DARKTRACE

How AI is Changing the Phishing Landscape

Everything you need to know to stay ahead of emerging email threats

Content

AI-powered threats are already the norm

Today, phishing has moved far beyond opportunistic email scams with suspicious payloads. Threat actors are increasingly leveraging AI to generate sophisticated social engineering campaigns across collaboration tools, social media, and mobile platforms.

While the basic mechanism of a phishing attack remains the same – exploiting human trust – AI makes it faster and easier for threat actors to launch complex campaigns that evade detection across multiple channels.

Security teams are encountering these AI-driven threats at every stage of the attack lifecycle, placing a growing strain on resources and resilience. In a recent Darktrace survey, 73% of organizations said AI-powered threats are already having a significant impact, while 92% reported being forced to upgrade their defenses to cope.

Against this backdrop, this guide explores how modern threat actors are using AI to turbo-charge their phishing engines, why traditional enterprise email security cannot keep up, and how a multi-layered, organization-specific AI defense can prevent successful phishing campaigns from causing bigger problems like:

- Account Takeover (ATO)
- CEO fraud
- Business Email Compromise (BEC) or Vendor compromise (VC)
- Ransomware and malware
- Supply chain attacks

393B

- **Emails get sent** every day[1]

3.4B

- **Phishing emails** get sent every day[2]

x2

- **AI assisted phishing text** doubled year-over-year[3]

1. Statista, 2023, "Number of sent and received e-mails per day worldwide," https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/
2. AAG, 2024, "The Latest 2024 Phishing Statistics," https://aag-it.com/the-latest-phishing-statistics/
3. Verizon DBIR 2026

AI gives attackers a potent force-multiplier

50%

- **of cyber professionals**
  cite hyper-personalized phishing using AI as a top concern[4]

Phishing is, at heart, a numbers game — and AI can profoundly change the math. Analysis of the 32 million phishing emails detected across Darktrace's global fleet shows a clear trend: email attacks grew significantly more sophisticated in 2025, with signs of AI assisted phishing still accelerating.

Leveraging AI makes it easier for threat actors to:

- Launch more convincing campaigns to a broader audience
- Harvest more data on targets, especially VIPs
- Increase the personalization of attacks
- Creating convincing multi-stage attacks without code

LLMs raise the bar

The Large Language Models (LLMs) found in CoPilot, Claude, or ChatGPT enable sophisticated, targeted phishing attacks at scale. LLMs can do the legwork of gathering data about intended targets, so they can craft more personalized and convincing emails that mimic the style of trusted senders and domains – at scale. Consequently, spear phishing has become the norm, particularly for VIPs with high levels of access within systems.

Within Darktrace's customer base, signs of AI usage increased year-over-year, with novel social engineering techniques rising from 32% to 38% and large-text, long-form messages increasing from 27% to 33% – reflecting a shift toward payload-less lures designed to evade traditional filters.

"Threat actors use generative AI to draft phishing lures, translate content, summarize stolen data, generate or debug malware, and scaffold scripts or infrastructure. For these uses, AI functions as a force multiplier that reduces technical friction and accelerates execution, while human operators retain control over objectives, targeting, and deployment decisions."

- **Microsoft - AI as Tradecraft**[6]

