WHY AI-BASED CYBERSECURITY WILL CONTINUE TO NEED THE HUMAN TOUCH

DAVID HECHLER

It strikes me as almost a foregone conclusion that artificial intelligence will transform cybersecurity. But it's far less clear, at least to me, whether the result will be a standoff between enemy forces that rely almost entirely on AI defenses.

It seems inevitable that there will be an AI arms race. There already is. The United States and China are the competitors mentioned most prominently in the media. Russia, North Korea and Iran are the other nation-states active in launching cyberattacks. They'll try to match the advances of their targets. Other countries could emerge in coming decades.

It's easy to argue that AI will figure into the equation more and more prominently—on both offense and defense. But that doesn't mean that the machines will be in control. AI will not be calling all the shots. At least not in the foreseeable future. Much about the way the competition evolves will depend upon the humans who collaborate with the technology. Just as it does when AI is used by the military (as I will discuss below).

A lot of the talk right now is about the astonishing technological advances. When the conversation turns to people, they are often engineers who are building the software, and leaders of companies that are funding it—and pushing the competition. These individuals are certainly enjoying a well-deserved moment. But they aren't the only ones who are important players in this realm.

Lawyers, philosophers, journalists, researchers and all kinds of academics have expressed concern about the dangers

AI may pose not only to our country, but to humanity. Far from being seen as our protector against cyberattacks, some people view AI as a grave threat to our future.

A widely cited **survey** produced by AI Impacts in 2022 asked researchers who had published papers presented at two large machine-learning conferences this question: "What probability do you put on future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species?" Based on 738 responses, the median respondent said the chance was 5%. But the number that many news accounts cited was double that number because 48% of respondents said the chance was 10%, and that's the statistic almost everyone used.

"AI systems can neither design themselves nor clean their own data, which leads us to conclude that increased reliance on AI will make human skills even more important..."

"Would you work on a technology you thought had a 10% chance of wiping out humanity?" New York Times columnist Ezra Klein **wrote** in March 2023. Klein explained his deep concerns while acknowledging that the train has already left the station. And the challenge of slowing, much less stopping, its progress seems daunting at best. As apprehensions about ChatGPT have mounted, a chorus of voices joined his.

It's possible that politicians may try to gain some measure of control through legislation. But even if they were convinced of the need, the likelihood of success seems highly problematic. The work is in the private sector, and the funding is from companies like Microsoft, Google and Facebook. So government doesn't control all the purse strings. And if the government tries to create legal roadblocks, critics will almost certainly accuse it of handing China a devastating, and potentially deadly, gift.

But let's return to cybersecurity, where the aim is to use AI to safeguard our safety. The machine learning will need to be directed by humans who study the threats and feed relevant information into the technology. In my research, the article I came across that shed the most light on this subject was **Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War**, by Avi Goldfarb and Jon R. Lindsay (this is where my earlier reference to the military comes in). Writing in the journal International Security, the authors did touch on cybersecurity and cyberwar, but that's not why I found it relevant. When we're talking about cybersecurity in the broadest sense—including battles between nation-states—then war is more than an analogy.

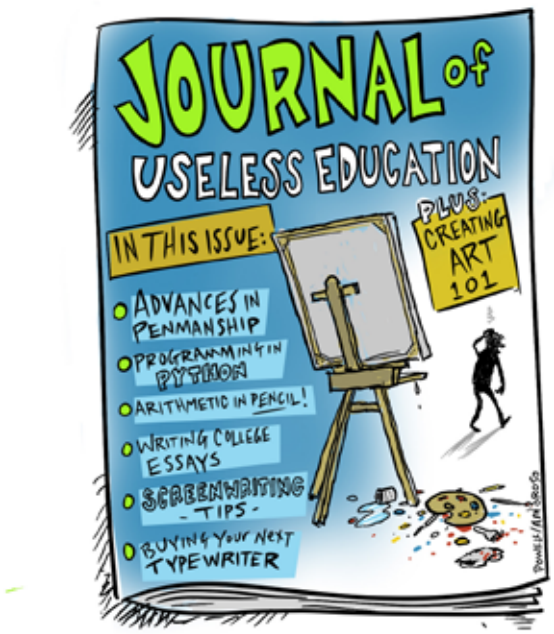
Goldfarb and Lindsay don't address the cybersecurity challenges we're addressing here, but they do talk about the ways corporations and even doctors use AI. The authors see great value in the technology. They expect it to transform the world in which we live. But they don't see it substituting for humans. They anticipate a collaborative relationship that builds on the strengths of each. "A well-specified AI utility function has two characteristics," they write. "First, goals are clearly defined in advance. If designers cannot formally specify payoffs and priorities for all situations, then each prediction will require a customized judgment. This is often the case in medical applications. When there are many possible situations, human judgment is often needed upon seeing the diagnosis. The judgment cannot be determined in advance because it would take too much time to specify all possible contingencies. Such dynamic or nuanced situations require, in effect, incomplete contracts that leave out complex, situation-specific details to be negotiated later."

The authors go on: "AI adoption may radically change the distribution of judgment by altering who in an organization makes decisions and about what, but in all cases, humans are ultimately responsible for setting objectives, making trade-offs, and evaluating outcomes.... AI systems can neither design themselves nor clean their own data, which leads us to conclude that increased reliance on AI will make human skills even more important..."

There's another important factor concerning cybersecurity based on AI. The debate over ChatGPT may not involve the government, but the government is very much involved in the world of cybersecurity. And it will inevitably be deeply involved in budgetary and strategic decisions that involve AI. When Goldfarb and Lindsay write that "seemingly trivial procedures can become politicized when budgets and authorities are implicated," it's easy to see how this applies to cybersecurity. "Even in the absence of parochialism," they continue, "the complexity of administrative systems introduces interpretive challenges for personnel."

In the case of cybersecurity, there's plenty of personnel. The Cybersecurity and Infrastructure Security Agency, the National Security Agency and the Department of Justice all play important roles. The heads of those organizations and other appointed cybersecurity leaders don't report to AI. And their judgments affect how AI is deployed. When it comes time for lobbyists and government agencies to press representatives in the House and Senate to approve appropriations for cybersecurity tentatively slated to be included in the annual National Defense Authorization Act, they aren't likely to be glad-handed by ChatGPT.

Finally, let's not forget that the political winds in the United States have been shifting from administration to administration. There are no guarantees that new leaders will continue to support AI or a robust cybersecurity budget. A new administration's strategy could certainly change course. And the same could be true in other parts of the world. As hard as it is to predict the advances of the technology, it can be just as challenging to gauge the path that politics will take.



(Reprinted from the TAG Cyber Security Annual, 2nd Quarter 2023.)