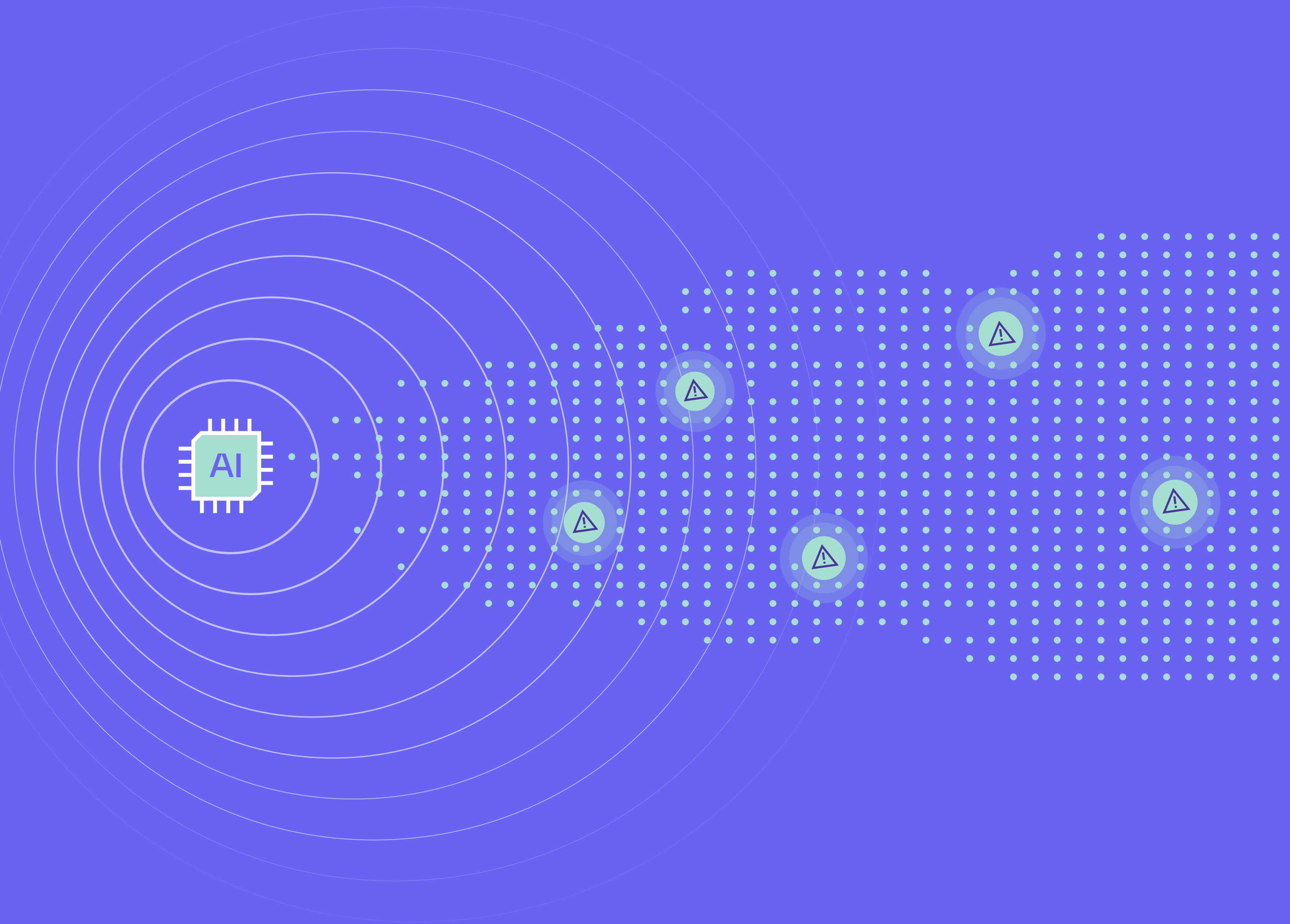
Abnormal

RESEARCH REPORT

AI Unleashed: 5 Real-World Email Attacks *Likely* Generated by AI in 2023



Executive Summary

97%

of security stakeholders are concerned about the risks of generative AI in the next twelve months.

The State of Email Security in an AI-Powered World, 2023

98%

of cybersecurity professionals believe that AI-powered security solutions are needed to detect and block AI-generated threats.

The State of Email Security in an AI-Powered World, 2023

$1.75M

in savings experienced by the average organization that invests in security solutions that use AI.

IBM Cost of a Data Breach Report, 2023

The past year witnessed revolutionary advancements in the field of generative artificial intelligence, or generative AI. Shortly after their launch, platforms like ChatGPT and Google Bard gained significant popularity due to their ability to understand and generate human-like text. Their user-friendly design also made them accessible to individuals without a deep understanding of machine learning, contributing to widespread adoption—with ChatGPT reaching its first 1 million users in only five days.

The unfortunate news, especially for those in the information security space, is that the accessibility of generative AI also created opportunities for cybercriminals to exploit the technology to create sophisticated cyberthreats—often with email as the first attack vector. To complicate matters, malicious platforms like FraudGPT and WormGPT have surfaced alongside legitimate open-source applications, making it even easier for threat actors to enhance their attacks. And with generative AI at their disposal, these bad actors can now create malicious attacks with more sophistication and at greater volumes than ever before.

To illustrate the transition to this new AI-powered threat landscape, we've collected five real-world examples of likely AI-generated malicious emails our customers have received in the last year. These examples demonstrate how bad actors are already leveraging AI across a variety of email attack types, including credential phishing, malware delivery, and payment fraud. They also point to a startling conclusion: **threat actors have clearly embraced the malicious use of AI.** This means that organizations must respond in kind—by implementing AI-powered cybersecurity solutions to stop these attacks before they reach employee inboxes.

Table of Contents

How and Why AI Is Being Weaponized for Email Attacks

While generative AI holds incredible potential for innovation and efficiency, it also has a dark side. Indeed, 91% of security professionals reported experiencing AI-enabled cyberattacks in the past six months. And no surface is more vulnerable to AI-powered attacks than email.

Previously, many cybercriminals relied on formats or templates to launch their campaigns. Because of this, a large percentage of attacks share common indicators of compromise that can be detected by traditional security software, as they use the same domain name or the same malicious link. Generative AI, however, allows scammers to craft unique content in milliseconds, making detection that relies on matching known malicious text strings infinitely more difficult.

Generative AI can also be used to significantly increase the overall sophistication of social engineering attacks and other email threats. For instance, bad actors can abuse the ChatGPT API to create realistic phishing emails, polymorphic malware, and convincing fraudulent payment requests. And even as OpenAI has placed limits on what ChatGPT can produce, cybercriminals have responded by creating their own malicious forms of generative AI.

For example, WormGPT lacks the guardrails found in the open source tools that prevent bad actors from using them unethically. It is a platform designed specifically for criminal activities, including the creation of effective phishing emails. And it doesn't stop there. Similarly, FraudGPT is a subscription-based malicious generative AI tool that uses refined machine learning algorithms to generate deceptive content. This platform acts as a cyberattacker's starter kit by capitalizing on existing attack tools, such as custom hacking guides, vulnerability mining, and zero-day exploits. Unfortunately, cybercriminals are unlikely to stop with these two, as they continue to find new and innovative ways to abuse generative AI models.

Over the past year, cybersecurity professionals have become increasingly aware of the dangers of malicious AI, with 98% of security stakeholders saying they are at least somewhat concerned about the cybersecurity risks posed by ChatGPT, Google Bard, WormGPT, and similar tools. This highlights the need for organizations to develop robust defenses, enhance detection capabilities, and remain vigilant against emerging threats. The first step in doing so is to understand what this threat is.

98%

of security stakeholders say they are at least somewhat concerned about the cybersecurity risks posed by ChatGPT, Google Bard, WormGPT, and similar tools.

Real-World Attacks (Likely) Generated By AI

Over the past year, Abnormal detected a number of attacks that were likely generated by AI. As with all AI models, understanding whether an attack was created by AI with 100% certainty is nearly impossible. However, certain tools provide strong indications of AI involvement.

