Generative AI: Impact on Email Cyber-Attacks

DARKTRACE

↑ 135%
increase in novel social engineering attacks from January to February 2023

! 82%
of global employees concerned about attackers using Generative AI to create scam emails

↑ 70%
of global employees notice increase in scam emails and texts in the last 6 months

Monday, 3rd April 2023

Social engineering – specifically malicious cyber campaigns delivered via email – remains the primary source of an organization's vulnerability to attack. Popularized in the 1990s, email security has challenged cyber defenders for almost three decades. The aim is to lure victims into divulging confidential information through communication that exploits trust, blackmails or promises reward so that threat actors can get to the heart of critical systems.

Social engineering is a profitable business for hackers – according to estimates, around 3.4 billion phishing emails get delivered every day.[1]

As organizations continue to rely on email as their primary collaboration and communication tool, email security tools that that rely on knowledge of past threats are failing to future-proof organizations and their people against evolving email threats.

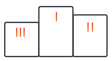
Widespread accessibility to Generative AI tools, like ChatGPT, as well as the increasing sophistication of nation-state actors means that email scams today are more convincing than ever.

Humans can no longer rely on their intuition to stop hackers in their tracks; it's time to arm organizations with AI that knows them better than attackers do. In new data published today, Darktrace reveals that email security solutions including native, cloud and 'static AI' tools, **take an average of thirteen days from an attack being launched on a victim to that attack being detected**, leaving defenders vulnerable for almost two weeks if they rely solely on these tools.

In March 2023, Darktrace commissioned a global survey with Censuswide, to 6,711 employees across the UK, US, France, Germany, Australia, and the Netherlands to gather third-party insights into human behavior around email, to better understand how employees globally react to potential security threats, their understanding of email security and the modern technologies that are being used to transform the threats against them.

Key findings include:

82% of global employees are concerned that hackers can use Generative AI to create scam emails that are **indistinguishable from genuine communication**



The top three characteristics of **communication that make employees think an email is a phishing attack** are: being invited to click a link or open an attachment (68%), unknown sender or unexpected content (61%) and poor use of spelling and grammar (61%)

Nearly 1 in 3 (30%) global employees have **fallen for a fraudulent email or text** in the past

70% of global employees have noticed an **increase in the frequency of scam emails and texts** in the last 6 months

87% of global employees are **concerned about the amount of personal information available** about them online that could be used in phishing and other email scams

Almost 4 in 5 (79%) company spam filters **incorrectly stop important legitimate emails** getting to their inbox

Over a third of people (35%) have **tried ChatGPT** or other Gen AI chatbots

1. AAG, 'The Latest 2023 Phishing Statistics' March 2023.

The email threat landscape today

Darktrace researchers observed a **135% increase in 'novel social engineering attacks' across thousands of active Darktrace/Email customers from January to February 2023**, corresponding with the widespread adoption of ChatGPT.[2] A novel social engineering phishing email is an email attack that shows a strong linguistic deviation - semantically and syntactically - compared to other phishing emails. The trend suggests that Generative AI, such as ChatGPT, is providing an avenue for threat actors to craft sophisticated and targeted attacks at speed and scale.

In addition, threat actors are rapidly exploiting the news cycle to profit from employee fear and uncertainty. The latest iteration of this is the collapse of Silicon Valley Bank (SVB) and the resulting banking crisis, which has presented an opportunity for attackers to spoof highly sensitive communication, for example seeking to intercept legitimate communication instructing recipients to update bank details for payroll. **73% of employees working in financial services organizations have noticed an increase in the frequency of scam emails and texts in the last 6 months.**

Innocent human error and insider threats remain an issue. **Many of us (nearly 2 in 5) have sent an important email to the wrong recipient** with a similar looking alias by mistake or due to autocomplete. **This rises to over half (51%) in the financial services industry and 41% in the legal industry,** adding another layer of security risk that isn't malicious. A self-learning system can spot this error before the sensitive information is incorrectly shared.

2. Based on the average change in email attacks between January and February 2023 detected across Darktrace's email deployments with control of outliers.

What does the arms race for Generative AI mean for email security?

Picture this: Your CEO emails you to ask for information. It's written in the exact language and tone of voice that they typically use. They even reference a personal anecdote or joke. Darktrace's research shows that **61% of people look out for poor use of spelling and/or grammar as a sign that an email is fraudulent**, but this email contains no mistakes. The spelling and grammar are perfect, it has personal information and it's utterly convincing. But your CEO didn't write it. It was crafted by Generative AI, using basic information that a cyber-criminal pulled from social media profiles.

The emergence of ChatGPT has catapulted AI into the mainstream consciousness – **35% of people have already tried ChatGPT or other Gen AI chatbots for themselves** – and with it, real concerns have emerged about its implications for cyber defense. **82% of global employees are concerned that hackers can use Generative AI to create scam emails indistinguishable from genuine communications.**

Emails from CEOs or other senior business leaders are the third highest type of email that employees are most likely to engage with, with over a quarter of respondents (26%) agreeing. Defenders are up against Generative AI attacks that are linguistically complex and entirely novel scams that use techniques and reference topics that we have never seen before. In a world of increasing AI-powered attacks, we can no longer put the onus on humans to determine the veracity of communications. This is now a job for artificial intelligence.

Self-Learning AI in email, unlike all other email security tools, is not trained on what 'bad' looks like but instead learns *you* and the normal patterns of life for each unique organization.

By understanding what's normal, it can determine what doesn't belong in a particular individual's inbox. Email security systems get this wrong too often, with **79% of respondents saying that their company's spam/security filters incorrectly stop important legitimate emails from getting to their inbox.**

With a deep understanding of the organization, and how the individuals within it interact with their inbox, the AI can determine for every email whether it's suspicious and should be actioned or if it's legitimate and should remain untouched.

This approach can stop threats like:

- CEO fraud
- Business email compromise (BEC)
- Invoice fraud
- Phishing scams
- Data theft
- Social engineering
- Ransomware & malware
- Supply chain attacks
- URL-based spear-phishing
- Account takeover
- Human error
- Insider threat