Campaign in a Box

Generative AI

livingsecurity

Welcome Message for Program Owners

Hello!

Welcome to your **Generative AI** Campaign in a Box! This box is designed to help your team understand the security concerns surrounding this powerful new tool. Generative AI can help us all work faster and easier, but it has to be used safely!

Via Giphy

In this month's box, you will find:

- **A blog post from Living Security.** You can share a link to our website or republish this article on your own intranet.
- **A campaign announcement email to send to your users about this month's theme.** Just a little introduction to kick-start the campaign!
- **Weekly email tips to keep users learning throughout the month.** If you're unable to share these via email, feel free to share them through another channel that's more appropriate for your organization.
- **Weekly chat messages that you can send via Slack, Teams, etc., to reinforce the lessons.** Depending on your messaging client, you may need to save the provided GIFs to your computer to attach them to your chat messages.
- **Complementary Resources.** See Living Security training modules and Family First content to support this month's topic.

You are absolutely free to edit and customize the content we send as you see fit—make this campaign your own! Please don't hesitate to let us know if there's something you would like to see in a future campaign.

Best wishes,

The Living Security Team

Blog for End Users

Can you *imagine* traveling back in time and explaining generative AI to someone in the ‘90s? They just barely invented text messaging before the end of the decade and here we are with robots that’ll write our emails or paint us digital paintings beyond our imagination. That’s wild!

Via Giphy

Generative AI tools like ChatGPT and DALL-E are incredibly powerful and can do wonders for productivity, but like anything that awesome…there’s a catch. They come with their own assortment of security concerns. The good news is that these security concerns can be navigated, so long as we understand generative AI best practices.

Privacy concerns are perhaps the biggest security risk when interacting with generative AI. Oversharing with AI could result in a data breach if a user shares confidential information with an AI, then it turns around and uses that information in an answer to another user. Personally Identifiable Information (PII) such as full names, home addresses, social security numbers, and personal contact information should never be shared with AI, especially if the PII in question is not your own. Other types of sensitive data, such as proprietary or financial data belonging to your organization, should also be kept out of your chats with AI. Basically, if you wouldn’t share it with the public, you shouldn’t share it with AI.

Keep in mind that privacy isn’t just a security concern; it’s also a *legal* issue. Regulations like the GDPR and CCPA set guidelines for how organizations are obligated to protect consumers’ personal data. If that personal data is fed to a generative AI tool and subsequently leaked, it can land the organization in legal trouble.

Beyond the risk of the generative AI tool sharing your information with other users, it’s also possible for an unauthorized party to gain access to your account and view your chat history. For this reason, it’s a good idea to regularly clear your account’s chat history.

Via Giphy

If you want some AI assistance on a project that involves sensitive information, you're not entirely out of luck. You just need to get a little *creative*. Try replacing the information you can't or shouldn't share with either anonymized or generic information.

- **Example 1:** Rather than typing a client's name, use a common pseudonym like "John Doe" or "Mary Sue."
- **Example 2:** Instead of using your organization's real name, use "Company Name" or "My Organization."
- **Example 3:** Replace an address with "1234 Street Name" or a phone number with "(555)-123-4567."

Even if you're anonymizing data or sharing information that isn't considered sensitive, every interaction you have with generative AI involving work data (or on a company device) needs to be in line with your organization's security policies. If you're unclear on these policies, reach out to your security team for help.

Via Giphy

Privacy issues aren't the only security concern surrounding generative AI. Unfortunately, these tools are available for anyone to use - *including* those with malicious intentions. Cybercriminals have found some pretty effective ways to deliver cyber attacks using AI.

Social engineering attacks are nothing new, but AI has increased both their effectiveness and their prevalence. Sometimes referred to as "human hacking," these attacks of manipulation often begin with research on a specific target. Rather than split their time and attention between researching a target and crafting their attack, threat actors are now able to have AI launch phishing attacks for them. This allows them to find more of a person's personal information, which can be used to win their trust, *and* it allows them to launch many more attacks in the same amount of time. You're likely already familiar with one of the most common social engineering attacks, phishing - those pesky emails, text messages, and phone calls that come from cybercriminals lying about who they are. As always, avoid clicking links, opening attachments, or sharing personal information with unknown senders.

