magix

Phishing Evolution in the AI Era

The Magix R&D Lab

Primary Author: Tim Butler

Co-Author: Floyd Tshoma

Co-Author: Hlayisani Shondlani

Phishing Evolution in the AI Era

The Magix R&D Lab

Authors

T. Butler (Primary)

H. Shondlani (Co-author)

F. Tshoma (Co-author)

Abstract

As generative AI reshapes the digital landscape, phishing, the world's most persistent cyber threat, has entered a new era of speed, scale, and sophistication. Phishing remains a primary source of initial access to business networks, often serving as the launchpad for broader cyber intrusions. With the average phishing-related data breach costing businesses USD 4.88 million, underscoring this attack vector's dominance.

Threat actors are weaponizing AI tools, such as SpamGPT, WormGPT and FraudGPT, leading to a surge in phishing sophistication and volume. Some metrics reveal an increase of 1,265% since the launch of generative AI tools, and defences are struggling to keep pace, making adaptive human-AI defence collaboration essential. The next generation of defence depends not just on stronger filters, but on integrating human intuition with AI precision. Human-AI cooperation will be crucial in anticipating, detecting and mitigating this evolving threat landscape.

Introduction

What is Phishing

Phishing comes in many forms, but the most common is one we are all familiar with: the email. This is traditionally a deceptive message appearing to come from a trusted source, designed to create urgency that pressures the receiver to act quickly without verifying the sender's identity. Sometimes a malicious link or attachment is added to harvest sensitive data such as financial information or credentials. Once obtained, this information can be used to conduct financial fraud, identity theft or data breaches.

Everyone can recall those poorly worded emails from a supposed prince in a far-off land, promising riches. Yet, as AI tools become weaponized, that era is ending. Generative AI now enables highly convincing and tailored phishing messages, crafted with flawless grammar and cultural nuance, leading to dramatically higher success rates. Extending beyond email into other vectors such as vishing (voice phishing), smishing (SMS phishing) and social-media based scams, broadens the overall attack surface.

76% Of businesses email compromise emails in Q2 were AI-generated, as confirmed by multiple AI text detection tools

86% Of organizations reported at least one AI-related incident that included AI-powered phishing or social engineering

42% AI-aided social engineering incidents impacted of organizations

48% of employees understand how threat actors use AI for phishing

History of Phishing

The origins of phishing are difficult to pinpoint, as precursors existed long before it became notoriously popular. Notably in 2000, the infamous ILOVEYOU work spread malware through email with a malicious attachment disguised as a text file, overwriting image files and replicating itself across the victims address book.

A brief timeline:

- **2000s:** The rise of mass email scams exploiting eCommerce and online payment systems, such as PayPal.
- **2010s: I**ntroduction of targeted attacks, including spear-phishing and whaling, with the emergence of ransomware.
- **2020s:** Integration of AI and cryptocurrency enabling professionalised cybercrime and anonymous financial transactions.

The rise of social media and AI weaponization has led to exponential growth in both the quantity and quality of phishing campaigns.

Traditional Tools

Before AI, phishing relied on manual construction or legitimate security tools repurposed by threat actors. One such legitimate tool is SET (Social Engineering Toolkit).

SET is an open-source Python-driven penetration testing suite for social engineering. Popular amongst many cybersecurity professionals, it includes modules for spear-phishing, website cloning, SMS spoofing and more.

Using SET, the process of creating, implementing and monitoring the unfolding campaigns is a straightforward matter. This allows professionals to gather valuable intelligence and proceed with further penetration testing on a client’s environment.

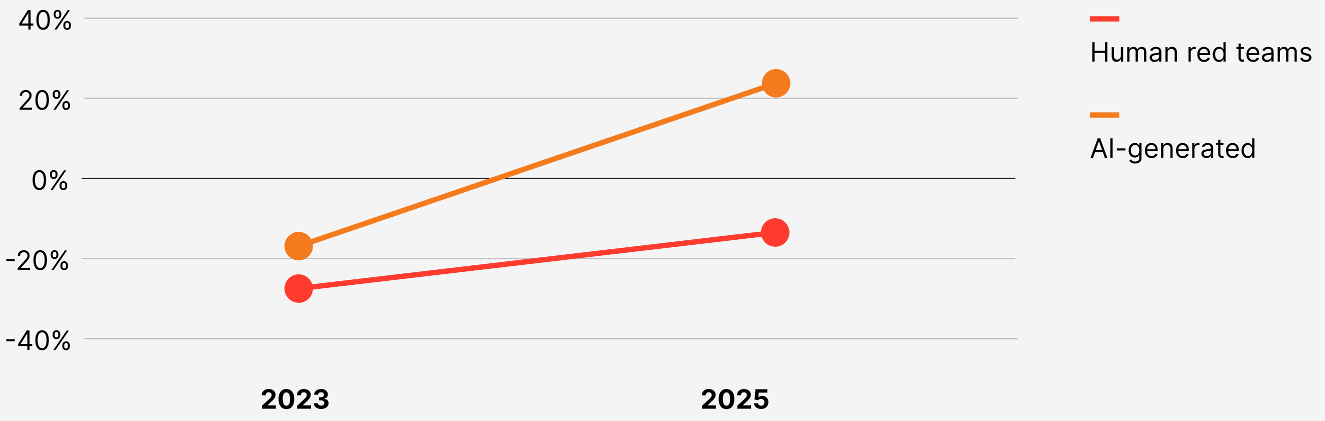
However, when repurposed by bad actors, SET is a powerful tool for running malicious phishing campaigns.

