STATE AFFAIRS
WISPOLITICS
Wisconsin's Premier Political News Service

UW expert warns of AI cyber threats, urges skepticism about defenses

Posted on Tuesday, Aug 22, 2023

A computer science professor at UW-Madison says certain AI tools pose a cybersecurity threat when in the wrong hands, and urges skepticism about potential defenses.

In a presentation held as part of Madison’s Forward Festival, Prof. Somesh Jha yesterday discussed how experts in the field are looking to preempt the use of these tools by hackers and other adversaries. Some AI programs are likely already being used in this way, he said, pointing to large language models like ChatGPT that can create sophisticated written content.

“But the problem is that we don’t have a good way of understanding where is that threat landscape right now,” he said. “If you talk to people who are in this threat intelligence community, they’ll tell you that, yeah you’re seeing already evidence that they have started using it, quite heavily.”

Emerging AI systems raise the issue of the “dual-use dilemma,” Jha explained, noting many scientific and technical advances have had both positive and negative implications for the world. As a historical example, the process for mass-producing ammonia being discovered led to the use of modern fertilizers for farming, as well as the creation of chemical weapons in World War I.

In the modern era, cryptography can be used to enable secure messaging and improve data security, Jha noted. But on the other hand, those same techniques are being used in ransomware attacks by hackers holding valuable information hostage.

He also referenced the use of AI software in creating convincing “deepfake” images and videos from a simple prompt, which could be used to spread misinformation.

And while some solutions are being developed for identifying AI-produced content, Jha argued these measures have their drawbacks. They include statistical programs for analyzing written text for abnormalities, as well as the practice of watermarking, which could allow content to be independently verified using a “secret key” code.

Some of the top U.S. tech companies including Google and Microsoft recently pledged to develop mechanisms for ensuring AI-generated content can be identified as such, the White House announced last month.

But Jha said “we have to be very careful” about trusting defenses like these.

“This is like a cat-and-mouse game,” he said. “Somebody puts it out there, some attackers says, ‘No, I can jiggle my image a little bit so that your deepfake detector doesn’t detect it.’ … We have to be very skeptical.”

Watch the video:
https://wiseye.org/2023/08/21/forward-fest-madai-and-large-language-models/

See the White House AI announcement here:
https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/

–By Alex Moe