One of the best ways you can prevent these attacks is by limiting what personal information

be found about you online, including your personal email address and phone number. The less a social engineer can find about you, the harder it will be for them to convince you they can be trusted. Your social media is a great place to start. Consider going through your post history and removing anything that contains personal information.

Via Giphy

There's another huge concern surrounding AI that has little to do with the user and more to do with the tool itself - misinformation. These tools may *seem* like they know everything, but what they're actually doing is pulling huge amounts of data, finding patterns, and using those patterns to make something new. Generative AI may be a speedy and powerful tool for research and content creation, but at the end of the day it relies on pattern recognition and lacks human judgment. This means it frequently (and very confidently) will generate information that is just plain wrong. It's even been known to make up fake sources to back up its incorrect claims.

Obviously, this means you should take what generative AI tells you with a grain of salt. What may be *less* obvious is how much more prevalent this may make misinformation on the internet. Generative AI tools are increasing in popularity, and more and more organizations are using them to create online content. Rarely do these articles, blogs, videos, etc. state that they're generated by AI, so it's likely you'll begin to interact with a lot of AI generated content and never even know it. That's why it's more important than ever to double-check the credibility of listed sources and independently fact check claims that aren't backed up by a source.

Advancements in generative AI continue to make deep fakes more convincing and easier to make, so don't overlook video footage of public figures when you're on the hunt for misinformation. Recordings of a person's face or voice can be used to create media (video and/or audio) depicting them doing or saying anything a user would like. This may sound alarming, but deep fakes *can* be identified. If a video or audio recording seems shocking, play it again and pay a little extra attention for anything that feels uncanny.

Via Giphy

Society's only beginning its exploration into the possibilities of generative AI, and with continued technological advancements will come new security concerns. Staying informed will be your best form of defense against misinformation and threat actors who seek to use this technology maliciously. If you don't follow one already, considering adding a tech news publication to your reading list so you can stay on top of generative AI best practices!

Week 1

Email Message

Subject: 🤖 AI is the talk of the town 🤖

Hello, hello!

I don't know about you, but there's a buzzword I've been hearing *a lot* lately. It starts with an 'A' and ends with an 'I.' 👀 That's right, this month we're diving into the magical world of AI - specifically, generative AI such as ChatGPT and DALL-E.

These tools can make our work easier, faster, and more efficient, but they can be dangerous if used incorrectly. The information shared with chatbots and other generative AI tools may be added to the AI's databases and used to generate content for other users, so it's important to be careful about what we share.

Via Giphy

When it comes to work data, confidential or sensitive information should never be shared with a third party, which *includes* AI tools. Sharing private data with AI can result in a data breach, which can hugely damage our organization's reputation and financial resources. It can also put our organization in violation of legal regulations surrounding people's personal data, such as the GDPR. Personally identifiable information (PII), proprietary data, and any other sensitive information is best out of your conversations with AI.

Instead, try anonymizing or genericizing the data you feed to AI tools. This could look like using "Person 1" or "John/Jane Doe" instead of a real client name or replacing the name of our organization with "Company Name."

Don't forget that, like basically any other program or webpage on the internet, AI tools like chatbots are gathering information about you as a user beyond what you're giving it on purpose. The parties hosting these AI tools can learn all kinds of things about you based on the information you provide to create an account and the way you interact with the AI. Beyond limiting what you share, you can protect your privacy by using temporary accounts or using a pseudonym when you create your main account. It's also a good idea to clear your chat history periodically, in case your account is ever compromised.

And, as always, make sure you're familiar with the privacy policies of the platforms you create accounts with!

