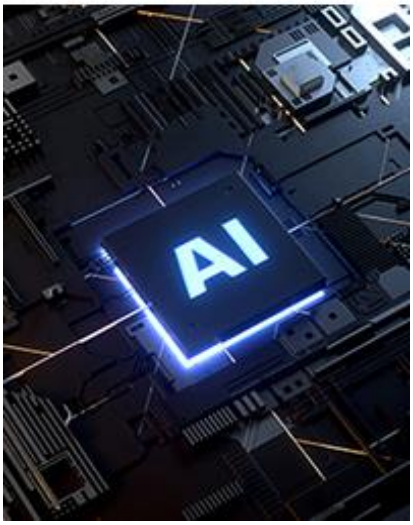
www.pipelinepub.com
Volume 19, Issue 8

Generative AI Attacks Require Dynamic Defense

By: Mark Cummings, Ph.D., William Yeack, CSE

Knives and Bombs



Generative artificial intelligence (AI) is giving cybersecurity attackers a huge jump in capability. Today's widely deployed defending technology is unable to stay ahead of it. It's as though attackers are acquiring bombs, while defenders are still wielding knives. (See Illustration #1.) Without effective defenses, the defenders may not be able to protect themselves and a lot of innocent people on the periphery may be hurt. Effective defenses must be developed quickly. Today's defenses can be characterized as "static." To be beneficial, these new systems must instead be dynamic and they cannot depend on Generative AI themselves. Western governments need to band together with innovators in commercial industries to support R&D focused on the rapid development of defensive systems, so it can respond effectively to attacks created by Generative AI systems.

Static vs. Dynamic

Today's defenses are static. That is, they use predetermined (static) patterns to identify attacks and scripts to apply responses (often called remediation). Because of their static nature, these can be classified as **S2** (Static attack ID/Static remediation) systems. They work well against classes of attacks that are employed repeatedly and change relatively slowly.

Pattern recognition defenses can be thought of as a series of sieves. Of those, each sieve will only allow a very specific shape to pass through. Then, all the data in a system are poured through the

sieve and if any come through, it is recognized as an attack. For each type of attack that has been found and analyzed, a sieve is created and data is poured through all the sieves one after the other. When a new type of attack appears, it is not immediately recognized. But when the damage becomes obvious, professionals analyze it and create a sieve.

Illustration #1: Advantage of Generative AI Created Attacks

click to enlarge

Scripted remediation, meanwhile, can be thought of as a series of recipes. When a type of attack has been identified, it is analyzed to determine how best to counteract it – stop any further damage and repair what has happened. This remediation generally involves a series of steps and often brings in new or corrected system components to replace damaged ones, reconfigures others, etc. A recipe is a good metaphor for this, because it also is generally a series of steps based on a series of ingredients.

As long as the 'shapes' of attacks don't change too rapidly, these kinds of defenses can limit damage. But, if the 'shapes' change very rapidly, there are not the necessary sieves readily available to catch them in time. Because the type of damage changeswith the change in type of attack, trying to use a scripted recipe from a different kind of attack for remediation is like trying to use a recipe with the wrong ingredients. Against this background our digital world has grown bigger, more complex, and more all-encompassing. In doing so, it has opened a large and growing opportunity for attacks. And, the damage that attacks can cause has also increased. Up until now, it has been primarily human attackers finding new types of attacks, and human defenders creating new sieves and recipes in response. Attackers have used automation to increase the number of attacks and speed with which they can deploy new ones. That has been a challenge for defenders. But, something of a balance has been achieved.

Many argue it's not a *good* balance, as demonstrated by the size of the financial cybersecurity losses, etc. But still, something of a balance. Generative AI is fundamentally changing this balance, however. It has the ability to rapidly identify a very large number of new generic attack

types. Plus, it can customize these for a particular target. The cost of each launch is relatively low. So, not every attack has to be successful -- that is, pay off. These attacks can be loaded into the automated attack systems and launched. As a result, the number and variability of attacks will accelerate dramatically.