How an Attack is Determined as *Likely* AI-Generated

After initial email processing, the Abnormal platform determines the probability that an attack was AI-generated by utilizing CheckGPT—a newly-launched tool that leverages a suite of open source large language models (LLMs) to analyze how likely it is that a generative AI model created the message. The system first analyzes the likelihood that each word in the message has been generated by an AI model, given the context that precedes it. If the likelihood is consistently high, it's a strong indicator that the text was potentially generated by AI.

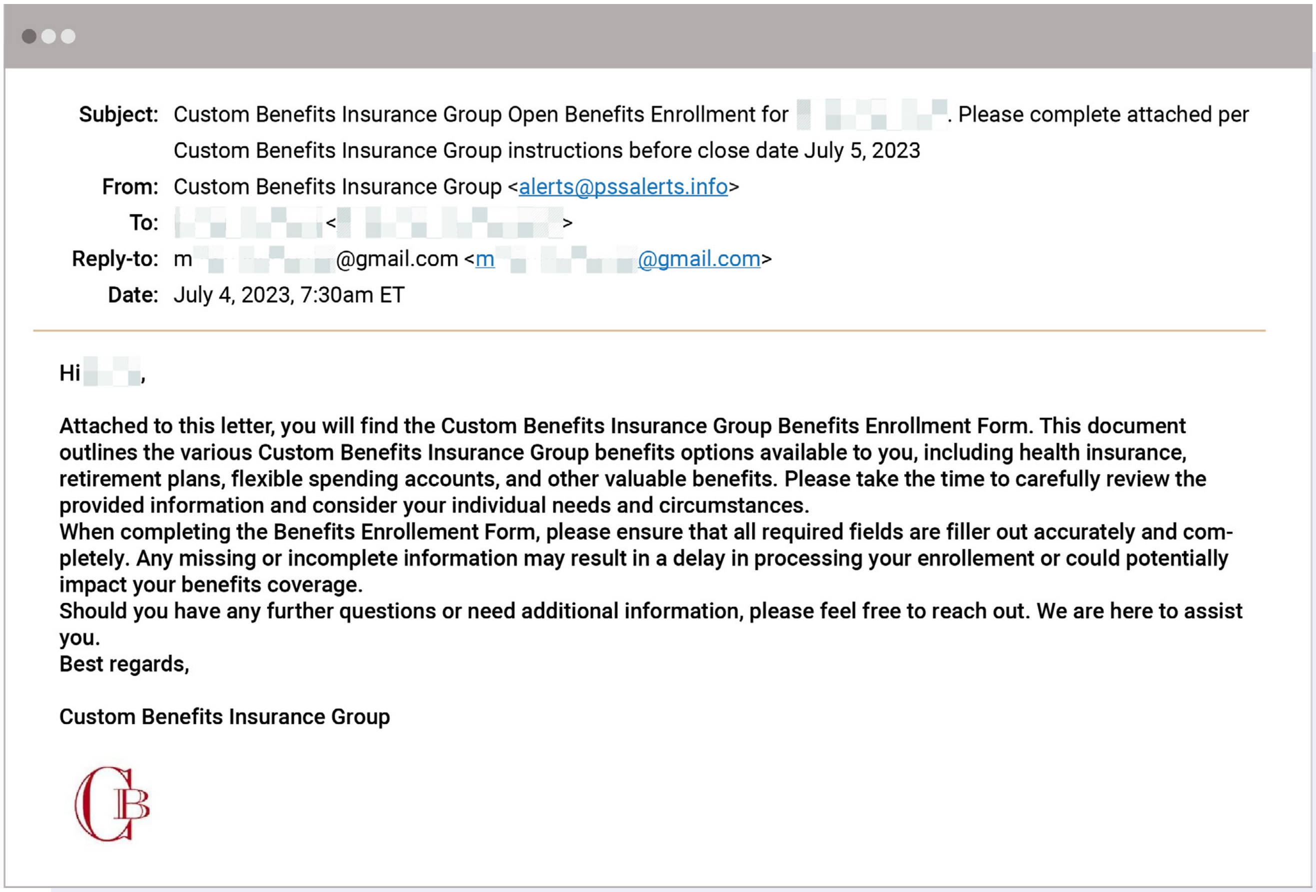
Giant Language Model Test Room, best known as GLTR, is then used to show this process visually. The models color-code the words based on how likely each word would be predicted given the context to the left. Green indicates a word is one of the top 10 predicted words while yellow indicates a top 100 predicted word. Words in red are ranked among the top 1,000 predicted words, with all other words shown in purple.

Throughout this past year, Abnormal identified thousands of attacks as likely AI-generated. The following five are some of the most interesting—proving how threat actors are exploiting AI and showing how the Abnormal platform stays one step ahead of attackers.

01 Attacker Poses as Insurance Company to Attempt Malware Delivery

In this malware attack, the threat actor poses as an insurance representative and informs the recipient that the email attachment contains benefits information, as well as an enrollment form that must be completed in its entirety and returned. If the recipient fails to do so, they are told they may lose coverage.

The perpetrator uses a seemingly genuine display name (“Customer Benefits Insurance Group”) and sender email (“alerts@pssalerts[.]info”), but replies are redirected to a Gmail account controlled by the attacker. Despite a professional facade, the Abnormal platform determined the attachment likely contains malware, putting the recipient’s computer at risk of viruses and credential theft.



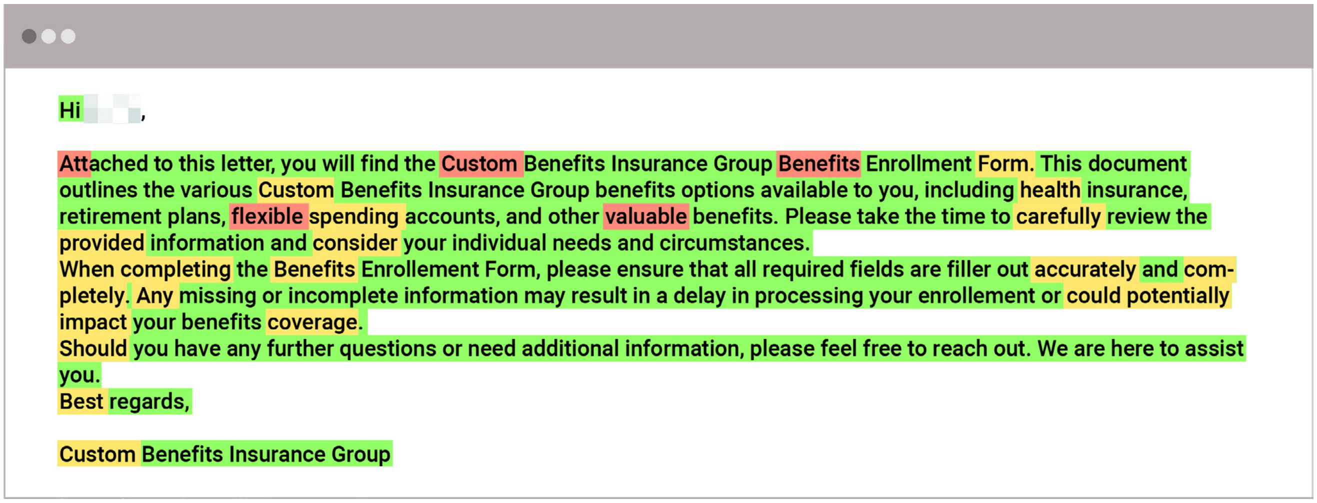
Subject: Custom Benefits Insurance Group Open Benefits Enrollment for [redacted]. Please complete attached per Custom Benefits Insurance Group instructions before close date July 5, 2023
From: Custom Benefits Insurance Group <alerts@pssalerts.info>
To: [redacted]<[redacted]>
Reply-to: m[redacted]@gmail.com <m[redacted]@gmail.com>
Date: July 4, 2023, 7:30am ET

Hi [redacted],

Attached to this letter, you will find the Custom Benefits Insurance Group Benefits Enrollment Form. This document outlines the various Custom Benefits Insurance Group benefits options available to you, including health insurance, retirement plans, flexible spending accounts, and other valuable benefits. Please take the time to carefully review the provided information and consider your individual needs and circumstances.
When completing the Benefits Enrollement Form, please ensure that all required fields are filler out accurately and completely. Any missing or incomplete information may result in a delay in processing your enrollement or could potentially impact your benefits coverage.
Should you have any further questions or need additional information, please feel free to reach out. We are here to assist you.
Best regards,

Custom Benefits Insurance Group

GLTR Analysis



Hi [redacted],

Attached to this letter, you will find the Custom Benefits Insurance Group Benefits Enrollment Form. This document outlines the various Custom Benefits Insurance Group benefits options available to you, including health insurance, retirement plans, flexible spending accounts, and other valuable benefits. Please take the time to carefully review the provided information and consider your individual needs and circumstances.
When completing the Benefits Enrollement Form, please ensure that all required fields are filler out accurately and completely. Any missing or incomplete information may result in a delay in processing your enrollement or could potentially impact your benefits coverage.
Should you have any further questions or need additional information, please feel free to reach out. We are here to assist you.
Best regards,

Custom Benefits Insurance Group

As you can see, the majority of the text is highlighted green, indicating that it was likely generated by AI rather than created by a human. You'll notice that there are also no typos or grammatical errors—signs that have historically been indicative of an attack.

