Yes. I’m not an expert on it, though.

Normal spam says “Hello, dear occupant” or something to that effect. Spear phishing is an email apparently from your boss that says, “It was great to talk with you last Wednesday and have lunch over at the Cascade restaurant. We should do that more often, but I’ve got an important request to make.” This sounds like something coming from your boss, who you did have lunch with last week, so you are likely to follow whatever orders you’re given.

For spear phishing to work, the bad guys have to learn a lot about you. Typically, they tap into your company's email, and they know who you converse with and how you converse with them. They can mimic people you normally interact with. For a human to do that takes a lot of work. You have to get into their computer to begin with. You have to read all the traffic going on. You have to analyze all that traffic and come up with a good story. AI can do that much faster and make much higher quality, easier-to-produce spear phishing.

You may have heard of the cyberattack on the MGM hotels in Vegas [in 2023]. Supposedly, one of the key elements of it was a phone call to the help desk of the IT group at MGM. It was from one of the lead IT people saying, "I'm sorry, my iPhone is broken, but there's something important I have to fix in the computer, so please override the multi-factor identification and let me in." And sure enough, the help desk helped him. He was a phony person, but the voice, the mannerisms, clearly, were convincing.

How do you ever know who you are talking to, whether it be over the phone, by email, or by a Zoom link? The AI systems, to some extent, know more about you than you do. If someone asked you, "What was the name of your sixth-grade math teacher?" Do you remember it? The AI system probably found that out. If you were on one screen and the AI on another screen, the odds are people would more likely believe the AI system.

You've listed all these potential threats from AI. Do you think organizations are prepared for those kinds of threats? And if not, then why not, and what can they do to prepare?

The issue is the good guys are working hard, but the bad guys are working a lot harder and a lot faster—partly because they have a lot to gain. There is this tug-of-war, or arms race as I mentioned, going on. Not all companies either are aware enough or moving fast enough. A lot of companies say, "Well, yes, other companies have had cybersecurity problems, but we've been OK so far, so things look like we are OK." I don't want to overgeneralize, but unfortunately, there's a lot of that syndrome out there. Until you've been really hit hard, it's easy to ignore these things, because people don't like to deal with unpleasant things.

The question people often ask me is: What do you do about it? These things are coming at us from every direction under the sun. Some of these tricks are very compelling, and it's easy to click on the link and at least go halfway down the rabbit hole. What I tell people is: Yes, be as alert, as careful as you can be, but you really need to change your mindset from being a matter of if you will suffer a cyberattack to when you will suffer a cyberattack. That's why we do a lot of research on cyber resilience. I've used the analogy that if you have an office building, you don't want to have just a single key to the outside of the building, and once someone gets in the building, everything, including the safe, is wide open. That's the way a lot of our computers are built. It's all or nothing. If you make it through the front gate, anything you want you can get.

" What I tell people is: Yes, be as alert, as careful as you can be, but you really need to change your mindset from being a matter of if you will suffer a cyberattack to when you will suffer a cyberattack. "

—Stuart Madnick

We want to minimize the amount of damage that a cyber attacker can do if—or when—the defenses are breached. People realize, yes, you want to be as secure as possible, but you also want to be as resilient as possible. That is the wake-up call just beginning to sound.

I want to go back to something you mentioned when talking about using AI to help improve cybersecurity. You mentioned that you weren't as optimistic as some others are. Why is that?

I don't claim to be an AI expert, although I did take a class from Marvin Minsky, one of the AI pioneers, when I was a student. I've been in a number of meetings with other faculty where this issue has come up and I see widely divergent views. I've lived through at least two prior AI winters, so I tend to be a bit more cautious. I will say I am much more impressed with the latest round than I have been in the previous rounds.

But they've been working on this issue of anomaly detection for well over a decade. Although you'll see very fancy reports, if you read into most of them, they are typically fabricated data with fabricated AI systems, and they're doing fabricated analysis on it. I haven't seen any example actually in use, in practice, in real companies, with quantified real results. There's a tendency to overhype.

Getting back to AI doing vulnerability detection—that I can believe. Because there are certain patterns of things people do in software that are not safe. It's just very time-consuming for you and me to go through huge programs line by line. Having an automated system that can do that makes a lot of sense. Likewise, supplementing or aiding humans in doing whatever they're doing, that I can believe. The question is: How much and how fast?

What other cybersecurity threats do you see emerging?

Let me give you a non-MIT—by non-MIT, I mean nonscientific—piece of data. In the past year, I've given four presentations at different conferences around the world, medium-sized, anywhere from 50 to 200 people. These are all people in the cybersecurity field. I've asked the audiences the following question: Ten years from now, do you think the cybersecurity situation will be much better than now, about the same, or worse? Across those maybe 400 or 500 people in total: 1% said it will be better, about 9% said probably about the same. If you do the math, about 90% said it will be worse than it is today. The feeling is it will get worse long before it gets better.

So they’re thinking it will get much worse. Do you see more threats emerging from other types of technologies?

One thing that our research, and that of others, has shown is that cyberattacks are most likely to happen when there’s a change: technological or organizational. One of the examples we use was the cyberattack on TJX—TJ Maxx, HomeGoods, Marshalls [in 2007]. They were one of the first retail stores to install Wi-Fi. They didn't realize that early Wi-Fi had vulnerabilities in it. Whenever we get a new technology, it has a lot of shiny things we look forward to, but almost always there's a cloud there somewhere, and it takes us a while to look for it or recognize those clouds.

“ Whenever we get a new technology, it has a lot of shiny things we look forward to, but almost always there's a cloud there somewhere, and it takes us a while to look for it or recognize those clouds. ”

—Stuart Madnick

Maybe 20 years ago, I was on sabbatical at the University of Nice in France—which, by the way, is very nice—and the IT group I was affiliated with there was adjacent to their automotive telemetrics research group. This was before autonomous vehicles. They were mostly focused on putting in better emission controls, all kinds of stuff we've seen over the following decades. The thing I observed at that time, they would say things like, “Well, we're trying to put computer systems into a very harsh environment—the automobile—and we've got to make them very small. We've got to make them very cheap. We've got a lot of constraints. These are really hard problems. Once we solve all those problems, we'll come around to looking at the issues involving cybersecurity.”

Well, about two years ago, I had meetings in Singapore. They arranged to have me meet various research groups, including a group that was working on autonomous vehicles. And I had pretty much the exact same conversation I had 20 years ago. “What we’re trying to do is really, really hard. Cybersecurity we agree is important. It's number n plus one on our list of n important things to get to, and once we’ve got those other n solved we will get to that one.” I've heard that same thing in many different ways.

Now, more companies are saying, “We want to see cybersecurity incorporated into your product from the start”—called secure by design. That’s still the exception, not the rule, but some government agencies are now making that a mandate for products that they buy. How well they actually enforce the mandate, I don't know. The trouble is that cybersecurity, although we all acknowledge its importance, when it comes to priorities, it isn't always priority No. 1. But we see evidence that it is rising in importance, which is one of the missions of our Cybersecurity at MIT Sloan initiative and our corporate partners. All of us need to work together to improve our individual, our corporate, and our society's cybersecurity.

Learn more

- The future of cybersecurityLink opens in new tab
- Hacking humans: How AI can enhance phishing attacksLink opens in new tab

Massachusetts Institute of Technology
Cambridge, MA 02139