The result is a large number of attacks that change very rapidly – too rapidly for patterns to be identified and installed in today's defensive tools. It will be extremely difficult for these systems to defend against these Generative AI attacks. It is similar to Covid-19 mutating faster than scientists could produce a vaccine, in order to fully protect against the spread of the disease.

These types of Generative AI created attacks can be characterized as dynamic. Because the attacks are ever-changing, an effective response can't be easily anticipated and scripted. A different approach is necessary. One that can respond to the dynamic nature of the attacks.

Generative AI Attacks

Generative AI is on a rapidly accelerating course. There are attempts to control it so that it's not used for nefarious purposes. But, there are also well documented ways of bypassing those controls. Attackers now have these new Generative AI "bombs" used to attack information systems.

This is occurring rapidly. So quickly that at RSA (the largest cybersecurity conference held annually in San Francisco) this year, there were no presentations on the above threats as detailed. But the conversations in the hallways were dominated by them. That means that in the few months before the conference while people were preparing their formal presentations, the threat was not fully recognized. But when the few people who understood it started talking about the threat in the hallways, it spread like wildfire.

Generative AI Can't Defend Against Generative AI

Some defenders will try to use Generative AI to defend and this may set off an arms race. But, defending Generative AI's will always be at a disadvantage no matter how powerful and fast they are. This is because many of the Generative AI attacks will be targeting the network edge or the middle. The defending Generative AI in a data center will have to:

1. Gather information at the point of attack (network edge, middle, etc.);
2. Deliver it to the Generative AI to be combined with large amounts of data from other parts of the network (collection and handling of this amount of data will be time-consuming);
3. Process all of the data to locate, identify, and characterize the attack;
4. Determine the correct remediation for that attack;
5. Finally, transmit instructions to the network edge or middle to perform remediation.

The time inherent in these steps will give the attacker enough time to achieve the attack's objective. It is a fundamental latency problem that, for technical reasons, the data center Generative AI's will not be able to overcome. On top of this, experts inGenerative AI technology

are afraid of losing control over Generative AI systems themselves. That is, Generative AI systems acting like the sorcerer's apprentice. This can be as simple as receiving unexpected results, or as dangerous as actions that threaten lives. Some go so far as to be concerned about the survival of the human race. Governments are taking these concerns very seriously.

Dynamic Defensive Tools

To defend against Generative AI created attacks, we have to develop alternatives to the sieve and the recipe. We need ways of recognizing that an attack is underway that do not depend on patterns. Similarly, we need ways of determining remediation without relying on recipes. One way is to follow the dynamic and adaptive methods nature has used in biology, but adapted to digital information systems. Because of:

- Latency issues, they must work at the point of attack;
- Remediation involving more than one part of a system, they must enable cooperation between system components for recovery;
- Volume of information problem, they must have only the data needed, where it is needed, when it is needed;
- Concerns about systems getting out of human control, they must have flexible ways for manual intervention to guard against unexpected outcomes.

These dynamic and adaptive defenses would be able to dynamically identify attacks and dynamically determine and implement remediation. They can be characterized as **D2** systems.

S2-D2 Defenses

It is tempting to imagine a scenario where we move rapidly from our current S2 defenses to D2 defenses. D2 systems can defend against S2 attacks as well as D2 attacks. But, organizations have large sunk investments in S2 defenses. Attackers also have large investments in S2 attack systems. We are beginning to see an upsurge in Generative AI-created attacks. But, there continue to be large volumes of static attacks. Thus, it makes sense to have defenses that can take advantage of the sunk investments in S2 defenses while also fielding D2 defenses. Such systems can be characterized as S2-D2. It is these S2-D2 systems we need to quickly develop.

Responding to the Challenge

Western governments need to band together with innovators in commercial industries to support R&D focused on the rapid development of S2-D2 systems. Governments in the US, Europe, Canada, Australia, and New Zealand have programs to provide financial support for R&D and entrepreneurship. These programs come from both the national defense and the industrial policy departments of these governments. Because of this, they have tended to be narrow in focus—often more concerned with improving national competitive positioning against the rest of the world. But, the threat from Generative AI cyberattacks is a global one. The Western world needs an effective response—and quickly. This can best be achieved through established cooperation between the leading Western governments while working closely with innovators in the commercial industry.