What Makes This Attack Difficult to Detect

Due to the use of a known domain, the lack of malicious links, and legitimate-looking body content, this email has a high probability of bypassing legacy email security solutions.

GLTR's color-coding illustrates the likelihood of a word being generated by an AI model.

🟩 Top 10 Predicted Words

🟨 Top 100 Predicted Words

🟥 Top 1,000 Predicted Words

🟪 All Other Words

How Abnormal Stopped the Attack

Abnormal detects attacks by analyzing tens of thousands of unique signals to identify anomalies in email messages. In this case, Abnormal categorized this email as malicious due to the following factors:

Reply-To Email Mismatch: The reply-to email is different from the sender's email. This discrepancy is a red flag for the Abnormal detection models, as attackers often use this tactic to transition users to another email for continued conversation.

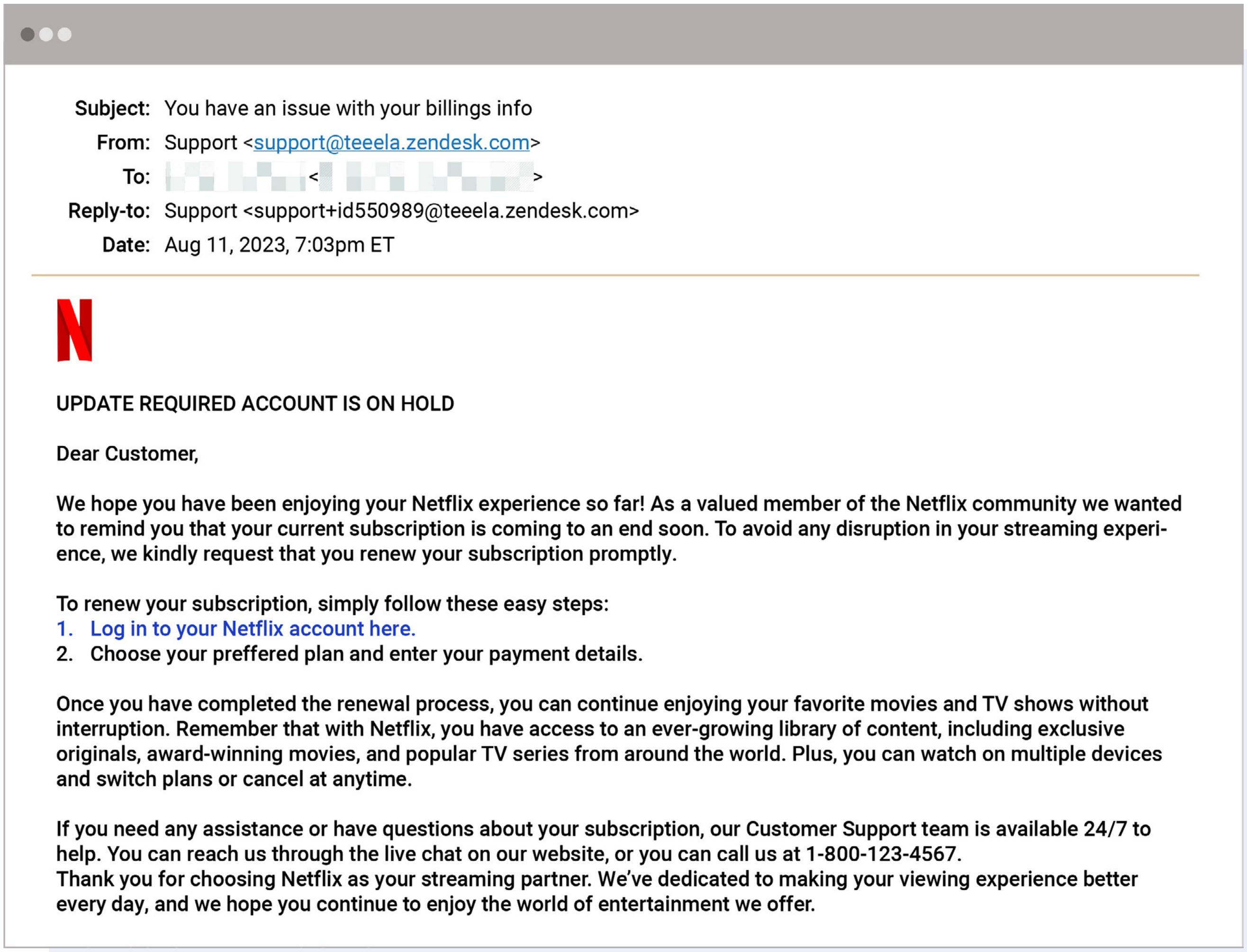
Domain Age: The sender's domain is only four months old. Since new domains are often created specifically to send attacks, Abnormal considers the age of the domain as a factor in determining the email's legitimacy.

Uncommon Attachment Type: Abnormal scans all attachments and flags those that deviate from the norm. In this instance, the recipient rarely receives emails with HTM attachments, making this email anomalous from known behavior.

02 Netflix Impersonator Compromises Legitimate Domain in Credential Phishing Attack

In this phishing attack, the threat actor poses as a customer service representative from Netflix and claims that the target's subscription is expiring. To continue service, the recipient is told they need to renew their subscription using the provided link. However, the URL leads to a malicious site where sensitive information is at risk.

The attacker employs social engineering to create a sense of urgency. They also add sophistication to the scam by leveraging what appears to be an authentic helpdesk domain associated with Teeela, an online toy shopping app. The use of an email hosted on Zendesk, a trusted customer support platform, may deceive recipients into thinking the email is legitimate and thus increase the attack's effectiveness.



Subject: You have an issue with your billings info
From: Support <support@teeela.zendesk.com>
To: [redacted]
Reply-to: Support <support+id550989@teeela.zendesk.com>
Date: Aug 11, 2023, 7:03pm ET

UPDATE REQUIRED ACCOUNT IS ON HOLD

Dear Customer,

We hope you have been enjoying your Netflix experience so far! As a valued member of the Netflix community we wanted to remind you that your current subscription is coming to an end soon. To avoid any disruption in your streaming experience, we kindly request that you renew your subscription promptly.