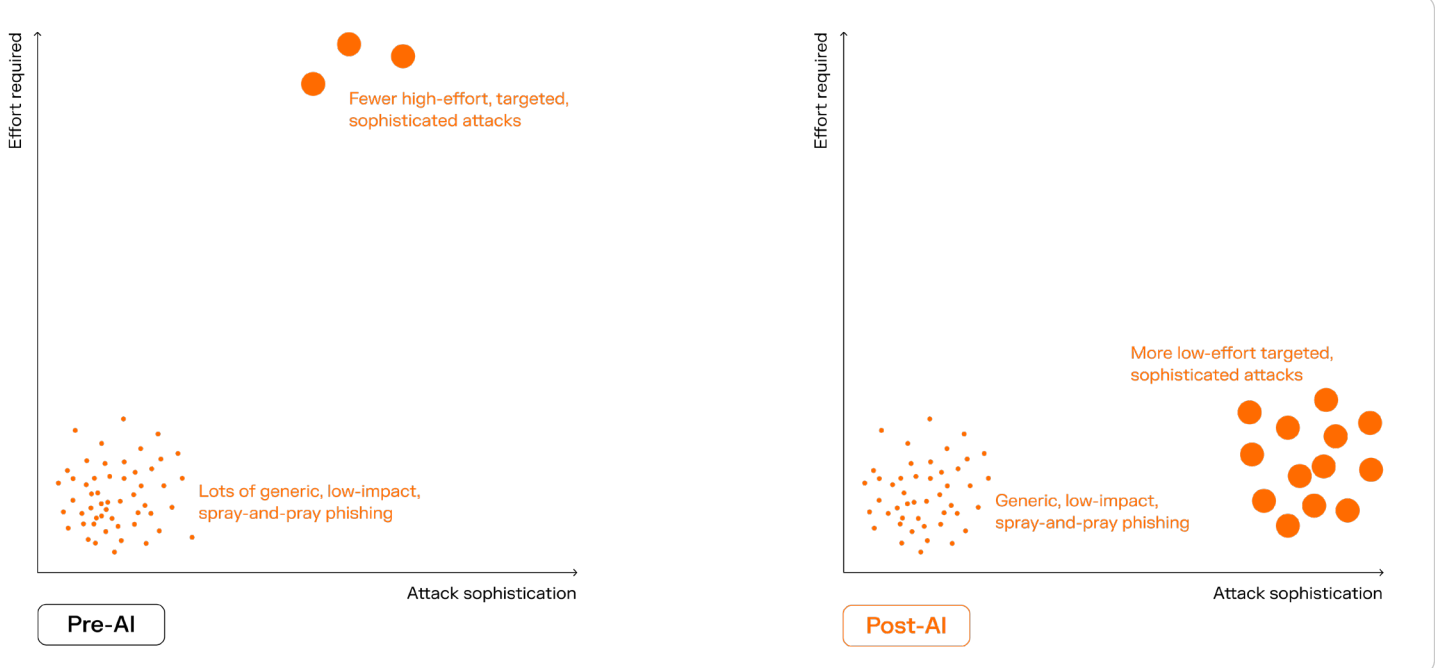


Figure 01: As opposed to traditional "spray and pray" phishing tactics, AI makes it easier for attackers to launch sophisticated and personalized attacks at a greater scale.

4. https://www.darktrace.com/resource/the-state-of-ai-cybersecurity-2026-threat-landscape
5. https://www.darktrace.com/resources/annual-threat-report-2026
6. https://www.microsoft.com/en-us/security/blog/2026/03/06/ai-as-tradecraft-how-threat-actors-operationalize-ai/

How to catch an AI-generated phish

AI has lowered the barrier to entry for successful phishing. Tasks that once required experienced developers – such as building convincing phishing websites that closely mimic trusted services like Microsoft 365 – can now be created quickly with the assistance of generative AI.

Let's take a look at the tone, structure, and intent of a fake phishing email generated by AI.

Immediate Action Required: Account Security Verification

From: IT Support Team <it-support@company.com>

To: Jane Smith <jane.smith@company.com>

Dear Jane Smith,

We hope this message finds you well.

As part of our ongoing commitment to maintaining a secure and resilient digital environment, we are reaching out to kindly request your immediate attention to an important matter. EXCESSIVE POLITENESS

Our systems have detected unusual activity associated with your account following your recent trip to San Diego to collaborate on the Q4 strategic planning initiative and we would be grateful if you could verify your credentials at your earliest convenience to ensure uninterrupted access to services. PERFECTLY TAILORED PERSONALIZATION EXCESSIVE POLITENESS

Please complete the verification process within the next 2 hours to avoid any potential disruption to your workflow and continued access to internal resources. UNEXPECTED URGENCY

We kindly ask that you treat this matter as a priority, as delays may impact your ability to access essential tools, systems, and shared project environments. LLM THREE-ITEM GROUPING

You can securely review your account details via the link provided below, which has been designed to ensure a seamless and secure verification experience. OVER-STRUCTURED REASSURANCE LANGUAGE

Kind regards,

IT Security Operations Team
Global Technology Services Division

It's not just email – it's the whole messaging ecosystem

While phishing attacks have evolved in email, they have also evolved outside of email.

Pre-texting – highly personalized, AI-generated lures that build victims trust – can be adapted and deployed across collaboration platforms such as Microsoft Teams, Slack, and Zoom, creating multi-channel attack chains that target whichever communication channel has the weakest protection.

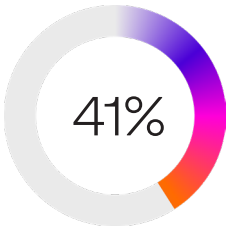
These social engineering threats can emerge and evolve rapidly across platforms undetected, evading rules-based and siloed email and collaboration security approaches.

Where humans fall short, email security needs to fill the gap

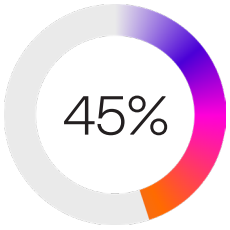
For now, there are still identifiable signals that can help flag AI-generated phishing – but these indicators are not static. As LLMs continue to evolve, the quality of generated content is improving exponentially, likely reducing the reliability of these "tells" over time.