All the best,
{{ SIGNATURE }}

Chat Message

Generative AI tools like ChatGPT are all the buzz lately. 🐝 I don't know about you, but I don't think I've gone 24 hours without hearing or reading about AI in...I don't even know! It's not the talk of the town for no reason. Used right, generative AI can be an incredibly helpful tool that allows us to work smarter, faster, and easier! 💪 But if used wrong, it opens up individuals and organizations to all kinds of risk. We'll explore that risk and how to mitigate it more throughout the month, but for now - *be sure you're familiar with our organization's policies surrounding the use of AI.* If you're not up to date, reach out to your manager or a member of the security team!

Via Giphy

Week 2

Email Message

Subject: 🧰🔪 AI's not just a tool for *you*...

Hi, Everyone!

If there's one thing I've learned from the movies, it's that great power in the wrong hands can be...less than ideal. Generative AI is no exception. While we've been figuring out how to use it to write better emails and organize our calendars, cybercriminals have been learning to use it for cybercrime. With the use of AI, threat actors can launch attacks that are both wider-reaching and more convincing than before. But don't panic, because if you understand how AI is being used maliciously, then you can understand how to stay safe.

Via Giphy

They may be using fancy new tools, but cybercriminals still have a favorite trick - *social engineering*. Sometimes referred to as "human hacking," these attacks aim to manipulate a person into acting against their own best interest. You're likely familiar with arguably the most common type of social engineering attack - phishing. Threat actors will often gather information about you on social media in order to craft messages specifically to trick *you*...and with AI writing these messages, they're getting more convincing. This makes it more important than ever to be careful what you post on your social media. If you haven't done so in a while, it may be a good idea to comb through old posts and remove anything that feels like an overshare.

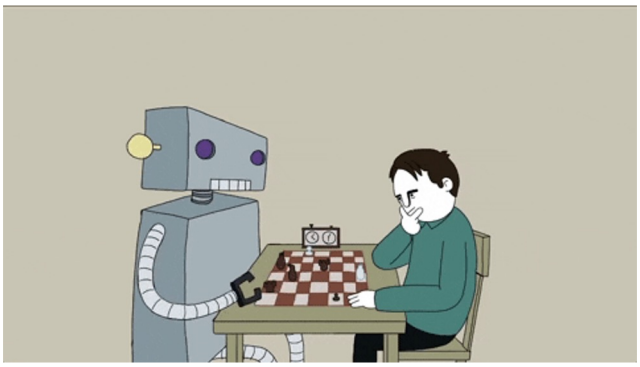
AI allows threat actors to automate attacks, which means you can expect an uptick in phishing emails, texts, social media DMs, and phone calls. If you receive communication from someone you don't know, slow down and look for red flags before sharing any personal information, clicking a link, or downloading an attachment. A sense of urgency, odd grammar choices, or a sender email address that's slightly off from one you trust can alert you that something isn't right.

Remember, if you ever get a suspicious message to a work account, be sure to report it to the security team so they can help keep others safe, too!

Stay safe out there!
{{ SIGNATURE }}

Chat Message

AI might not be better than people at most things quite yet, but there *are* some things it can do much, much better than we can - *like fool other AI.* 🤖 AI continuously pools large amounts of data and analyzes it for patterns, making it an incredible tool for security systems (like email spam filters, for instance). Unfortunately, with the growing popularity of generative AI, it's become easier for cybercriminals to slip malicious code or illegitimate emails past AI security systems. 🦹 Remember that technology isn't the end-all-be-all of cybersecurity. *You* have the power to protect yourself *and* our organization from cybercrime. If any suspicious activity slips past our security tools, be sure to let the security team know!



Via Giphy

Week 3

Email Message

Subject: 🧍🧍 Looks can be deceiving

Hi again!

Last week we talked about social engineering and how to stay safe from AI-powered phishing attacks. Today we're exploring what happens when AI gets a little more...creative. Deep fakes are pieces of media - video or audio - created by AI, and cybercriminals are finding new ways to use them for personal gain.

