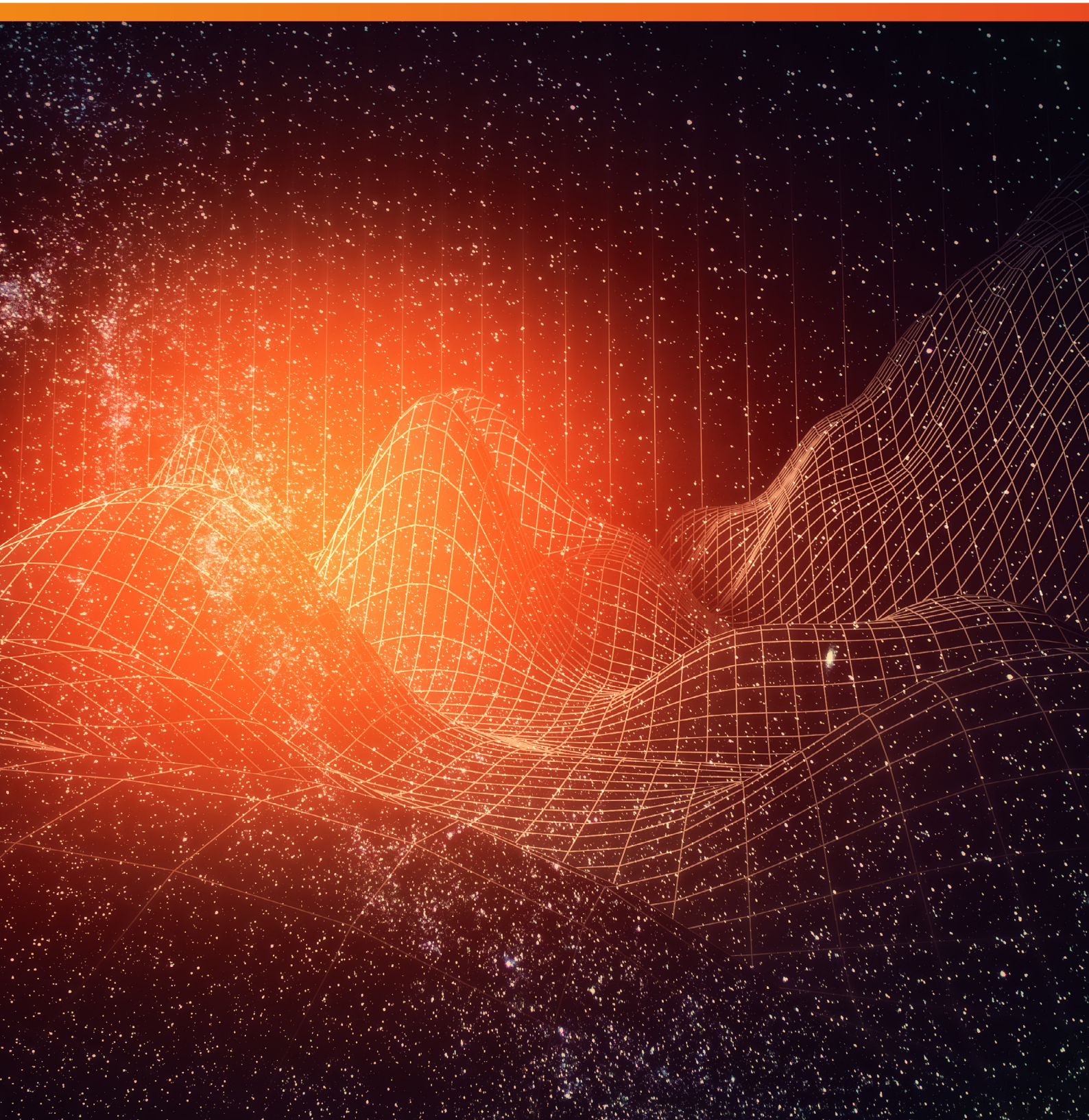
DARKTRACE

How AI is Changing the Phishing Landscape

And How Defenders Can Prepare Themselves



INTRODUCTION

The Race to AI is on

Phishing attacks have evolved from simple email scams to sophisticated campaigns leveraging social media, mobile platforms, and, increasingly, Artificial Intelligence (AI). While the basic mechanism of a phishing attack remains the same, AI makes it faster and easier for threat actors to launch complex campaigns that evade detection.

