# ****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.

## ****#1: Employees Are Uploading Sensitive Company Data to ChatGPT****

AI tools 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.

As covered by [DarkReading](https://www.darkreading.com/risk/employees-feeding-sensitive-business-data-chatgpt-raising-security-fears) and [Cyberhaven](https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt/) articles, over 4% of employees have pasted sensitive data to ChatGPT. Data provided to ChatGPT has included sensitive information regarding company strategy documents, patient information, software code, to intellectual property that is not publicly accessible. This is a unique threat to companies as it is not easy to classify the data that may be sent to an AI platform as it can be copied and pasted from documents or manually transcribed from screenshots or other documents. All this data is utilized to improve responses and information provided back to end users which could result in data leakage through properly worded questions to the platform.

## ****#2: ChatGPT is Being Used to Write More Convincing Phishing Emails****

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.

Since the launch of ChatGPT, while attacks have stayed steady according to an article by [The Guardian,](https://www.theguardian.com/technology/2023/mar/08/darktrace-warns-of-rise-in-ai-enhanced-scams-since-chatgpt-release) it has lowered the entry into phishing for attackers. Not only can the tool help to cut across language and culture, but it can also assist with increasing the complexity of the phishing attack, making it more difficult to detect. A recent study by [HoxHunt](https://www.hoxhunt.com/blog/chatgpt-vs-human-phishing-and-social-engineering-study-whos-better) shows that the failure rate between a campaign developed by a human vs. AI are nearly identical however, there is still a ways to go for AI-generated campaigns to be as successful as human phishers.

One example of how this is currently being used as an attack vector, according to [Bitdefender](https://www.bitdefender.com/blog/hotforsecurity/bitdefender-labs-warns-of-fresh-phishing-campaign-that-uses-copycat-chatgpt-platform-to-swindle-eager-investors/), is simply through the utilization of spam or unsolicited emails that appear to be marketing ad. However, since the links are controlled by malicious actors, it redirects users to a copycat version of ChatGPT that even interacts in a similar fashion. After answering a few questions, the copycat version will ask for the user’s email and phone number and then proceed to prompt them to make money off of the tool. Shortly after this, the user will receive a call where a person on the other line will then ask how much money the user is able to invest in order to start making money on the tool.

Graphical user interface, text, application

Description automatically generated

## ****#3: ChatGPT is Being Used To Develop New Malware****

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.

Below are some examples of malware being built or modified with ChatGPT.

InfoStealer

In an darknet forum, a malicious actor made a post regarding how he was building an information stealing tool with the assistance of ChatGPT. As can be seen from the below example, pulled from a [Checkpoint](https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/) writeup, the malicious actor utilized ChatGPT to help with the python based code.

Graphical user interface, text

Description automatically generated

Code Mutation

As can be referenced in the [CyberArk](https://www.cyberark.com/resources/threat-research-blog/chatting-our-way-into-creating-a-polymorphic-malware) research article, malicious actors can use ChatGPT to mutate or alter code to make it harder to detect or to alter the behavior. This becomes especially useful when trying to create new variants that may avoid detection. As part of this, the malicious actors are able to mutate as many times as desired, until the desired results are achieved or enough obfuscation is obtained to make it harder to attribute the attack to an individual.

Text

Description automatically generated

Encryption Tool

In another example in the [Checkpoint](https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/) article, a user claims to have built an encryption tool with no experience writing code.

Graphical user interface, text

Description automatically generated

## ****#4: There Are Trojan Scams Posing as ChatGPT****

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.

There are multiple reports of malware in the Apple, Google, and Microsoft stores that are masquerading as ChatGPT apps. According to articles by [bleepingcomputer](https://www.bleepingcomputer.com/news/security/hackers-use-fake-chatgpt-apps-to-push-windows-android-malware/) and [TrendMicro](https://news.trendmicro.com/2023/03/09/fake-chatgpt-app-website/), these applications are coming in various forms with varying target actions. In some cases the application is used a loader application for other malicious applications, will subscribe the user in premium SMS services, act as a spyware, or attempt to steal data from other applications.

Another form of malware or trojan applications that is popping up are browser plugins. One specific case, as covered by [DarkReading](https://www.darkreading.com/application-security/chatgpt-browser-extension-hijacks-facebook-business-accounts) covers a specific case where users install a Chrome plugin to access ChatGPT as part of their browser to have their business Facebook account taken over by the extension.

## ****#5: Security Leaders Believe ChatGPT Cyberattacks are Coming****

A recent study conducted by [Blackberry](https://blogs.blackberry.com/en/2023/03/the-growing-influence-of-chatgpt) highlights a growing concern within the IT industry related to the exposure or threat posed by AI tools. 51% of respondents hold the perception that AI based tools will lead to a breach or assist with a breach in the next year while 71% believe that AI will be harnessed for malicious acts.

While there is a belief that AI tools will add additional layers of complexity and threat to the cyber industry, it has significantly assisted with defensive capabilities. A [Global Threat Intelligence Report](https://www.blackberry.com/us/en/solutions/threat-intelligence/2023/threat-intelligence-report-jan) published in Jan. 2023 showed that AI based security tools have stopped over 1.7 million malware attacks over a 90 day period. This data point shows why there is such a large investment in AI based tools even considering the belief that it will assist with a breach or cyberattack in the near future.

## ****#6: ChatGPT Affects Security With Copyright Headaches****

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.