Commercial tools such as Phishing Box and uSecure have also emerged, helping organizations simulate phishing and reinforce staff awareness through gamified learning. However, the same innovation cycle benefits attackers, who now employ automation and AI to increase precision and scale.

How has phishing evolved

AI vs Human Phishing Efficiency

Hoxhunt’s experiment results



AI-generated phishing has surpassed human-crafted attacks on tricking users

With generative AI, phishing has entered a new sphere of influence defined by speed, automation and near-flawless deception.

AI has transformed phishing in five key ways:

- Data-Driven Targeting: AI scrapes social media, professional profiles and company websites and other public data to gather detailed context on a target’s role, interests and behaviours. Using this, it can understand a person’s role in business, current projects, events, contacts, interests and, in many cases, even their writing style.
- Language Perfection: AI removes telltale grammatical errors that used to signal danger, and can be adapted to write in any language.
- Personalisation at Scale: Messages mimic the tone, writing style, and even the internal slang (such as short-names and nicknames of coworkers) to sound authentic and make the message appear uniquely relevant.
- Automation and Speed: Thousands of unique phishing variants can now be produced in minutes, instead of spending hours carefully crafting each one.
- Beyond Email: Voice cloning and deepfake technology extend phishing into calls, voice messages and even video conferencing. (Further details regarding this can be found in a previous white paper by the Magix R&D Lab: The Surge of Deepfake Cyber Threats.)

In 2024, IBM performed security research, where a team of experts pitted against AI to create a phishing campaign. The AI only needed 5 prompts and 5 minutes to build an email nearly as effective as the one that took the experts’ 16 hours to craft. This is an enormous time disparity and can lead to massive scale of phishing campaigns.

The creation and sending out of phishing email is not the only scope of the AI-PhaaS (AI-powered Phishing-as-a-Service) era, but more sophisticated models are being implemented, allowing attackers to do far more. Tycoon 2FA, a phishing kit as emerged in August of 2023, was specifically designed circumvent two-factor and multi-factor authentication measures on Microsoft and Gmail accounts. This demonstrates the escalation in threat sophistication from Threat Actors, and the cybersecurity landscape is proving to be more challenging than ever before.

Lowering the Barrier to Entry

PhaaS (Phishing-as-a-Service) has commoditized cybercrime. Instead of needing advanced technical skills and extended periods of time, even inexperienced actors can now rent fully functional phishing infrastructure. Hosted on the dark web and illicit forums, these services offer subscription-based access to templates, spoofed domains, AI-driven customization and analytics.

Platforms such as Caffeine, Greatness, and EvilProxy demonstrate how professionalized this industry has become. PhaaS mirrors legitimate SaaS models, with tiered pricing, dashboards tracking campaign success rates, and even customer support channels. Combined with SpamGPT, WormGPT or FraudGPT, phishing has evolved into a scalable fraud ecosystem where deception is an automated, multilingual criminal enterprise.

Case Examples

April 2025

Illinois Department of Human Services (IDHS)

In April 2025, the Illinois Department of Human Services (IDHS) reported a breach that resulted from a phishing campaign. The campaign targeted IDHS employee accounts and led to unauthorized access to personal information on their database. This affected roughly one-million individuals, where approximately 4,700 people had their Social Security numbers exposed.

June 2025

HMRC

In June 2025, a large-scale phishing attack allowed scammers to access roughly 100,000 users accounts. After gaining their personal details, the scammers then created fraudulent accounts, posed as taxpayers and managed to seal £47 million from HMRC through claiming rebates. The HMRC managed to identify and lock down the affected accounts, followed by notifying the compromised users, many who did not have an online tax account.

Some attacks do not rely on phishing alone. In July 2025, researchers exposed phishing campaigns that impersonate popular brands. These campaigns tricked their targets into calling phone numbers that are operated by these threat actors. This popular social engineering is known as a Telephone-Orientated Attack Delivery (TOAD) or callback phishing. By coaxing the victims into calling the number in the enticing phishing email, the attacker then poses as a legitimate customer representative and tricks the victim into either disclosing sensitive information or installing malware on their devices.

