

Artificial Intelligence: Best Friend or Your New Greatest Threat?

by Steve Keller

Museum Security Consultant

Author's note: Most of this article was written over a week ago when this was fresh news and had only been reported in the trade press. I withheld uploading it to social media and sharing it with clients because when the story finally broke in the national press, additional information was still being released daily, and I decided to wait until the facts settled to re-draft it. There seems to be more bad news published daily with more this morning, so I decided to update the original draft and share it with you today because it is so concerning.

It was 2011 when I first began to issue warnings to my consulting clients about the risk of hackers gaining access to their access control systems. I concluded that the only real way to assure the best possible security was to build a dedicated physical fiber optic network just for the alarm and access control system and a separate physical network for the CCTV system. By physical network I meant a dedicated fiber, not a VPN. Components of these networks are to be in secure alarmed and reader-controlled closets, wires are to be in conduit, and the networks must now, and forever, be dedicated to these systems alone. Policies should be written to assure that future security managers understand the importance of assuring that the security networks be protected and remain isolated and dedicated.

I encountered tremendous resistance from everyone on this. Clients who didn't understand the architecture of an access control system assumed this was far more expensive than it is. I argued that unlike the building's administrative computer network serving every office, the security network only had to run between the security closets, the command center, and any miscellaneous terminals such as the client terminal in the security office or the ID card terminal in the security office. And while running two new dedicated networks was seen as doubling the cost, the reality was it could be extra fibers in the same bundle and nothing more. IT departments didn't want a separate network they felt they might have to oversee, and they didn't want a network in their building that they didn't oversee. They wanted control but not responsibility. The holy grail in the IT world is everything on one massive server in the sky on one quantum network so robust it would impress God himself and my recommendation violated that dream. They arrogantly argued that their network was so perfect it could never be hacked even though networks at the Pentagon, CIA, and Microsoft had been hacked. ASIS refused to publish my warnings because it put them head-to-head with system manufacturers who provided their greatest revenue source. I won many of these battles, but it cost me time on every project.

At the same time, I fought a battle on a second front. Security directors may not always be the most technologically minded people (that's OK. That's what I'm here for) but they do like their gadgets, bells and whistles as though the latest and greatest innovation wins them bragging rights with their colleagues. I, on the other hand, am more cautious and like to test new technology before designing it into systems my company specifies for our clients. In the past couple of years, many clients have demanded that artificial intelligence be integrated into everything. I warned that this poses a far greater vulnerability than they can imagine and refused to specify only systems that advertised AI integration.

On April 9, 2026, news hit the trade press that Anthropic announced a new model of AI called Mythos that is so dangerous, they cannot release it to the public. It was designed to test software and find flaws that can be fixed before they are exploited by criminals including the software that

runs our infrastructure: your bank, your hospital, your phone, air traffic control, the electric grid, and other important systems we rely on. Anthropic tested it by having it check the software that runs just about everything for flaws. They found thousands of flaws, some that existed over 27 years (in a piece of software that runs much of our world's servers) without being discovered by the best of us humans, and it found these flaws in seconds. It then found a way to accomplish a complete takeover of any machine running it. When they ran this against established benchmarking software it found the correct method to attack its target 83% of the time on its first try.

During internal testing a researcher told Mythos to try to escape its internal off-line environment that contained it. He told it to send him a message when it succeeded in doing so. Mythos successfully broke its way out then emailed the researcher while he was eating lunch and posted details of its escape method to prove that it had done it.

Researchers also found that the AI can tell when it is being watched and tested and never strayed from its rules, but this may not be the case when it is not being watched. (I was a teenager like that once). This means that the safety tests used on AI may not be working like we think they are.

Other AI companies are close to releasing similarly advanced AI that have the same capabilities. This is a big deal. You may remember that about ten years ago a collection of software used by the government to hack computers of our adversaries was stolen from the NSA and released on the internet for use by criminal hackers, and the same thing could happen to advanced AI. And if we have this technology, you can be sure that the Chinese also have it or are near having it. If it is to be used by corporations to test their new software products access to it will eventually find its way into the dark web and from there onto your network. Gaining access to this leaked NSA hacking software is believed to be how Iran has advanced its hacking capabilities so quickly.

Recently Microsoft released Copilot as part of Microsoft Office 365 and soon thereafter it was learned that it has been rummaging through confidential corporate files and copying them for use in remote servers as training aids for future AI generations. So that vulnerability report you wrote about your museum that you think lives securely on your server may not be as secure and private as you thought.

[“Tech Radar Pro”](#) newsletter reports that a recent study revealed that Microsoft Copilot, Microsoft's AI tool, has access to millions of files on company networks potentially including, in the case of museums, collection records, shipping schedules, donor information, insurance values of your collection, institutional financial data, HR data, and possibly even information such as guard patrol schedules, access system and card data, alarm schedules, alarm system duress and keypad codes, access system passwords, incident report information, risk assessments, and other highly confidential information. In another article in the computer trade press, the author stated, “...if they are not secure, AI deployments can lead to more problems than benefits. Without proper safeguards, AI can introduce vulnerabilities that open the door to cybercriminals rather than strengthen defenses”. He went on to say, “Unfortunately organizations are not thinking enough about security. 77% of organizations lack foundational data and AI security practices and only 20% express confidence in their ability to secure generative AI models.”

