COFENSE

The New Era of Phishing: **THREATS BUILT IN THE AGE OF AI**

Introduction

Artificial intelligence (AI) has accelerated the speed and sophistication of modern phishing threats, overwhelming security teams with high-volume and high-variance attacks. Threat actors no longer experiment with AI in isolated ways. Instead, they use it as a core capability to generate, test, and deploy phishing campaigns at scale. AI enables attackers to rapidly refine language, adjust delivery methods, and iterate on campaign elements. The result is phishing that is faster, more adaptive, and more convincing than ever before, giving rise to polymorphic, multi-channel campaigns that continuously change their appearance while preserving the same malicious intent. These attacks evade perimeter defenses with increasing success and reach user inboxes at scale.

Every day, Cofense analysts leverage AI alongside human expertise to examine real-world phishing reports from millions of trained users, creating a phishing-specific dataset that captures how attack tactics evolve over time. Investigation of this data reveals consistent changes in phishing behavior, including how actors scale campaigns, improve personalization, and blend seamlessly into legitimate business workflows. The trends outlined in this report examine how phishing infrastructure, delivery models, social engineering techniques, payloads, and domain usage continue to evolve. Together, these patterns demonstrate why phishing must be analyzed after delivery, where behavioral context and human validation expose threats that evade static, perimeter-based controls.

By the Numbers

The Cofense Phishing Defense Center saw, on average, **1 malicious email every 19 seconds throughout 2025.**

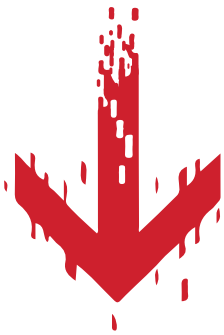
There was a **204%** **increase in phishing emails delivering malware,** compared to 2024.

Legitimate remote access tool abuse increased almost **900%** by volume from 2024 to 2025.

Campaigns spoofing the Social Security Administration increased **372%** in 2025.

In 2025, **76%** of initial infection URLs identified were unique and not seen across other customers, yet **94%** reused the same infrastructure.

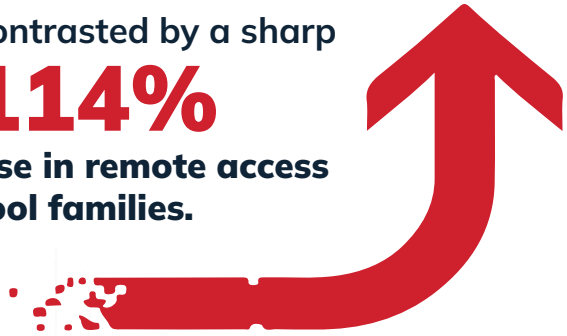
In 2025, **18%** of malicious emails relied on plain text conversations rather than URLs or attachments, ranking as the **second-highest** threat vector.