Phishing in 2025: Key Threat Statistics

36%
of data breaches start with phishing

$4.88m
average breach cost

1,265%
surge in phishing emails since generative AI

What can we do to prevent it

Remaining aware, vigilant and suspicious is essential. Although this seems like a simple answer, it can be incredibly difficult to implement in busy work environments. Employees are often dealing with multiple tasks and projects, with multitudes of emails that tie to those tasks and others that do not. Coupled with the phishing emails premise of urgency, this often leads to a momentary slip in this vigilance.

Awareness training is the most effective form of keeping employees vigilant but, as mentioned, workloads can prevent them from giving the training the proper attention it deserves. Simple and consistent reminders and practices can help alleviate this and grow a culture of awareness. These practices can include:

Checking URL Links

The links and URLs in phishing campaigns are often long and confusing, in order to disguise the true destination and mislead users.

As a base point, always ensure any link contains HTTPS for secure connections. If in doubt, using a url safety checker, such as ESET Link Checker and Trend Micro Safety Center, allows a user to check suspicious links.

Check Redirect Chains

Some phishing emails have multiple redirects, stemming from legitimate sources that eventually lead to a malicious site. Sometimes this might be to download a malicious file through these links.

Copying the link into a document (such as Microsoft Word) allows you to view it in its entirety, pinpointing where it will ultimately lead you. Alternatively using url safety checkers allows you to confirm and view the data before clicking on it. When in doubt, do not click on the link.

Inspect the Site

Most malicious URLs have a sign-in or login page, where they attempt to capture your login details. Clues to a malicious site can be found in certain areas, such as a missing Favicon or a mismatched page title.

Verifying the domain is also recommended to ensure you're on a legitimate site. Sometimes this will be glaringly obvious and other times more subtle. An example is login.microsoftonline.com as a legitimate site, where a malicious site might be login.miscrosoft-online.com or login.microsoftinline.com. These subtle differences should post some red flags in a user's mind.

Report to your IT Team

If in any doubt, always turn to your IT Team. Forward the suspicious mail to them, where a trained eye can look through and verify whether it is legitimate or not. This not only provides reassurance to the user but allows the IT Team to get ahead of the curve and implement blocks and other security measures to prevent such an attack from rapidly expanding through the rest of the business.

AI-Enhanced Defences

Using machine learning tools can assist in identifying linguistic or behavioural patterns of AI-written text. The behavioural analytics, especially, can assist in detecting anomalies in message timing, tone or sender-recipient relationships. AI can also begin implementing adaptive awareness, where it uses microlearning training triggered by real user behaviour. Finally, implementing a zero-trust verification process, where sensitive transactions require a secondary confirmation, can apply an additional layer of defence against sophisticated phishing exploits.

Conclusion

Phishing has evolved from a numbers game into a precision strike. AI has made social engineering faster, smarter, and more scalable. Its integration into cybercrime ecosystems ensures continued evolution. Businesses recognize that no single filter or training program is sufficient. True resilience demands collaboration between technology, process and people, combining AI's detection speed with human judgement.

The next generation of cybersecurity resilience will depend not just on blocking phishing, but on anticipating it.

magix

This white paper was compiled by the Magix Lab team

Sources

https://deepstrike.io/blog/Phishing-Statistics-2025

https://www.strongestlayer.com/blog/ai-generated-phishing-enterprise-threat-2025

https://thehackernews.com/2024/09/expert-tips-on-how-to-spot-phishing-link.html

https://thehackernews.com/2024/08/how-phishing-attacks-adapt-quickly-to.html

https://thehackernews.com/2025/03/new-morphing-meerkat-phishing-kit.html

https://urlsafetychecker.com/

https://pages.egress.com/whitepaper-email-risk-report-01-24.html

https://www.checkpoint.com/cyber-hub/threat-prevention/what-is-phishing/ai-phishing-attacks/

https://ccgrouppr.com/blog/wormgpt-fraudgpt-the-dark-side-of-ai/

https://www.ibm.com/think/x-force/ai-vs-human-deceit-unravelling-new-age-phishing-tactics

https://zerothreat.ai/blog/deepfake-and-ai-phishing-statistics

https://www.rapid7.com/blog/post/ai-goes-on-offense-how-llms-are-redefining-the-cybercrime-landscape/

https://www.varonis.com/blog/spamgpt

https://www.kaseya.com/blog/history-of-phishing/

https://hoxhunt.com/guide/phishing-trends-report

https://nordlayer.com/blog/data-breaches-in-2025/

https://www.brightdefense.com/resources/recent-data-breaches/

https://thehackernews.com/2025/07/hackers-using-pdfs-to-impersonate.html

https://cybersecuritynews.com/attack-techniques-of-tycoon-2fa-phishing-kit/

https://cybersecuritynews.com/ai-tools-promoted-by-threat-actors/

https://www.airosoftware.com/phishing-attacks-all-you-need-to-know/