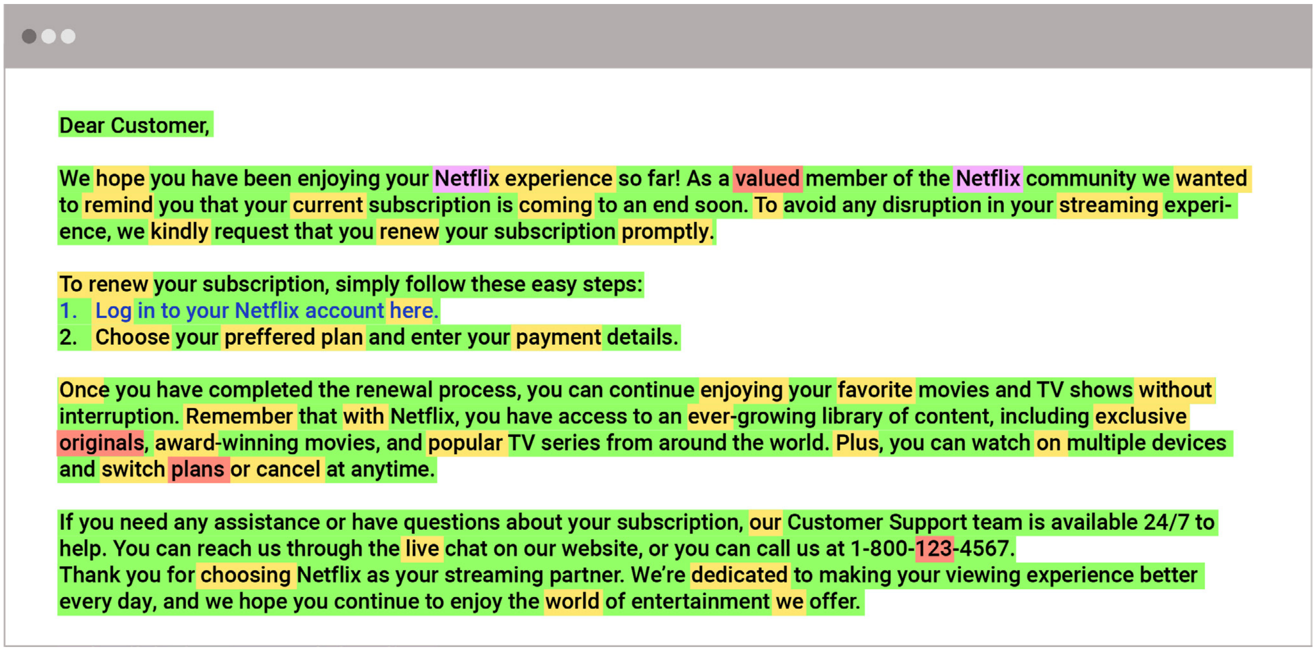
To renew your subscription, simply follow these easy steps:

1. **Log in to your Netflix account here.**
2. **Choose your preffered plan and enter your payment details.**

Once you have completed the renewal process, you can continue enjoying your favorite movies and TV shows without interruption. Remember that with Netflix, you have access to an ever-growing library of content, including exclusive originals, award-winning movies, and popular TV series from around the world. Plus, you can watch on multiple devices and switch plans or cancel at anytime.

If you need any assistance or have questions about your subscription, our Customer Support team is available 24/7 to help. You can reach us through the live chat on our website, or you can call us at 1-800-123-4567.
Thank you for choosing Netflix as your streaming partner. We've dedicated to making your viewing experience better every day, and we hope you continue to enjoy the world of entertainment we offer.

GLTR Analysis



Dear Customer,

We hope you have been enjoying your Netflix experience so far! As a valued member of the Netflix community we wanted to remind you that your current subscription is coming to an end soon. To avoid any disruption in your streaming experience, we kindly request that you renew your subscription promptly.

To renew your subscription, simply follow these easy steps:

1. Log in to your Netflix account here.
2. Choose your preffered plan and enter your payment details.

Once you have completed the renewal process, you can continue enjoying your favorite movies and TV shows without interruption. Remember that with Netflix, you have access to an ever-growing library of content, including exclusive originals, award-winning movies, and popular TV series from around the world. Plus, you can watch on multiple devices and switch plans or cancel at anytime.

If you need any assistance or have questions about your subscription, our Customer Support team is available 24/7 to help. You can reach us through the live chat on our website, or you can call us at 1-800-123-4567.
Thank you for choosing Netflix as your streaming partner. We're dedicated to making your viewing experience better every day, and we hope you continue to enjoy the world of entertainment we offer.

Again, the majority of the text is highlighted green, indicating that it was likely generated by AI, rather than created by a human. One thing of interest here is that the AI review highlighted the interesting phone number, where the attacker seems to have missed updating the phone number to a legitimate one.

What Makes This Attack Difficult to Detect

This email was sent from a compromised address on a legitimate domain, contains no attachments, and leverages social engineering tactics rather than exploiting technical vulnerabilities. As a result, it is likely to elude detection by a traditional email security platform.

How Abnormal Stopped the Attack

Modern, AI-native security solutions like Abnormal analyze the sender, content, and links for signs of an advanced attack. In this case, Abnormal categorized this email as malicious due to the following factors:

Link Analysis: The attacker attempted to obscure the destination of the phishing link with a URL shortener, but Abnormal is still able to analyze the link and determine that it leads to a malicious website where sensitive data is at risk.

Unusual Sender Domain: Abnormal compares the domains of links contained in the email with the sender domain. Because the included link was not hosted on the same domain as the sender's address, Abnormal flagged this as suspicious.

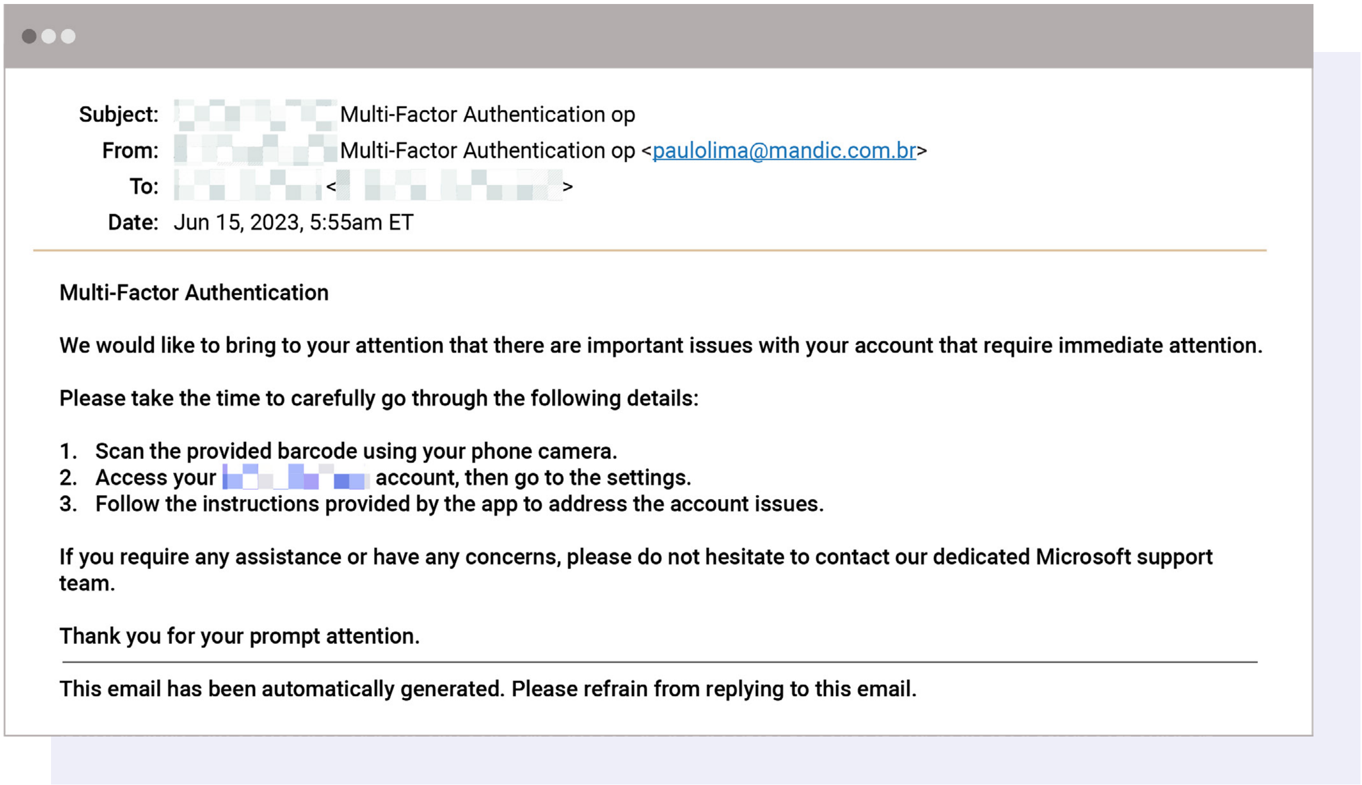
Content Analysis: The detection engine analyzes the content of every email for signs of phishing or other social engineering techniques. The use of urgent language and request for payment details in this email indicates a phishing attempt.