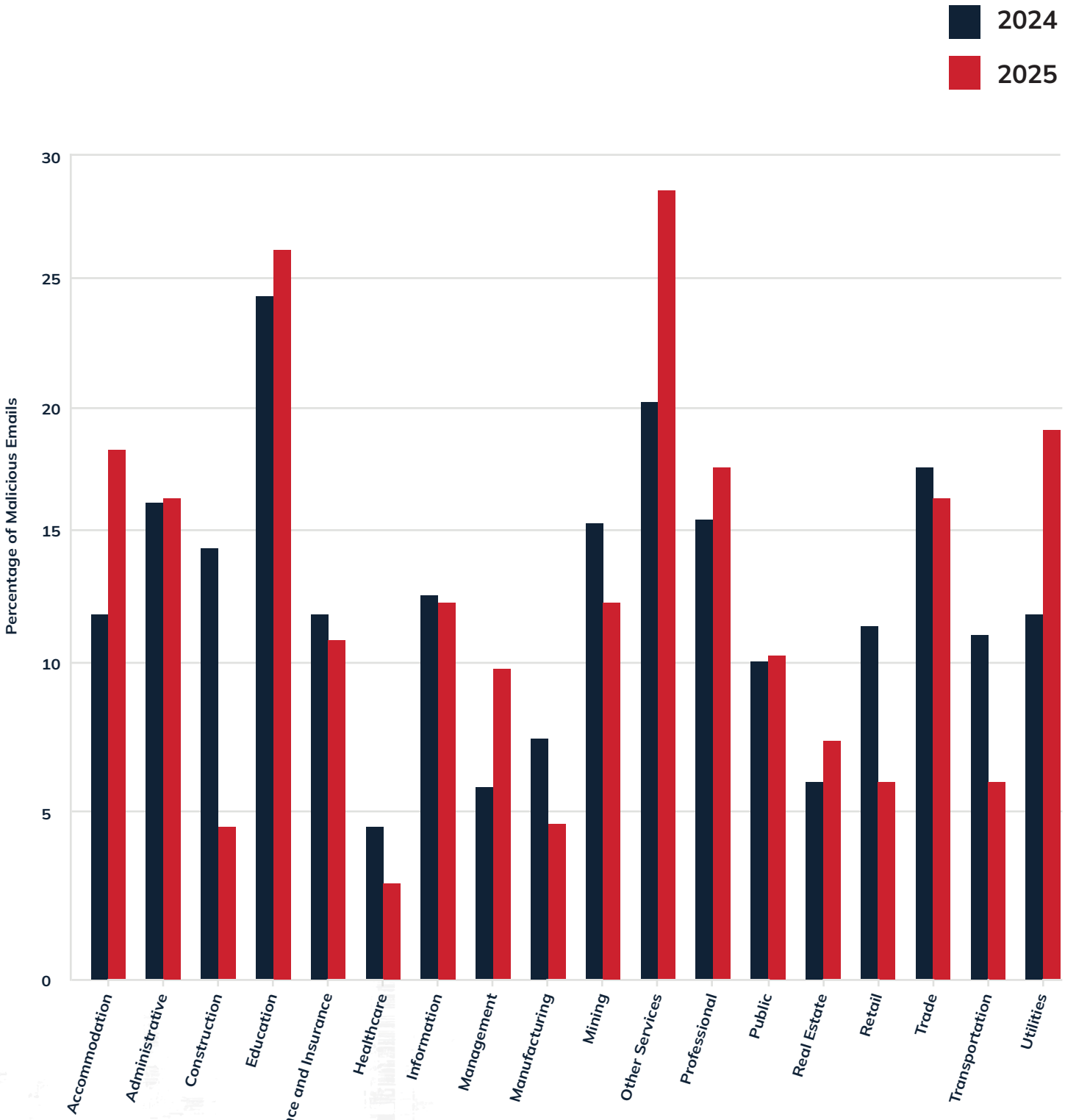
In 2025, the Cofense Phishing Defense Center identified a **27%** **reduction in overall malware family diversity,**

ClickFix attacks increased **4.7 times** between 2024 and 2025.

contrasted by a sharp **114%** **rise in remote access tool families.**



Percentage Change in Malicious Emails Bypassing Email Security Perimeters Year Over Year



Top Trends to Watch

Trend #1: Phishing Pages Grow More Adaptive and Analysis-Aware

Cofense Intelligence identified a significant rise in malware campaigns that dynamically adjust what they deliver based on the victim's browser and operating system. In multiple cases, the same phishing website delivered different payloads depending on how it was accessed. When visited from a Windows system, the site downloaded a Windows executable that installed a ConnectWise remote access tool (RAT). When accessed from a Mac, the same site delivered a macOS package of ConnectWise RAT that communicated with the same command-and-control infrastructure.

This adaptive behavior has expanded beyond desktop systems. In 2025, Cofense Intelligence observed more Android application package files delivered through phishing campaigns than in the previous three years combined. Many of these Android files appear to originate from kits traditionally used to distribute Windows-based ConnectWise RAT or GoTo RAT payloads, indicating that threat actors are repurposing existing infrastructure to target additional platforms more efficiently. The kits used to deliver the Android application package files have the same general themes, such as impersonation of Adobe Document Cloud, Microsoft Teams, or Zoom, but often feature differences in the exact content displayed, a hallmark of AI-generated content.

Credential phishing pages are also becoming more sophisticated. While browser-based customization and analysis detection were once relatively uncommon, their use increased significantly throughout 2025, nearly doubling between the first and second halves of the year. Earlier examples typically involved minor cosmetic changes, such as rearranged input fields. More recent campaigns show advanced behavior, including presenting different spoofed brands depending on the browser or optimizing credential harvesting pages specifically for mobile users. Threat actors continue to deploy complex, advanced, and targeted phishing pages at a faster rate than human developers can support. This fact, along with other indicators such as excessive commenting of code, indicates that threat actors are likely incorporating AI more directly into the creation and deployment of phishing kits.

