**ChatGPT is Already Posing New Cybersecurity Risks**

AI will likely make a large impact on the quality of life for many people over the coming years after we get over the fear of the new technology. Much like it took time for much of the world to embrace combustion engines rather than horse/mule-powered wagons, this too will take time. At first, we will go through phases of fear and worry about the weaponization of AI, some will even have images of Terminator running through their heads. But as time progresses, people will start to embrace and implement AI into their daily lives. There is already evidence of this happening across the security industry as companies race to see how ChatGPT and other AI tools can further improve their tools and detection capabilities.

While time will tell how useful the AI tools are, we are already seeing indications of malicious actors and researchers taking the time to evaluate how things can or will go wrong. It’s almost impossible to miss the next big fear or disaster scenario in these early months after release but in this article, we will look at some of the concerns organizations and individuals may have come across, what this means, how AI is being used to help with this use case.

**Loss of intellectual property**

To fully understand the risk posed by an AI chat tool, we must first level set how these tools truly learn. Learning takes place based on the input of information, feedback, and corrections that take place over the course of hundreds of thousands of people using the tool. If a user provides input, say code, to the tool to ask a question, this will allow it to save this code for future reference to provide better answers to another user (or even the same user) at a future date.

Let’s go a little more abstract here. Say your company is building a new widget for the market that will solve a major pain point or disrupt a sector. Your developers decide to utilize ChatGPT, or another AI tool, to assist with writing code that they are struggling with. By uploading snippets of your code and asking for help with solving runtime failures, your developers are training the AI tool in how your widget works. While the full code base may not be uploaded to the platform, by uploading snippets or small sections at a time, your developers are providing the AI tool with components to assist with building your widget.

It must be acknowledged that one or two uploads may not allow the AI tool to fully build your awesome widget. However, given enough input, and even input from competitors, it’s possible that if asked the right question, the AI tool could put together a functioning code base that would be similar, if not the same, as your widget. This may seem like a push, but we already know that ChatGPT, and other AI-based tools, are able to write fully functioning code for good and evil. So, do not assume that it is not possible that someone would be asking how your widget works and could potentially get a near replicable of your widget code.

**Phishing lures made easy**

Where this is a tool that can make life easier for a malicious actor, they will find it and use it. In this case, there are already attacks in the wild that are using ChatGPT to generate content for phishing attacks to make them more successful. These AI tools only help malicious actors to be more efficient in an already prevalent attack mode against organizations. The reality is that phishing is a low-cost, low-effort way for a malicious organization to gain access to corporate networks, attain information, or commit fraud. So, there is no reason not to expect them to utilize these tools to assist in this attack pattern.

The AI tools streamline the creation of content by allowing malicious actors to close the gap in language and culture for the target organization. Further, the creation of content is streamlined using questions and input, allowing them to create customized payloads that are highly targeted for an organization in a matter of minutes. The increase in speed to craft the payloads will allow for more emails to go out to targeted victims which then increases the odds of catching someone clicking a link or downloading the malicious file.

**Say it ain’t so: mAIware**

AI developers have attempted to implement protected controls to prevent the use of their AI tools for malicious or illegal activity. However, researchers ([CyberArk](https://www.cyberark.com/resources/threat-research-blog/chatting-our-way-into-creating-a-polymorphic-malware), [Check Point](https://blog.checkpoint.com/2023/02/07/cybercriminals-bypass-chatgpt-restrictions-to-generate-malicious-content/)) have already been able to create a proof of concept malware that bypasses these security controls through the rephrasing or manipulation of the platform. This should not be surprising as the controls to prevent malicious content are still written by humans and, as these researchers have found, are largely focused on a negative list of words to prevent the generation of the content.

While this may be scary to think about, it is important to point out that this has not been identified as malware being fully written by an AI tool but rather only improved. Further, this will not necessarily introduce any new attack vectors against individuals and organizations but will just provide another tool for malicious actors to use to generate their content. Due to the generation of the malware requiring thoughtful and well-structured requests into the platform, malicious actors may likely stick to using the AI platform to partially generate code and help to speed up the generation of malware that will be introduced into the wild.

**Trojan AI scam**

Where there is money to be made, or systems to hijack, malicious actors will find a way. Recent reports have indicated that malicious actors have pivoted to taking advantage of the lack of availability for ChatGPT to offer up services for mobile devices and plugins for browsers that claim to bypass or provide additional availability of the service. In these cases, these tools are simply malware designed to take over your system, steal data, or attain credit card information for fraudulent purchases. These attack vectors have been highly successful in recent months and will most likely continue to proliferate across different versions, different actors, and with different end goals.

This new vector of attack may be scary for individuals but should be especially worrisome for business leaders and security practitioners. To date, these tools have been targeting individuals but there are recently published variants that are taking over business Facebook accounts and could easily transition into other types of organizational attacks like malware/ransomware deployments.

**Copywrite/Opensource Concerns?**

There has yet to be a case brought in a major court regarding copyright or open-source license infringement for a text-based AI tool. However, a recent US Court decision regarding the AI tool Midjourney may provide a precursor to how this may affect individuals and businesses. In the ruling, related to author Kris Kashtanova, “Zarya of the Dawn”, a [letter](https://fingfx.thomsonreuters.com/gfx/legaldocs/klpygnkyrpg/AI%20COPYRIGHT%20decision.pdf) was released that the images generated by the AI tool could not be copyrighted as part of the book. As part of the ruling, the US Copyright Office determined that the images were not able to be copyrighted as they were not generated by a human and therefore were not eligible for copyright protection.

As for open-source licensing, every AI tool has an open-source license agreement within the platform that may stipulate how the material produced by the AI tool may be utilized. For an organization looking to use the results of the text-based tool, whether code, marketing material, or other use cases, it may be possible to be out of compliance with the open-source agreement if not thoroughly reviewed and understood. There are plenty of precedence court rulings in the US and outside of the US where a company was sued and subsequently lost due to not properly following open-source licensing agreements.

Both cases are lumped together as one of the biggest concerns for many security practitioners is related to the protection of assets for a business. Copywrite claims, or intellectual property claims could be severely weakened with the ruling against Ms. Kashtanova and the lack of understanding of the open-source agreement that is in place for the use of anything generated from the AI tools. While we are in the early stages of AI tools being utilized by companies, it can be expected that we will start to see more court cases that impact companies that have relied on the tools to assist with generating content and intellectual property.

**What Happens Next?**

Some may compare AI tools to the discovery of atomic fusion which was harnessed during horrific events that forever left a scar in history. While atomic and nuclear weapons still hold the world hostage, even more so in current events, we still found a way to harness the technology for good by using it to generate electricity to help power further inventions and capabilities efficiently.

As touched in on the introduction, a fatalistic mentality may have swept society regarding AI. However, there are plenty of people out there that are finding cool and interesting ways to leverage AI-based technology for good (and bad). It will take time to fully understand how AI will impact us and society, but we can take some lessons learned from history. Many times, in the past we have introduced new technologies that have scared us to death to only find that there is good to come from it. As the tools continue to mature, humans will find better and more interesting ways to leverage the tools to produce fantastic outcomes. Or we will end up living through “The Terminator” for real, either way, things are getting interesting.