AI Could Take Hacking to New Levels

In future elections and society in general, will you be able to believe your own eyes?
Kevin McCaney
Thu, 03/01/2018 - 13:42

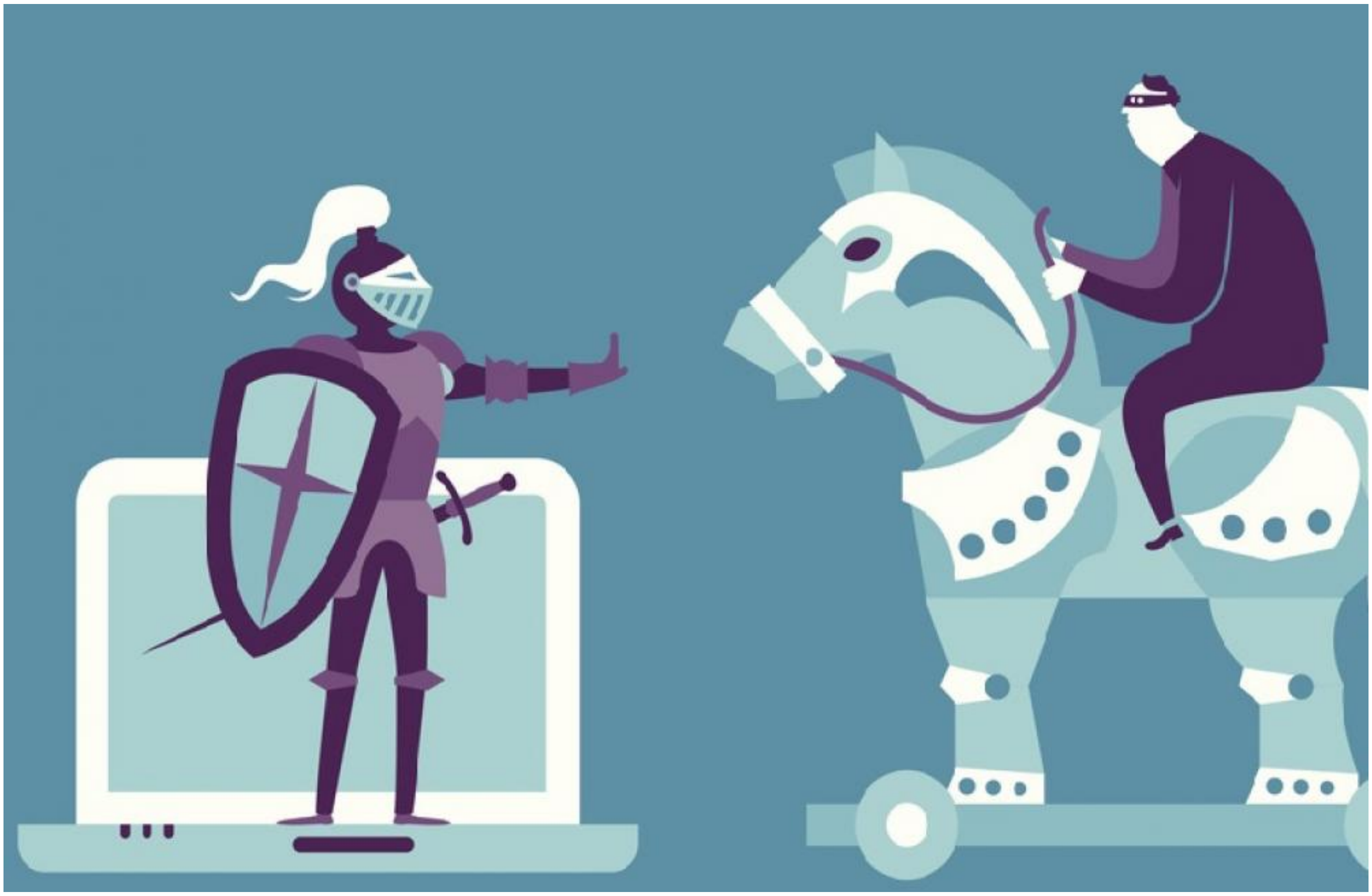

Illustration: akindo/iStock

If you thought Russian trolls were bad, wait until artificial intelligence is unleashed in the near future, whipping up faux “news” stories, comments, posts, emails, audio files, photos and videos of questionable pedigree, but with a quality equal to that produced by humans, all in the blink of an eye and around the clock. In future elections and society in general, will you be able to believe your own eyes?