Threat actors are increasingly combining these adaptive phishing pages with operating system-specific malware delivery to maximize the value of each visit. These sites often collect detailed information about the visitor, including browser plugins, language settings, geographic location, and screen size. This data allows attackers to tailor both the phishing experience and the malware payload to the individual user, increasing the likelihood of success.

In addition, the same detection techniques are being used to evade analysis. Phishing pages may redirect visitors to legitimate websites when common security tools are detected or display error pages when user behavior does not match expected device characteristics. This approach builds on earlier tactics such as CAPTCHA checks and, if successful, is likely to become a standard component of phishing toolkits going forward.

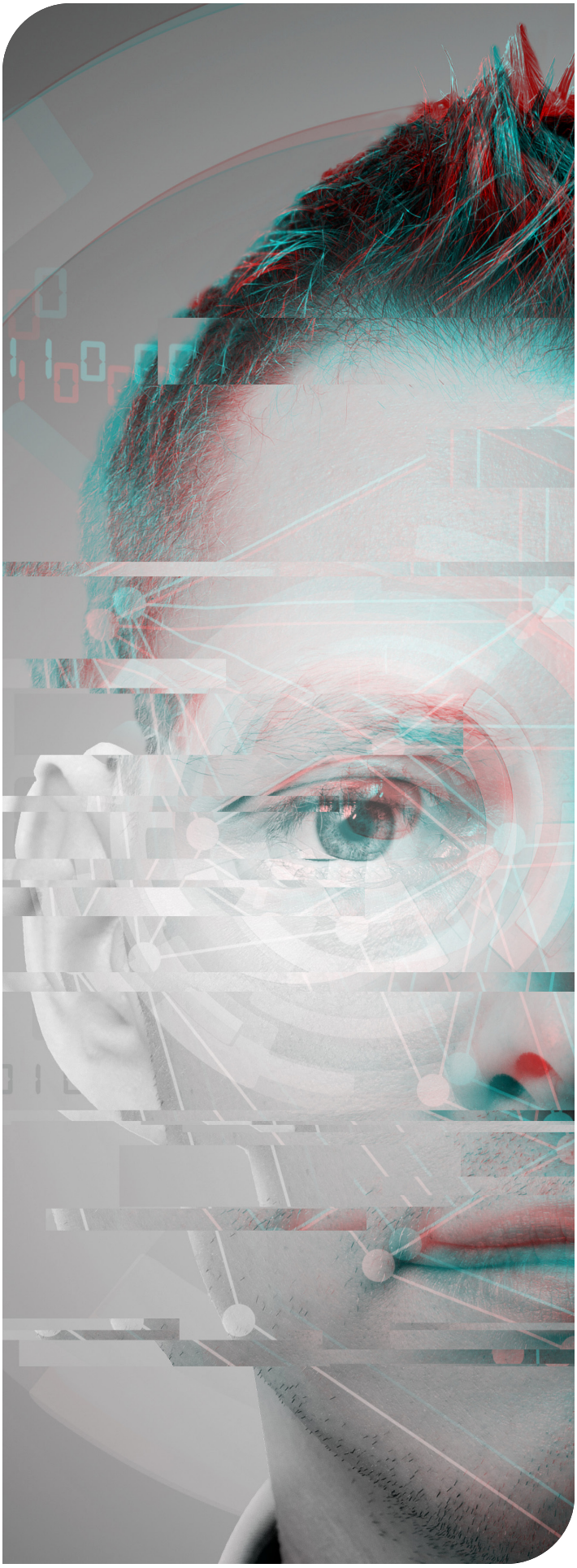
Trend #2: Polymorphic Phishing Becomes the Default Delivery Model

The growing use of AI throughout the phishing process has enabled threat actors to launch mass-produced attacks that are also highly unique and polymorphic. Attackers are no longer limited to sending a single, generic phishing template to large numbers of users. With AI, they can rapidly generate countless variations of the same campaign, each customized with different logos, signatures, wording, or even entirely different URLs and files used to deliver the malicious activity. In many cases, these variations are tailored to individual victims, making each phishing email appear distinct.