At the same time, today's employees are operating under a huge cognitive load, processing high volumes of emails and notifications at a rapid rate to remain productive. In that environment, even well-trained users will miss things, and it only takes a single click to trigger an identity-based breach. Organizations cannot realistically expect individuals to identify these sophisticated phishing attempts.

Closing that gap requires an email security solution that can detect and block AI phishing attempts before they reach the user, even if they have never been seen before and carry no known threat indicators.



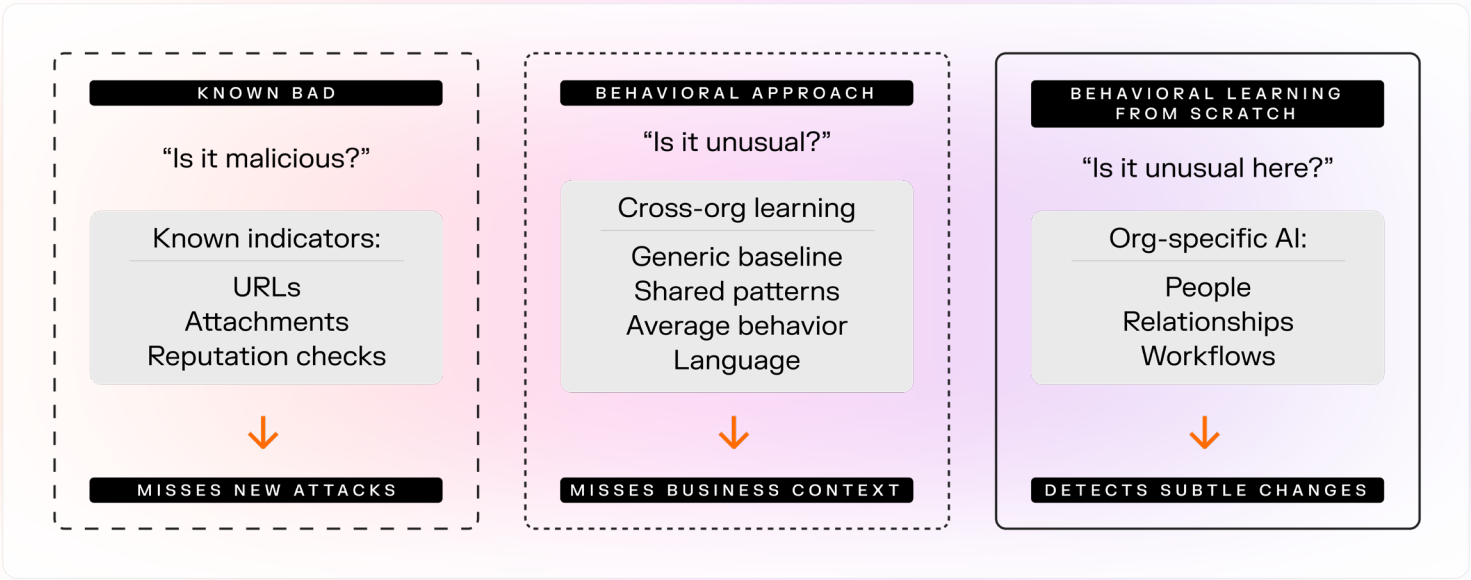
41% of emails observed by Darktrace were personalized spear-phishing attempts targeting individuals.[7]



45% of organizations say they are not prepared for AI-powered cyber threats.[8]

7. https://www.darktrace.com/resources/annual-threat-report-2026
8. https://www.darktrace.com/resource/the-state-of-ai-cybersecurity-2026-threat-landscape

Why legacy approaches aren't designed for AI-generated phishing



Known-bad detection is stuck in the past

Many organizations still deploy a Secure Email Gateway (SEG), assuming it will catch whatever their native email security cannot. However, nearly one in six of the riskiest inbound emails still evade the native + SEG layers on the first pass – 17% is the average SEG miss rate after Microsoft filtering, as calculated by Darktrace. The attacks most missed by SEGs? Solicitation, phishing links, and user impersonation. All threat types that evade static controls and reputation checks via a lack of obvious payload, social engineering, and routine workflow context. In short, the threats that AI is adept at producing.

Behavior-based approaches still look for known indicators

Modern vendors that claim to use a behavioral approach fall into the same trap. They might apply AI learning to look for attacker behaviors across huge datasets of customers, resulting in a global "known good" that then applies to every organization's communication activity. The outcome is a biased understanding of threat indicators which doesn't reflect the unique, day-to-day business context of each specific org.

Building an understanding from scratch with AI

The most effective approach is to build an understanding of each organization's comms individually, rather than applying a model of "normal" learned across thousands of businesses. Every organization has unique communication patterns, relationships, workflows, and language. By learning these from scratch, security can identify subtle deviations that would appear perfectly legitimate in another environment.