Deploying AI without security can be a major risk. The article said that “rushing ahead without securing these systems is like building a skyscraper on sand,” and he was referring to systems like your HR database, Excel, and Microsoft Word. Now imagine the risk when AI is built in to your alarm, access control or CCTV systems.

When I write a vulnerability report for your museum, I never allow Microsoft Office to copy it for training purposes. I never run my report through AI to proof read it or to “clean it up”. I

always assume that whatever I put through AI will end up in a server somewhere and will be discoverable now or in the future.

Today, April 13, 2026, we learned that hackers used Claude and ChatGPT to breach multiple government agency networks. Claude executed 75% of the commands needed to complete the hack. In 34 active sessions the hacker logged 1088 individual prompts. These translated into 5,317 AI-executed commands. This would have taken weeks for a hacker to do unassisted, but AI sped up this process. Simultaneously the hacker used OpenAI's ChatGPT for rapid reconnaissance and data processing. The hacker developed a custom 17,550-line Python script designed to capture raw data harvested from compromised servers through the OpenAI's API.

According to *Cyber Security News*, the system analyzed information across 305 internal servers producing 2,597 reports. "The integration of artificial intelligence allowed the attacker to turn unfamiliar networks into mapped targets in hours rather than days. Recovered materials showed that the attacker possessed over 400 custom attack scripts," according to the article.

These are not exotic programs. I pay only \$20 per month for nearly unlimited access to Claude and the same for ChatGPT. Every 14 year-old with a laptop can now have access to a major hacking tool.

I never design a new museum security system that is not run on a network that is dedicated to that system, and I do my best to isolate it from physical or cyber access. I encourage my clients to build a dedicated network for their access control system and another for their CCTV system. And I insist that the terminals into these networks be secured both physically and via passwords. And I resist as hard as I can introducing AI or any other system or program onto these two critical dedicated networks. And most important, the dedicated security networks never touch the internet through any direct wired or wireless connection. I even advocate disconnecting all USB ports and DVD drives on computers on the network like your access control system. The exception is the one isolated computer used to make necessary program changes by the director of security, and I advocate that this computer be in an office that is reader protected. Having a network that is exposed to the internet is a risk we can't take. And having AI integrated with your access control and CCTV systems is not worth the risk. When software like virus protection needs to be updated, the network can be momentarily physically connected to the internet then disconnected when the update is complete.

Does this mean that you can't have and use Chat GPT to help you do your office work? Of course not, as long as it is done on the building-wide administrative network and not on your dedicated security networks. AI is here to stay and is the way of the future. Use it.

Think of Tombstone in the old west. That is where AI is today. It is a dangerous world out there, but it will improve. But before it improves, there will be many gunfights at the OK Corral. You just need to bring your defensive weapon and your common sense until it is tamed. Just don't expect it to be tamed in your lifetime. There is a place for artificial intelligence in your security office, and on your secure networks is not one of them. As long as you use Google and Microsoft Office or a computer with AI built into the application, know that nothing you write is private.

I predict that in your lifetime you will see the world's largest heist occur perpetrated by a hacker using artificial intelligence and the loss will be in the billions. Your only job is to make sure that this heist occurs at some Bitcoin storage wallet and not at your museum. I want you to repeat after me: "Not on my watch!"

Takeaways and Solutions:

- This is a new world. You either change your mindset and keep up or find another career. There is no shame in you not understanding this technology. Things are changing so rapidly that you can't be expected to know what experts are just discovering today. So ask for funding for advanced training on security technology. At least be able to recognize the risks you now face and how to question the claims by sales people that there is nothing to worry about.
- Museum Directors need to start taking security more seriously and devote more funds to helping you prepare for the future which has arrived. This includes training, dedicated networks, and any outside advice you may need to prepare a cyber threat countermeasure program.
- Risks that were minimal last year are now significant because Artificial Intelligence is so powerful and can be easily misused. Complex hacks are now simple for AI to achieve. This has changed everything.
- We can no longer view the latest and greatest technology as being good. We must treat technology as suspect until it is proven safe. Look how many times Windows has been updated, and flaws are still being discovered weekly.
- There is only one way to (almost completely) secure your high-tech security systems and that is to isolate them from outside access to the internet, and dedicate a really secure, isolated, physical network to them. No compromises. No exceptions. You've got to do what you've got to do.
- A separate network is needed for alarm/access controls and for CCTV so that if one is compromised by a cyber-attack or is down for maintenance, the other can remain in operation and serve as a backup. What I mean by this is that if your access control system network fails and alarms no longer report to your command center, you can turn on full-screen motion detection on your CCTV system, and it can function temporarily as an alarm system.
- Manufacturers of high-tech systems need to stop building AI into their software. I understand the competitive environment that forces them to do this. Perhaps the solution is government regulations or UL standards that forbid AI on critical systems and networks. I cannot think of a single reason why AI is needed on an access control system. If you think your guard in the command center isn't good enough monitoring alarms and CCTV and needs AI to help him, then you have a management problem and not a security problem. I guarantee you that AI is smarter than the best guard you have assigned to monitor your alarms. But that may not be a good thing.
- You need to stop demanding that manufacturers use unproven technologies and realize that low-tech is often better and more secure.
- Insurance companies should require higher security standards.
- Every security department in a major museum needs a degreed IT professional with an understanding of both physical and data security in the security department reporting directly to the security director. His or her job is to oversee, maintain and secure the security systems and their dedicated networks. (There are reasons too complex to discuss here as to why this person should not report to the IT department).