Data from the Cofense Phishing Defense Center highlights the scale of this shift. In 2025, analysts observed that 76 percent of initial infection URLs identified in phishing attacks were unique and had not appeared in any other campaigns across the customer base. However, of those initial infection URLs, 94 percent of their IP addresses were seen multiple times. This is an indication of threat actors executing multiple phishing campaigns utilizing the same infrastructure. Similarly, 82 percent of malicious files identified had unique file hashes, even if they contained the same malicious payloads. The ability to create and manage such a high volume of unique indicators of compromise is only feasible through the use of generative AI, and it significantly reduces the effectiveness of traditional detection methods.

This level of diversification presents challenges not only for security tools that rely on pattern recognition, but also for employees tasked with identifying suspicious emails. AI-driven phishing tools now actively collect publicly available information about users and their organizations to personalize attacks. Threat actors routinely incorporate details such as home addresses, phone numbers, professional connections, organizational charts, and recent social media announcements to make messages appear legitimate and relevant. As these tactics become more advanced, organizations and individuals must reassess what information they share publicly and take steps to limit the data that can be exploited for malicious purposes.

Trend #3: AI Makes Business Email Compromise More Efficient and Convincing



Generative AI has had a profound impact on phishing defense, and its influence is especially clear in the evolution of Business Email Compromise (BEC) attacks. Unlike many other forms of phishing, most BEC emails do not rely on malicious URLs, file attachments, or QR codes. Instead, they use simple, conversational social engineering to persuade recipients to reply directly and provide sensitive information or take action.

In 2025, conversational attacks were the second-most-common threat vector observed by Cofense, accounting for 18 percent of all malicious emails identified. These attacks depend entirely on user interaction, which makes them particularly difficult for traditional security systems to detect and remediate. Many email security tools are designed to analyze links or attachments to determine whether a message is malicious. While some solutions attempt to identify BEC emails using language analysis, they often struggle because the content can be either extremely generic or highly specific to the organization. In many cases, the difference between a malicious message and a legitimate internal email may be a single, seemingly harmless sentence.

The use of generative AI has made BEC attacks even more dangerous. In the past, knowledgeable employees could often spot BEC emails based on poor grammar, awkward phrasing, or obvious spoofing. Today, AI tools enable threat actors to produce high-quality emails that are grammatically correct, convincingly written, and accurately tailored to individual users and organizations. These messages often spoof legitimate senders and align closely with normal business communication, making them difficult for both security systems and employees to distinguish from real emails. As a result, AI-driven BEC campaigns are becoming more precise, more credible, and far more effective at bypassing defenses and exploiting trust.

Trend #4: Legitimate Files and Remote Access Tools Become Primary Weapons

Over the past year, Cofense analysts observed a marked increase in threat actors using legitimate content to deliver malware. One of the most concerning developments is the growing abuse of legitimate RATs. Software such as ConnectWise ScreenConnect and LogMeIn's GoTo Remote Desktop, which are commonly used by IT teams for support and troubleshooting, are being repurposed by attackers to function as remote access trojans and provide full control over infected devices.

In 2025, the use of legitimate RATs by threat actors increased by 57 percent. These attacks are especially dangerous because nearly every step of the infection chain appears legitimate to most endpoint detection and response solutions. Files are often hosted on trusted cloud platforms such as Dropbox, Amazon S3, and OneDrive, making downloads difficult to flag. The binaries are typically signed with valid digital certificates, rendering many traditional detection techniques ineffective. Even when these tools are identified as potentially unwanted programs, alerts may be ignored if security teams are not closely aligned with IT, since the same software is frequently used for legitimate administrative purposes.

As the scale of the attacks using legitimate RATs increases, so does the burden on threat actors, as each infected system must be manually monitored and managed. Some of the RATs involved use AI to allow large-scale management by default, while others, like ConnectWise RAT, expose interfaces that can be controlled by AI-based tools. To continue to execute campaigns involving a large number of systems infected by legitimate RATs, threat actors increasingly rely on automation and AI in their workflows.

Network activity further complicates detection. Communications generally occur with well-established, trusted domains, and the only indication of attacker control may be a key passed through otherwise standard and legitimate network traffic. This allows threat actors to abuse free trials and gain persistent access to entire networks without leaving obvious indicators behind.

At the same time, attackers using traditional malware have also adapted their tactics. Cofense observed a sixfold increase in the use of legitimate files compared to 2024. This includes techniques such as side loading malicious content into legitimate processes, as well as delivering scripting packages like Python and AutoIT. As the use of legitimate tools and content continues to grow across all stages of attack campaigns, understanding the broader context of activity will become critical. In many cases, context may be the only reliable way to determine whether a legitimate file or tool truly belongs in an environment. Teams should use AI-driven detection with human supervision to gather and interpret contextual signals needed to determine whether a technically legitimate RAT is being abused or is in fact an authorized component of an organization's infrastructure.