Via Giphy

Perhaps one of the more alarming developments in the world of AI-driven cybercrime is that of deep fake vishing attacks. Recently there have been reports of people receiving calls from kidnapped family members being held for ransom - except these family members are *actually* safe and sound in their own homes. The calls are deep fakes, created using audio of the supposed victim's voice. These cruel scam calls are effective because they cause the family to panic, making them both less likely to notice mistakes in the voice clone *and* more likely to pay the scammer without thinking. If you receive one of these calls, breathe. Use another phone to call the loved one in question and make sure it's them you're actually talking to.

Vishing attacks aren't the only concern when it comes to deep fakes; they also open up huge concerns about disinformation. From political motivations to a twisted sense of humor, people spread disinformation for all kinds of reasons, and deep fakes offer a much more convincing way to do so than just an old-fashioned, text-based social media post. Next time you see a video on the internet that shocks you, look a little closer. Deep fakes might be convincing at first glance, but most have give-aways if you scrutinize them. Unnatural facial expressions, jumbled text in the background, and blurred hands or hair are common give-aways to look for.

Deep fakes may seem scary, but they're absolutely possible to distinguish from real media. If something seems hard to believe, take a second look.

{{ SIGNATURE }}

Chat Message

You already know that not everyone is who they say they are online. If that weren't the case, we wouldn't have one of MTV's best shows, *Catfish*. 🐟 Well now we have to *also* keep in mind that not every chatbot is who they say they are. Cybercriminals are able to harvest personal information by tricking people into talking to *their* chatbots instead of legitimate ones. They may imitate bots used for customer service, entertainment, or even work. 💼 So before you share personal or private information with a chatbot, double-check that the website the chatbot is hosted on it trustworthy.

Via Giphy

Week 4

Email Message

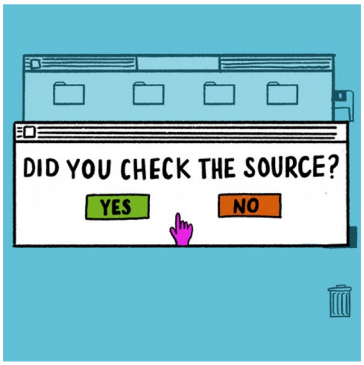
Subject: 👂🤔 Don't believe everything you hear...

Hey, Team!

We've spent the month covering the security concerns surrounding generative AI and how we can use it safely, but today let's take a step back. How does generative AI even work?

If you're interested in the nerdy nitty gritty, articles like this one can help you understand the finer details - but the big thing to know is that generative AI uses pattern recognition to turn a huge data set into something new. This pattern recognition helps the AI understand how humans communicate (whether with words, pictures, etc.) so the content it outputs feels as natural as possible. However, this super power is a bit of a double-edged sword.

Because it's searching for patterns in a data set rather than hand-selecting the most credible information, generative AI can't always be trusted. Content it creates may be biased, misleading, or downright untrue. What's worse is the facts and sources it makes up are often believable enough that they may not raise suspicion - meaning most users aren't fact-checking misinformation generated by AI.



Via Giphy

So what can you do to avoid being duped? As simple as it sounds, the answer is this: don't take AI at its word. Use trustworthy sources to double-check information given to you by generative AI, and if the AI itself is providing you sources, make sure to double-check that they're *real.*

Keep in mind that, with the rising popularity of generative AI, you might not always know when you're interacting with the content it produces. More and more organizations are using AI to generate articles, blog posts, video essays, social media posts - you name it. It's more important than ever to double-check the information you see online.

Keep lookin' out!
{{ SIGNATURE }}

Chat Message

Whew! 😪 We sure have covered a lot of info on generative AI this month...but the most exciting thing about it is that it's a relatively new frontier. For the next few years, there will be continuous developments in how, when, and why we use AI - and *also* in the security risks that accompany its widespread popularity. One of the best things you can do to maintain your cybersecurity superpowers during this age of AI exploration is to stay on top of the latest news and expert recommendations. 🧠 Whether it's podcasts, online publications, or good old-fashioned print...wherever you get your news, add AI security to the list of topics to keep an eye on.

Via Giphy