03

Threat Actor Aims to Steal Credentials via QR Code Phishing Attack

In this QR code phishing attack, the threat actor impersonates an internal department within the target's own company and informs the recipient of urgent but unspecified issues with their account. To resolve the issues, the target is told to scan an attached QR code and follow the instructions contained within the email.

In reality, the attacker is trying to gain login credentials from the recipient under the guise of setting up multi-factor authentication (MFA). By applying social engineering tactics and presenting the email as sent from a no-reply address, the threat actor increases the chances of the recipient immediately scanning the malicious QR code instead of pursuing alternate means of resolution.



Subject: [redacted] Multi-Factor Authentication op
From: [redacted] Multi-Factor Authentication op <paulolima@mandic.com.br>
To: [redacted] <[redacted]>
Date: Jun 15, 2023, 5:55am ET

Multi-Factor Authentication

We would like to bring to your attention that there are important issues with your account that require immediate attention.

Please take the time to carefully go through the following details:

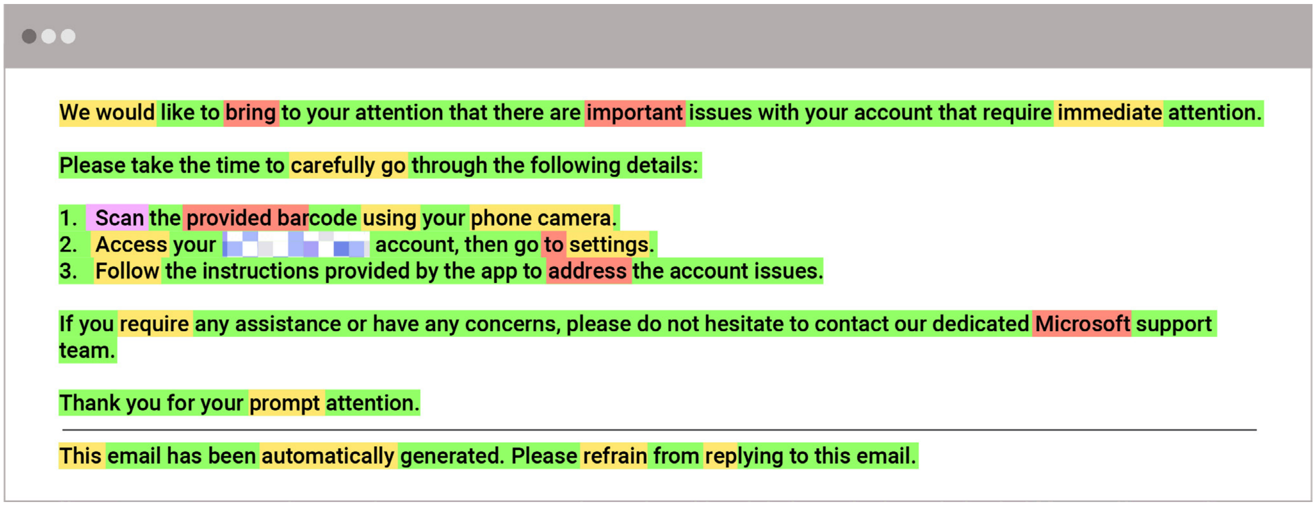
1. Scan the provided barcode using your phone camera.
2. Access your [redacted] account, then go to the settings.
3. Follow the instructions provided by the app to address the account issues.

If you require any assistance or have any concerns, please do not hesitate to contact our dedicated Microsoft support team.

Thank you for your prompt attention.

This email has been automatically generated. Please refrain from replying to this email.

Attached to the malicious email is a document that contains only the QR code, with no further information or instructions. When scanned, the QR code redirects recipients to a page designed to steal login credentials.



We would like to bring to your attention that there are important issues with your account that require immediate attention.

Please take the time to carefully go through the following details:

1. Scan the provided barcode using your phone camera.
2. Access your [redacted] account, then go to settings.
3. Follow the instructions provided by the app to address the account issues.

If you require any assistance or have any concerns, please do not hesitate to contact our dedicated Microsoft support team.

Thank you for your prompt attention.

This email has been automatically generated. Please refrain from replying to this email.

The majority of the text is highlighted green, indicating that it was likely generated by AI, rather than created by a human. Again, there are no grammatical or spelling errors to indicate an attack.

What Makes This Attack Difficult to Detect

Legacy email security tools have trouble detecting this as an attack because it was sent from a spoofed sender address, contains legitimate-looking content, and uses a payload that is not immediately identifiable as malicious. The use of the QR code is a fairly new technique used to bypass security measures, as many legacy systems cannot determine the resulting link by scanning the code.

QR code phishing or "quishing" attacks have become increasingly common as the use of QR codes has gained popularity. Attacks that contain QR codes often bypass traditional solutions because they are unable to parse the code to determine the final destination of the link.

How Abnormal Stopped the Attack

Abnormal applies advanced behavioral AI and machine learning to identify anomalies in email communication that may indicate fraudulent behavior. In this case, Abnormal categorized this email as malicious due to the following factors:

Unknown Sender and Domain: The email is sent from a domain from which this company has never received prior emails. The platform tracks and flags unknown domains and emails as potentially malicious.

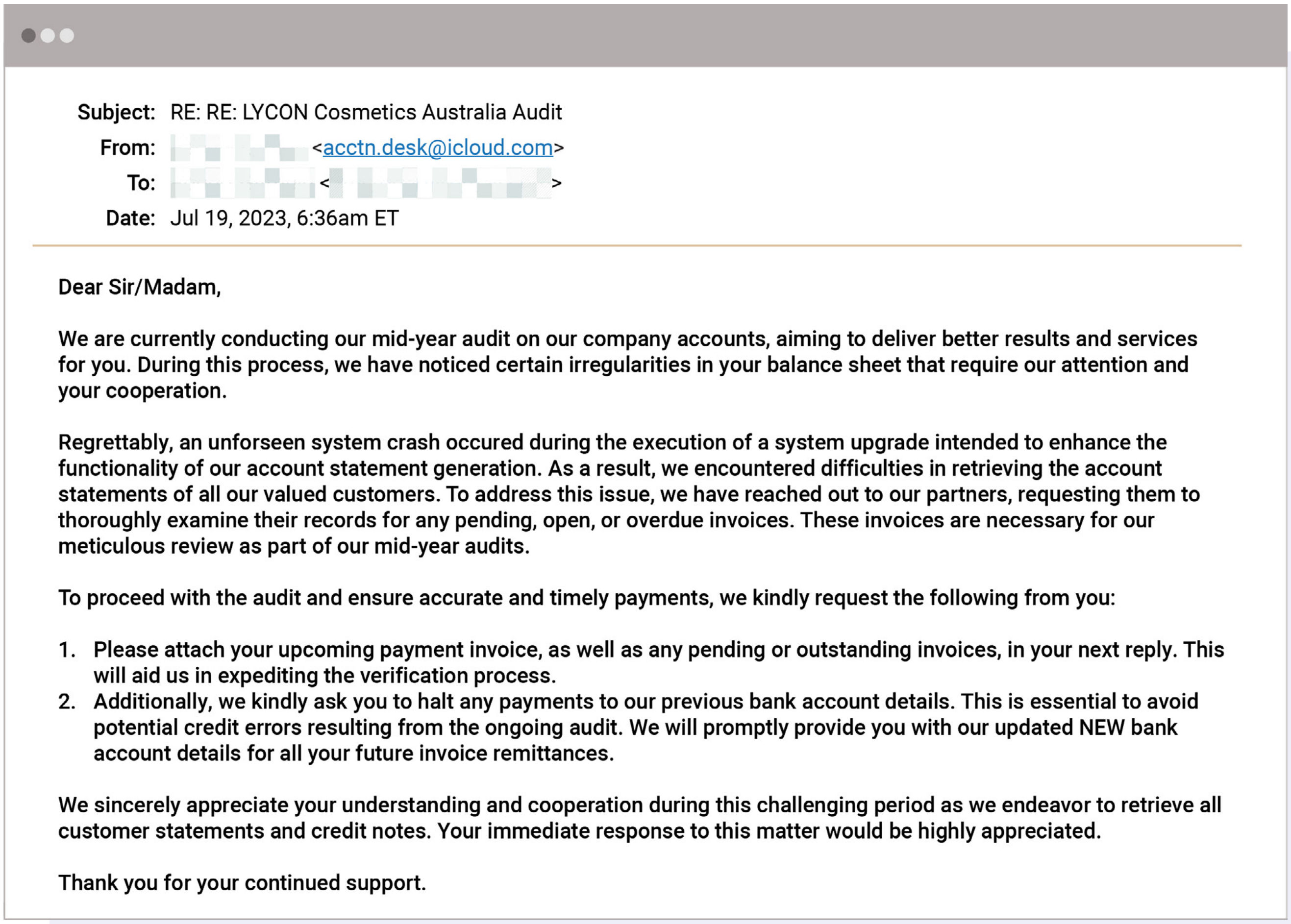
Suspicious Email Content: The platform uses natural language processing (NLP) to understand the context and intent of every email. In this case, it recognized the pattern of a phishing attack disguised as a multi-factor authentication (MFA) operation.