This detection method is especially effective against AI-generated threats that lack obvious indicators, because it is designed to detect subtle, sophisticated shifts in behavior for an individual user's inbox or interactions.

How to minimize the risk of AI phishing in your organization

As a security leader, the task can feel overwhelming, AI on every side. But there are practical steps you can take to minimize the susceptibility of your organization and users to targeted, sophisticated phishing attempts.

Questions to ask your email security provider

Ask about your provider's defensive AI.

- Is it trained on patterns aggregated across its customer base, or does it build an understanding of your organization's unique communication patterns from scratch?

AI-generated phishing can be subtle, especially with more employees using it for day-to-day emails.

- Does your product give a nuanced anomaly score for each suspicious email rather than operating block / allow lists?
- Can it take a range of autonomous actions to neutralize emails – including locking links, rewriting URLs, or adding warning tags and banners?
- Do VIPs automatically receive stronger protections, based on the nature and volume of phishing threats they are likely to receive?
- Can you extend your email protection across the entire messaging ecosystem – including Teams, Zoom, and Slack?
- Does your provider offer integrated security awareness training that supports employees in reducing human error and improves email detection?

Next steps

The shift towards AI-driven attacks that are fluent, context-aware and convincingly personalized exposes the limits of attack-centric email security.

The shift towards AI-driven attacks that are fluent, context-aware and convincingly personalized exposes the limits of attack-centric email security.

Protecting your organization now depends on understanding behavior, context, and intent within your own business, not just filtering out malicious messages.

For security leaders, the priority is clear: adopt email and collaboration security that can continuously adapt to novel, AI-generated threats across channels, while supporting users without adding friction to their day-to-day workflows.

Key takeaways for security leaders

- **AI-generated phishing** is reducing the effectiveness of traditional detection signals
- **Sophisticated attacks** increasingly rely on context, personalization, and trust
- **Humans cannot be relied on** to spot convincing AI-generated phishes
- **Threat intelligence and signature-based approaches** cannot keep pace with evolving threats
- **Detection** must focus on behavior and intent, not known attack indicators

Interested in strengthening your defenses against AI-driven phishing? Get a free demo to see how Darktrace can upgrade your detection and response.

About Darktrace / EMAIL

Email and collaboration security built for modern threats

Darktrace / EMAIL is designed to stop the attacks that legacy email security misses, those that abuse trust, mimic normal business behavior, and move across inboxes, identities, and collaboration channels.

Unlike approaches that rely on models built across many organizations, Darktrace's AI learns exclusively from your organization's activity. It builds an understanding of how your people communicate and collaborate, enabling it to detect subtle deviations without relying on rules, signatures, or prior examples.

- **58% of threats detected by Darktrace pass through other email filters**
  Identifies advanced BEC, supplier compromise, and payload-less social engineering that blend into normal business workflows.
- **Blocks novel threats up to 13 days faster**
  AI understands your unique email and messaging patterns to stop attacks before traditional tools can recognize them.
- **Automates response across email, identity, and SaaS**
  Reduces manual triage, alert fatigue, and console switching through coordinated, autonomous action.
- **Protects your brand, messaging, and collaboration tools end to end**
  Provides visibility and protection across users, domains, and collaboration tools (including Zoom, Slack, and Microsoft Teams) not just the inbox.
- **Improves signal quality by 60%**
  Empowers analysts and employees with AI-driven triage and contextual guidance, leading to better end-user reporting and reducing SOC investigations.
- **Unifies email defense and security awareness training**
  Delivers measurable behavior change through adaptive, bi-directional learning for employees.
- **Fastest deployment in the industry**
  Responds to threats up to **30x faster**, delivering value in days, not months.
- **Proven and trusted**
  Recognized as a Leader in the 2025 Gartner Magic Quadrant for Email Security and a Gartner Peer Insights Customers' Choice 2026, with a 4.8★ rating from customers.

Want to learn more?

View the product →

Calculate your potential ROI with Darktrace / EMAIL →

About Darktrace

Darktrace is a global leader in AI cybersecurity that keeps organizations ahead of the changing threat landscape every day. Founded in 2013 in Cambridge, UK, Darktrace provides the essential cybersecurity platform to protect organizations from unknown threats using AI that learns from each business in real-time. Darktrace's platform and services are supported by 2,700+ employees who protect nearly 10,000 customers globally. To learn more, visit www.darktrace.com.

North America: +1 (415) 229 9100 | Europe: +44 (0) 1223 394 100 | Asia-Pacific: +65 6804 5010 | Latin America: +55 11 4949 7696

darktrace.com | info@darktrace.com