Identity and Access Management (IAM) leaders understand all too well how dangerous AI and Machine Learning (ML) can be in the wrong hands. What remains to be seen is whether AI can turn the tables on phishing by enabling faster detection, automated response, and a pivot of security postures toward prevention.

This paper provides a mid-2024 update on how modern threat actors use AI to turbo-charge their phishing engines, why traditional enterprise email security cannot keep up, and how defenders can use AI in a multi-layered approach to prevent successful attack campaigns from causing **bigger problems like:**

- Account Takeover (ATO)
- CEO fraud
- Business Email Compromise (BEC)
- Ransomware and malware
- Outbound supply chain attacks

We'll see how the Darktrace ActiveAI Security Platform uses multi-layered AI to understand normal patterns of life and determine which emails belong in each individual user's inbox – something other email security solutions often get wrong (even with the use of AI).

PHISHING THREATS LEAD TO THE MAJORITY OF SECURITY BREACHES

350B

Emails get sent and received every day [1]

3.4B

Spam emails get delivered every day [2]

41%

Security incidents began with phishing as the leading attack vector [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. IBM, 2023, "IBM Security X-Force Threat Intelligence Index 2023," https://www.ibm.com/reports/threat-intelligence

How phishing works . . .

Email inboxes give hackers a place to hide threats in plain sight and trick employees into sharing their passwords, account details, and payment information. Once they obtain valid credentials, threat actors log into systems and instantly gain access to employees' rights, including the ability to change privileges, move sensitive data, and transfer company funds, before anyone notices.

Social engineering attacks may unfold quickly or in stages.

Left unchecked, they can progress to ATO, data being breached or held for ransom, and outbound attacks against your supply chain.

AI gives attackers a potent force-multiplier

Phishing is, at heart, a numbers game — and AI can profoundly change the math. The Anti-Phishing Working Group (APWG) declared 2023 "the worst year for phishing on record" with nearly five billion attacks occurring, but volume isn't the only challenge.[4] The greater potential impact of AI on phishing comes from adding complexity, speed, and scale – making what's worked well for threat actors in the past work even better in the future.

Leveraging AI makes it easier for threat actors to:

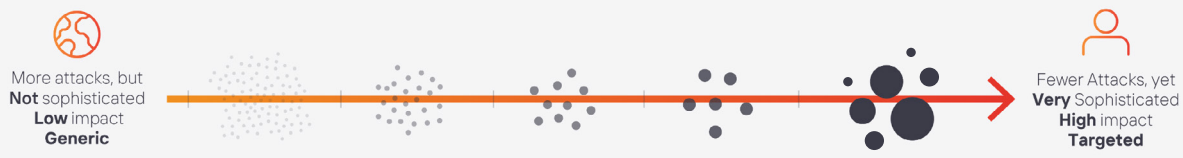
- Launch more convincing campaigns to a broader audience
- Increase the sophistication of attacks
- Repeat attempts with greater frequency, closer to what we've seen in the past with spam overall

> Threat actors are looking at AI, including LLMs, to enhance their productivity and take advantage of accessible platforms that could advance their objectives and attack techniques. Cybercrime groups, nation-state threat actors, and other adversaries are exploring and testing different AI technologies as they emerge in an attempt to understand potential value to their operations and the security controls they may need to circumvent.
>
> Microsoft, 2024

LLMs raise the bar

Phishing leads the charge in the use of offensive AI, as Large Language Models (LLMs) like ChatGPT enable sophisticated, targeted phishing attacks at scale. Massive social engineering attacks do the "legwork" of gathering data about intended targets so LLMs can craft more personalized and convincing emails that mimic the style of trusted senders and domains — and know things only those entities should know. Messages generated by AI tools also contain fewer telltale spelling and grammatical errors that raise a red flag for users.

AI MAKES IT EASY TO LAUNCH HARD ATTACKS FASTER



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

4. Anti-Phishing Working Group, 2023, "Phishing Activity Trends Report: 4th Quarter 2023," https://docs.apwg.org/reports/apwg_trends_report_q4_2023.pdf

Automation also supports mounting hard-to-detect attacks on a pay-as-you-go basis. Phishing-as-a-Service (PhaaS) kits lower barriers to entry by equipping novice attackers with everything they need – ready-to-use emails that impersonate known brands and fake hosted sites used to capture credentials.