Attachment Analysis: Abnormal analyzes all attachments for potential threats. Even though QR codes are common and usually harmless, the system detected that the QR code attached to this email leads to a malicious login page.

04 Cosmetics Brand Impersonator Attempts Invoice Fraud

In this billing account update attempt, the attacker poses as a business development manager for cosmetics company LYCON and informs the recipient of irregularities in their balance sheet noticed during a mid-year audit. They explain that due to a crash during a system upgrade, they no longer have access to account statements and must now request all open or overdue invoices. The attacker also advises a halt to payments to previous accounts and promises new banking details for future transactions once the audit is complete.

The scam aims to extract sensitive financial information and reroute payments to the attacker's bank account. No links or attachments are present, and the email is written in an official tone, utilizing several social engineering techniques to deceive the recipient.



Subject: RE: RE: LYCON Cosmetics Australia Audit
From: [redacted] <acctn.desk@icloud.com>
To: [redacted] <[redacted]>
Date: Jul 19, 2023, 6:36am ET

Dear Sir/Madam,

We are currently conducting our mid-year audit on our company accounts, aiming to deliver better results and services for you. During this process, we have noticed certain irregularities in your balance sheet that require our attention and your cooperation.

Regrettably, an unforseen system crash occured during the execution of a system upgrade intended to enhance the functionality of our account statement generation. As a result, we encountered difficulties in retrieving the account statements of all our valued customers. To address this issue, we have reached out to our partners, requesting them to thoroughly examine their records for any pending, open, or overdue invoices. These invoices are necessary for our meticulous review as part of our mid-year audits.

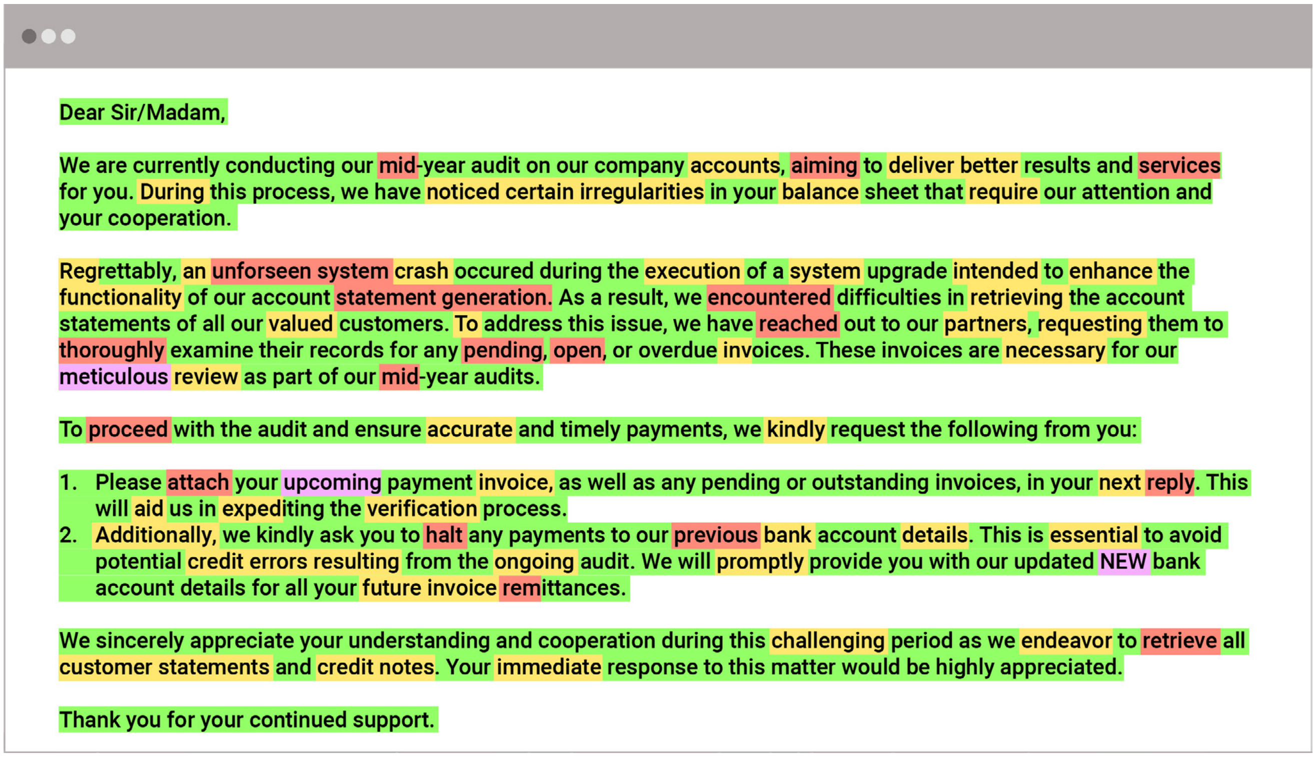
To proceed with the audit and ensure accurate and timely payments, we kindly request the following from you:

1. Please attach your upcoming payment invoice, as well as any pending or outstanding invoices, in your next reply. This will aid us in expediting the verification process.
2. Additionally, we kindly ask you to halt any payments to our previous bank account details. This is essential to avoid potential credit errors resulting from the ongoing audit. We will promptly provide you with our updated NEW bank account details for all your future invoice remittances.

We sincerely appreciate your understanding and cooperation during this challenging period as we endeavor to retrieve all customer statements and credit notes. Your immediate response to this matter would be highly appreciated.

Thank you for your continued support.

GLTR Analysis



Dear Sir/Madam,

We are currently conducting our mid-year audit on our company accounts, aiming to deliver better results and services for you. During this process, we have noticed certain irregularities in your balance sheet that require our attention and your cooperation.

Regrettably, an unforseen system crash occured during the execution of a system upgrade intended to enhance the functionality of our account statement generation. As a result, we encountered difficulties in retrieving the account statements of all our valued customers. To address this issue, we have reached out to our partners, requesting them to thoroughly examine their records for any pending, open, or overdue invoices. These invoices are necessary for our meticulous review as part of our mid-year audits.

To proceed with the audit and ensure accurate and timely payments, we kindly request the following from you:

1. Please attach your upcoming payment invoice, as well as any pending or outstanding invoices, in your next reply. This will aid us in expediting the verification process.
2. Additionally, we kindly ask you to halt any payments to our previous bank account details. This is essential to avoid potential credit errors resulting from the ongoing audit. We will promptly provide you with our updated NEW bank account details for all your future invoice remittances.

We sincerely appreciate your understanding and cooperation during this challenging period as we endeavor to retrieve all customer statements and credit notes. Your immediate response to this matter would be highly appreciated.

Thank you for your continued support.

Once again the majority of the text is highlighted green. While there is more red included than in previous examples, this is likely due to the more formal language.

What Makes This Attack Difficult to Detect

Because the email was sent from a legitimate email service provider, is entirely text-based, and relies on social engineering to compel the recipient to take action, it would be challenging for traditional email security solutions to detect this attack.

How Abnormal Stopped the Attack

Using AI, Abnormal establishes a baseline for normal behavior and then detects and blocks activity that deviates from this baseline. In this case, Abnormal categorized this email as malicious due to the following factors:

Email Origin: The email originates from an iCloud account, which is not a typical business email domain. Abnormal recognizes this as potentially suspicious, particularly when an email discusses business matters like audits and invoices.