The Cofense Phishing Defense Center saw a

106% increase

in the total number of remote access trojans reported (legitimate and non-legitimate) in 2025 compared to 2024.

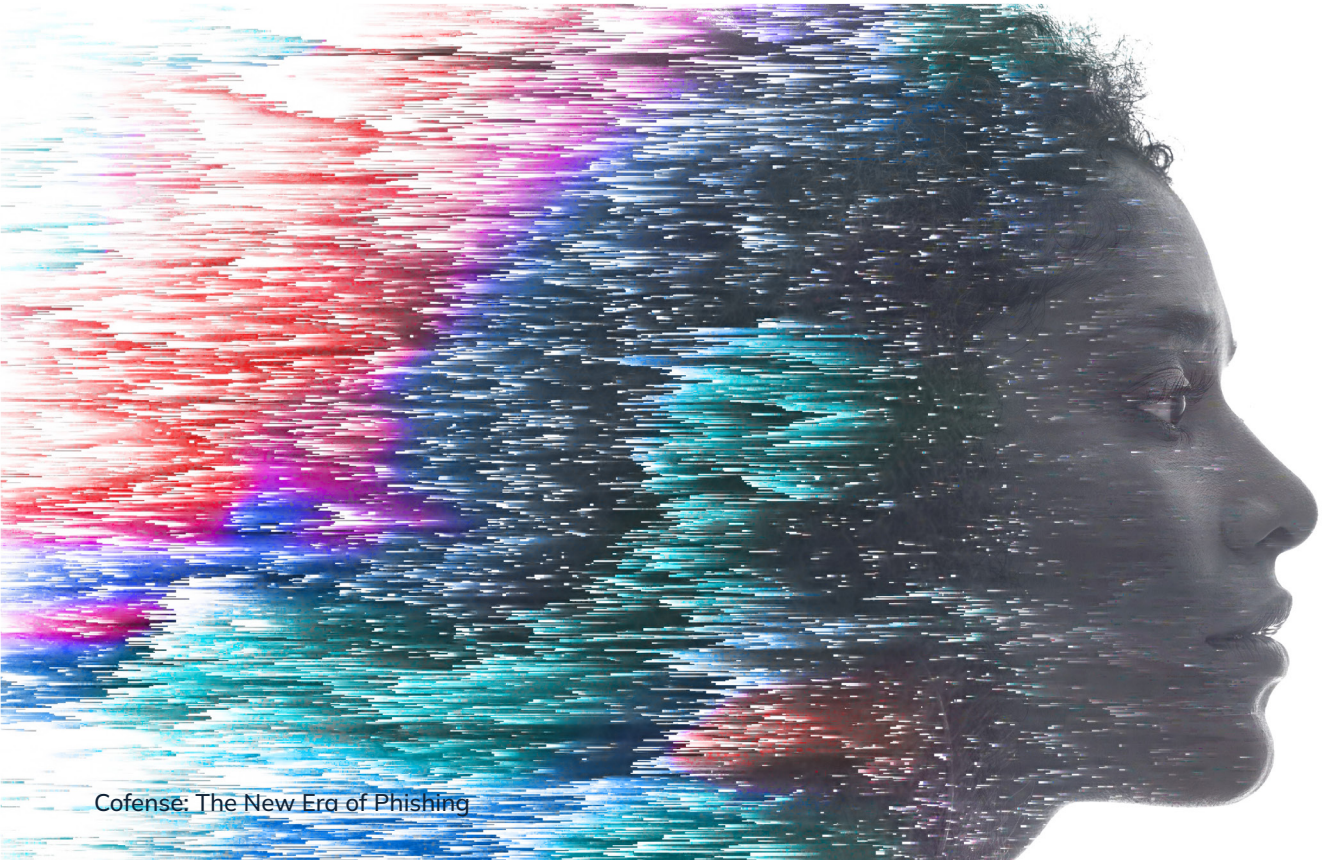
Trend #5: Expansion into Previously Underused Top-Level Domains

In 2025, Cofense Intelligence observed a dramatic shift in the top-level domains (TLDs) used in credential phishing campaigns, with the .es TLD emerging as one of the most heavily abused. Usage of .es domains in credential phishing increased 19 times from the fourth quarter of 2024 to the first quarter of 2025. When comparing all of 2024 to all of 2025, that increase rose to 51 times. As a result, the .es TLD moved from roughly 56th place to 3rd among all abused domains, representing the most significant change in TLD usage seen in credential phishing during this period.