That’s the warning — perhaps taken to extremes — contained in a new report on a study led by Oxford and Cambridge universities and authored by 26 leading AI researchers and technologists. The report, "The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation," released Feb. 21, goes into the ways artificial intelligence can supercharge hacking, describing how advancements AI, machine learning other software will set off a high-level “cat-and-mouse game” between attackers and defenders, with public opinion and possible political consequences in the balance.

The report, which also takes a broader look into the realms of digital, physical and political security, lauds the beneficial AI’s advancements in areas ranging from health care and scientific research to cybersecurity and driverless cars. But it points out those same advancements, like those in all technology, come with a double edge and can easily be turned against the greater good.

A Picture is Worth a Thousand Lies

One example is with image recognition, which now hovers at about 98 percent accuracy, but has led to similar improvements in image generation. The technology used in surveillance or weapons systems can be turned around to manipulate or create images as well.

“AI systems can now produce synthetic images that are nearly indistinguishable from photographs, whereas only a few years ago the images they produced were crude and obviously unrealistic,” according to the report.

New developments apply in other media as well, creating “the ability to generate synthetic images, text, and audio could be used to impersonate others online, or to sway public opinion by distributing AI-generated content through social media channels.” Video, too, can be manipulated to create synthetic but realistic-looking footage of, say, a candidate making an incendiary comment he or she didn’t make. AI bots also could spread propaganda via video chats or other means. To add to the confusion, the possibility of faked videos could lend a level of plausible deniability to actual footage of embarrassing or repellent comments.

The authors don’t want to go overboard with the-sky-is-falling scenarios. They point out AI systems generally have a narrow focus, excelling at one particular task at a time. And as impressive as AI has been, areas such as robotics haven’t leaped forward much over the past decade, so the authors don’t expect truly revolutionary breakthroughs to crop up tomorrow.

Nevertheless, AI technology is steadily advancing, and great leaps forward are likely at some point. Once a machine has gotten up to human levels in a game like chess, for instance, it has quickly exceeded the performance of the best humans. Machines are also getting passed games like chess and starting to excel at multifaceted games like Atari's old but difficult Montezuma's Revenge. The report cites a survey of AI researchers in which almost all of them said AI will eventually exceed human-level performance in the tasks surveyed, most likely within the next 50 years.

So what can we do about it?

Fight Fire with Fire

AI-fueled attacks of one kind or another will grow in number because of the decreasing costs and increasing availability of the tools, the report says. New threats humans couldn't pull off are also likely, as is their effectiveness, since AI can enable highly targeted attacks that will be difficult to attribute to a source and capable of getting around defenses, even in AI systems.

The report's authors make four high-level recommendations, essentially advocating for "baking in" responsibility. Policymakers need to confer with researchers on new developments in order to find ways to mitigate malicious use. AI developers need to take AI's dual-use power seriously, working to fend off predictable misuses. Best practices should be identified and followed. And the range of stakeholders should be expanded.

But organizations also can fight AI with AI, using the technology's ability to automate processes and learn from examples on the fly to identify vulnerabilities and malicious uses, and anticipate attacks in order to counter them. Defensive technologies and techniques need to keep pace with offensive methods.

Meanwhile, it has been rumored from time to time that you can't believe everything you see on the internet. People in general might want to keep that in mind.

View printer friendly version
Federal Cybersecurity
cyber
Artificial Intelligence
Eye on AI