Unusual Request: The email asks the recipient to halt payments to a previous bank account and promises to provide new account details. Due to the financial nature of this request, Abnormal flags this as a possible indicator of attack.

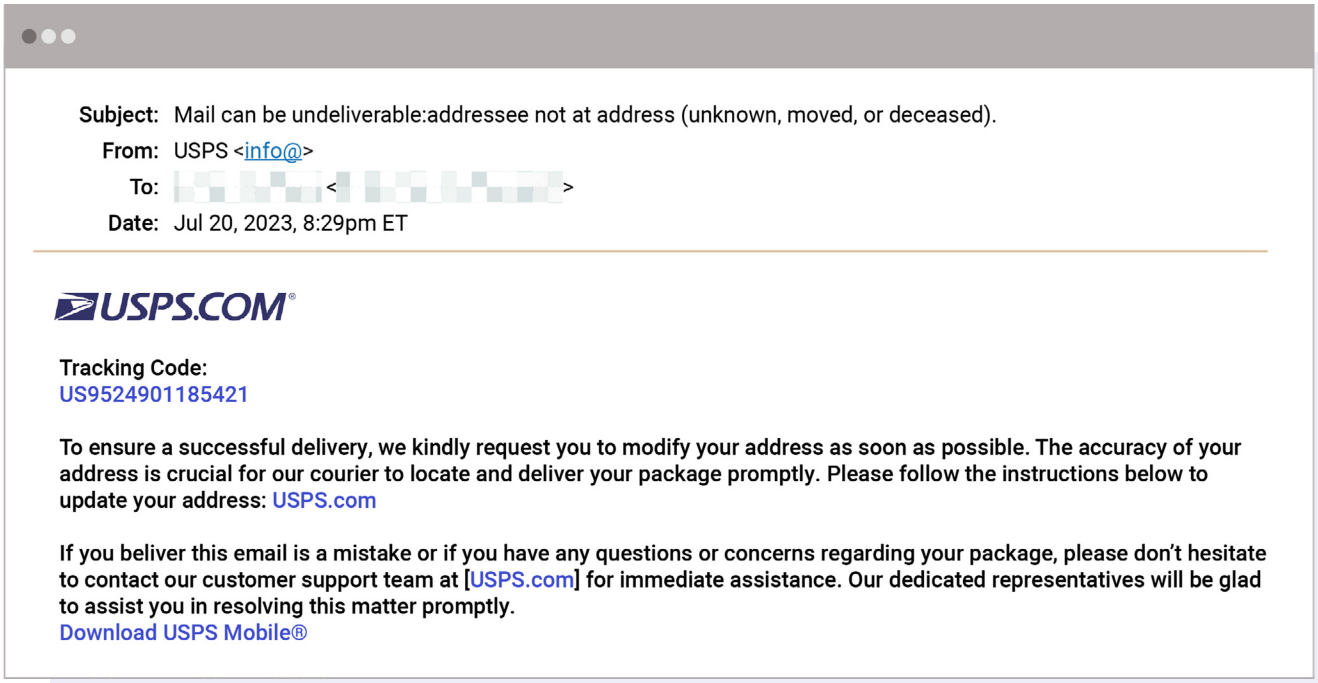
Business Context Analysis: Abnormal scans emails for content related to financial topics, including invoices and bank information. Because the sender is unknown to this recipient and the email content is financial in nature, Abnormal flags this as suspicious.

05

USPS Impersonator Seeks Credit Card Information in Multi-Layered Attack

In this attack, the threat actor impersonates the United States Postal Service (USPS) and sends what appears to be an automated notification about a package delivery issue. The threat actor informs the target they can use the provided link to update their address so that the courier can successfully deliver the package.

By exploiting the urgency associated with a delayed or potentially canceled shipment and posing as a known entity, the attacker aims to compel the recipient to click the link and enter sensitive information. However, the URL leads to a fraudulent USPS page on the domain "allesoverecommerce[.]com"—where personal information or credit card details are at risk of being stolen if entered into the form. What makes this attack especially convincing is the phishing page that appears to be an exact duplicate of a real USPS tracking page.



Subject: Mail can be undeliverable:addressee not at address (unknown, moved, or deceased).
From: USPS <info@>
To: [redacted] <[redacted]>
Date: Jul 20, 2023, 8:29pm ET

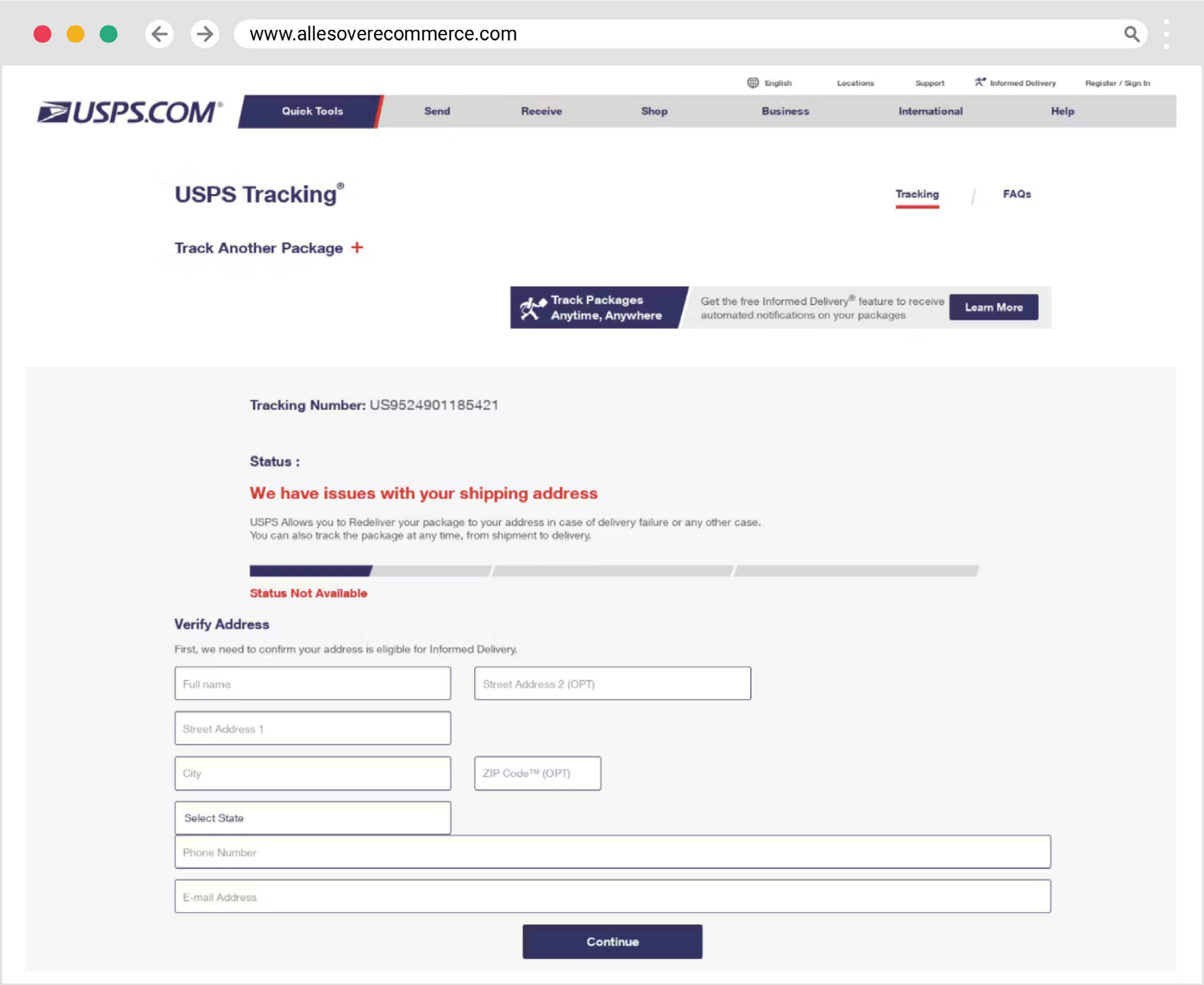
USPS.COM®

Tracking Code:
US9524901185421

To ensure a successful delivery, we kindly request you to modify your address as soon as possible. The accuracy of your address is crucial for our courier to locate and deliver your package promptly. Please follow the instructions below to update your address: USPS.com

If you beliver this email is a mistake or if you have any questions or concerns regarding your package, please don't hesitate to contact our customer support team at [USPS.com] for immediate assistance. Our dedicated representatives will be glad to assist you in resolving this matter promptly.
Download USPS Mobile®

Clicking on the link in the email redirects targets to this sophisticated phishing page, where the target is encouraged to enter their personal information. Despite showing in the text as USPS.com, this credential phishing page is actually hosted on “allesoverecommerce[.]com” where attackers are counting on end users not noticing the redirect.