There are several possible explanations for the sudden prioritization of the .es TLD. Threat actors may be taking advantage of easier deployment methods for credential phishing pages, or they may be using new phishing kits that automatically generate .es domains with subdomains and deploy phishing content at scale. If driven by tooling, this trend likely reflects the continued rise of AI-enabled kits that allow less experienced attackers to launch advanced credential phishing campaigns with minimal cost and effort.

Threat actors abusing the .es TLD most commonly relied on compromised or attacker-controlled subdomains hosted on seemingly random domains, such as post[.]bphsrpzyqg[.]es, challenge[.]rocarlo[.]es, and switch[.]bphsrpzyqg[.]es. This approach is notable because attackers often customize subdomains to closely match the email lure, such as references to voicemail or document notifications. In these campaigns, the subdomains appeared largely generic, suggesting automation rather than manual customization.

While the increase in .es TLD usage was observed to a lesser extent in links embedded directly within phishing emails, it was most prominent in second-stage domains. These domains were typically used for redirects or to host the final credential phishing pages. In most cases, the .es domains hosted fully developed and convincing phishing sites, supported by well-crafted emails that included branded layouts and logos and closely resembled legitimate communications rather than simple text-based messages.



Conclusion

The trends outlined in this report reveal that phishing has entered a new era. AI has fundamentally changed how attacks are created, scaled, and delivered, enabling threat actors to produce campaigns that are faster, more adaptive, and increasingly indistinguishable from legitimate business communications. The result is a threat landscape defined by polymorphism, personalization, and constant variation, where traditional perimeter-focused defenses alone are no longer sufficient.

Across each trend examined, from the abuse of legitimate RATs to the rise of conversational business email compromise and highly adaptive phishing pages, attackers are deliberately designing campaigns to bypass static controls and exploit trust after messages reach the inbox. The scale of uniqueness observed in indicators of compromise, files, and URLs further reinforces this shift. When nearly every attack appears different on the surface, exact-match detection and reliance on previously seen indicators cannot keep pace.

Cofense is purpose-built to address this reality. By combining post-delivery visibility, human-reported intelligence, and AI-driven analysis, Cofense delivers the context required to detect and stop threats that evade automated controls. Millions of trained users serve as distributed sensors, surfacing real attacks in real time. Our expert analysts validate, investigate, and enrich those reports, transforming individual phishing emails into actionable intelligence that can be used to identify campaigns, uncover emerging tactics, and protect organizations at scale.

This human-powered intelligence is amplified through the Cofense platform approach. Signals generated from one customer's reported threat can help protect others, creating a powerful network effect that extends beyond traditional indicator sharing. Rather than relying solely on static indicators or opaque models, Cofense provides explainable and actionable insight that security teams can trust and act on quickly.

As AI continues to lower the barrier to entry for attackers and accelerate the pace of phishing innovation, organizations must adapt their defenses accordingly. Success in this environment depends on seeing what gets through, understanding why it worked, and responding with speed and confidence. Cofense enables this shift by closing the gap between detection and response and by ensuring that human insight remains central to phishing defense.

In an era where phishing is built with AI, effective defense requires more than automation alone. It requires visibility beyond the perimeter, intelligence grounded in real-world attacks, and the ability to act decisively when trust is exploited. This is the value Cofense delivers and why it remains essential for organizations preparing for the next generation of phishing threats.

Contributors:

Josh Bartolomie
Chief Security Officer;
VP, Global Threat Services

Chance Caldwell
Sr. Director,
Phishing Defense Center

Max Gannon
Mgr, Intel Analysis

COFENSE

Cofense is the leader in post-perimeter phishing defense, built for the reality that phishing gets through. Cofense helps enterprises identify threats already in employee inboxes, remediate active attacks fast, and reduce future risk.

The Cofense platform combines phishing-specific AI with expert validation, automation, and collective insight drawn from millions of real-world user interactions. Human-supervised intelligence improves accuracy, eliminates guesswork, and accelerates response so security teams can act with confidence.

Global enterprises including Mastercard, Accenture, and Toyota trust Cofense to strengthen organizational resilience against the most persistent cyber threat: phishing.