USPS Tracking®

Track Another Package +

Track Packages Anytime, Anywhere

Get the free Informed Delivery® feature to receive automated notifications on your packages

Tracking Number: US9524901185421

Status :

We have issues with your shipping address

USPS Allows you to Redeliver your package to your address in case of delivery failure or any other case.
You can also track the package at any time, from shipment to delivery.

Status Not Available

Verify Address

First, we need to confirm your address is eligible for Informed Delivery.

Full name

Street Address 2 (OPT)

Street Address 1

City

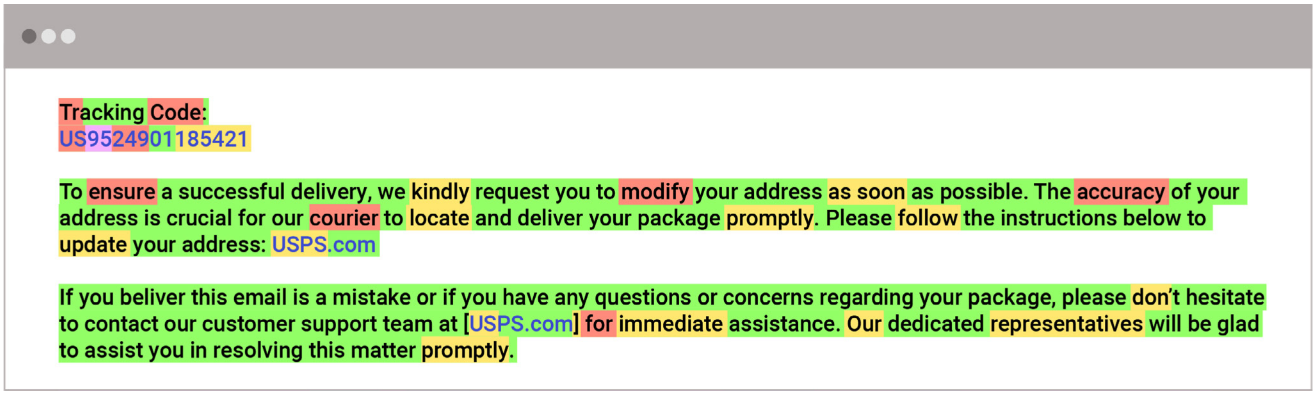
ZIP Code™ (OPT)

Select State

Phone Number

E-mail Address

GLTR Analysis



Tracking Code:
US9524901185421

To ensure a successful delivery, we kindly request you to modify your address as soon as possible. The accuracy of your address is crucial for our courier to locate and deliver your package promptly. Please follow the instructions below to update your address: USPS.com

If you beliver this email is a mistake or if you have any questions or concerns regarding your package, please don't hesitate to contact our customer support team at [USPS.com] for immediate assistance. Our dedicated representatives will be glad to assist you in resolving this matter promptly.

The majority of the text is highlighted green, indicating that it was likely generated by AI, rather than created by a human. One thing to note here is that the attacker asks the target to follow instructions, but then simply enters a link, with no further explanation given—prompting them to click on that link to understand the next steps.

What Makes This Attack Difficult to Detect

Legacy email security solutions have difficulty flagging and blocking this attack because it contains legitimate-looking content, is sent from an unknown domain, and aims to manipulate the recipient via social engineering—an element these tools can't detect. bypass security measures, as many legacy systems cannot determine the resulting link by scanning the code.

How Abnormal Stopped the Attack

By leveraging behavioral AI and ML to understand what is normal, Abnormal can block the email threats that bypass other solutions. In this case, Abnormal categorized this email as malicious due to the following factors:

Link Analysis: Abnormal analyzes all links within the body of an email. While some of the links in this message do lead to the real USPS site, the two in the body copy instead lead to a phishing page on "allesoverecommerce[.]com."

Content Analysis: The platform analyzes the content of every email for signs of phishing or other social engineering techniques. The use of urgent language and a request for sensitive information in this email indicates a phishing attempt.

Name Impersonation: While the sender's display name and signature match a known brand (USPS), the email address does not appear to be associated with USPS. Abnormal flags this as suspicious.

An Increasing Threat for 2024: How AI Will Change the Email Threat Landscape

As you can see in these examples, generative AI makes email attacks much harder to detect by the human eye, resulting in an increased need for security tools that can use multiple signals to detect anomalous activity. Despite the fact that generative AI has only been used widely for a year, it is obvious that the potential is there for widespread abuse. For security leaders, this is a wakeup call to prioritize cybersecurity measures to safeguard against these threats before it is too late.

Collaboration in the cybersecurity community—including ongoing research into AI transparency, ethical AI practices, and the development of resilient cybersecurity frameworks—will be crucial to addressing the risks associated with the rapid advancements in generative AI technologies.

Throughout the coming years, we predict AI will continue to advance in the following areas:

Advancements in Model Architectures: Continued progress in generative AI is likely to involve the development of more advanced model architectures. Researchers may explore variations of existing models like GPT or introduce entirely new architectures, both of which can then be used for cybercrime.

Fine-Tuning and Specialization: Generative models may become more specialized for specific tasks or domains. Fine-tuning pre-existing models for applications such as art generation, content creation, or specific industries could become more prevalent. By the same token, bad actors could fine tune their models to automatically generate and send millions of personalized email attacks each day, utilizing these tools to scrape LinkedIn profiles and create realistic-looking socially-engineered attacks at massive scale.

Increased Realism in Output: Expect improvements in the realism of generated content, including text, images, and even audio. Enhanced capabilities may lead to more convincing deepfakes or synthetic media, raising concerns about the potential misuse of such technology. Put these capabilities into the hands of bad actors, and you have an environment where targeted voice phishing and similar advanced techniques become commonplace.

So what do security leaders need to know moving forward? **AI is here to stay, and we must all be prepared to defend against it.** The attacks shown here are well-executed, but they are only the beginning of what is possible. We've reached a point where only AI can stop AI, and where preventing these attacks and their next-generation counterparts requires using AI-native defenses.

Conclusion

The emergence of generative AI has introduced both promise and peril. There is no denying that generative AI and other forms of artificial intelligence have provided modern businesses with several benefits.

But as these attacks demonstrate, its widespread adoption has also enabled cybercriminals to craft more complex email attacks at scale—without the grammatical and syntax errors of the past. These attacks are increasingly difficult for legacy solutions to detect and block due to their lack of traditional indicators of compromise, resulting in them reaching end user inboxes. Once there, humans struggle to determine whether an email is legitimate or malicious and often choose poorly—allowing attackers to steal credentials, access sensitive data, and reroute payments into their own accounts.

The email attacks of 2023 were more advanced than those of the past, but cybercriminals won't stop there. They will continue to experiment with generative AI to create their attacks and find ways to infiltrate organizations—especially those without the AI-powered defenses needed to thwart them. **Fortunately, it is still possible to win the AI arms race, provided that security leaders act now to prevent these threats.**

Abnormal

Abnormal Security provides the leading behavioral AI-based email security platform that leverages machine learning to stop sophisticated inbound email attacks and dangerous email platform attacks that evade traditional solutions. The anomaly detection engine leverages identity and context to analyze the risk of every cloud email event, preventing inbound email attacks, detecting compromised accounts, and remediating emails and messages—all while providing visibility into configuration drifts across your environment. You can deploy Abnormal in minutes with an API integration for Microsoft 365 or Google Workspace and experience the full value of the platform instantly, with additional protection available for Slack, Teams, and Zoom.

Interested in Stopping AI-Generated Email Attacks?

Request a Demo →

See Your ROI →