70B

email and identity attacks were recorded by Microsoft in 2022 [5]

30%

of employees surveyed had been fooled by fraudulent texts or emails in the past [6]

10.4M

Darktrace detected phishing emails in Q4 2023 [7]

AI carries more phish upstream

Modern campaigns extend beyond the traditional email inbox as cyber-criminals take aim at smartphones, tablets, and Internet of Things (IoT) devices on under-protected home networks. Concurrent with the adoption of LLMs, Darktrace email security experts have seen and predicted notable changes to the phishing threat landscape:

TARGETED ATTACKS GETTING MORE PERSONAL

Advanced social engineering campaigns make it easier to abuse trust and to attack the most desirable targets in spear phishing or "whaling" attacks. Modern BEC attacks impersonate C-Level executives to send emails containing urgent requests to transfer data or company funds. AI enhances these techniques by adding "deepfakes," life-like video, images, and audio that make phishing exploits even more compelling.

THE INEVITABLE RISE OF QUISHING

Darktrace observed a sharp rise in "quishing" attacks that use QR codes for obfuscation.[8] Often associated with marketing promotions and viewing restaurant menus, QR codes don't trigger the same suspicions in users as email attachments and unfamiliar links — which makes them an ideal mechanism for masking phishing exploits. Quishing may fly under the radar of email security as the use of QR codes directs recipients away from their secure inboxes and desktops to mobile phones or other personal devices that lack the same rigorous controls. Most email security solutions only check QR codes as images without evaluating them as potential redirect links. Darktrace takes investigations further to see whether scanning the code takes users to malicious login pages or veiled malware threats.

THREAT ACTORS SEEKING TO 'COLLABORATE' ON TEAMS

Many companies now extend their use of tools like Microsoft Teams outside their organization to collaborate and communicate with customers, partners, vendors, and other third parties. This trend opens another new venue for phishing attacks as threat actors commandeer Teams to impersonate trusted senders. As with QR codes, users are less likely to hesitate before taking the bait and accepting invitations to join Teams calls. Users default to trust since they expect Teams to be for internal use only and most companies do not have dedicated Teams security.

Darktrace addresses this growing threat by applying the full scope of its detection capabilities to spot attempts to gain initial access to Teams and prevent or shut down lateral movement that might follow such a compromise. Darktrace leverages the same behavioral AI techniques for Microsoft customers across Office 365 and Teams to detect threats. Rather than just analyze Teams content, Darktrace/Email-Teams looks at actual user behavior from both a content and context perspective to spot early symptoms of account compromise like early-stage social engineering before a payload is delivered.

PHISHING EXTENDS BEYOND THE INBOX

In **vishing** scams, attackers leave voice messages urging intended victims to call their bank or other trusted institution. Vishers provide a fake phone number and ask users to provide privileged account details when they call.

Smishing leverages the fact that people are 3x more likely to read and respond to a text message than they are an email. Smishing exploits the SMS protocol used in texting to trick users into sharing credentials and PINs that unlock mobile accounts.

Quishing prompts recipients to switch from their desktops or laptops to mobile devices to scan QR codes that take would-be victims to fraudulent sites requesting sensitive information.

5. Microsoft, 2022, "Microsoft Digital Defense Report," https://www.microsoft.com/en-us/security/business/microsoft-digital-defense-report-2022
6. IBM, 2023, "IBM Security X-Force Threat Intelligence Index 2023," https://www.ibm.com/reports/threat-intelligence
7. Darktrace, 2023, "End of Year Threat Report," https://darktrace.com/threat-report-2023
8. Darktrace, "Attacks are Getting Personal, but Cyber Security is Not," https://darktrace.com/blog/attacks-are-getting-personal

> The emergence of Artificial Intelligence (AI) has revolutionized the landscape of cybersecurity, empowering attackers to craft increasingly sophisticated and adaptable polymorphic malware.
>
> Medium: **Ashley Jackson, pentester**

Spear phishing represents

45%

of all phishing attempts

Darktrace End of Year Threat Report 2023 [7]

USERS UNDERSTAND THE RISK

A 2023 survey commissioned by Darktrace showed users around the world understand the phishing risk and need to use modern technology to optimize email security:

61%

of survey respondents say poor spelling and grammar in emails makes them suspect a phishing attack

82%

of employees worldwide are concerned about the use of generative AI to craft convincing phishing emails

79%

of respondents say their company's spam/security filters incorrectly stop important legitimate emails from getting to their inbox [9]

NOVEL THREATS AND OBFUSCATION ON THE RISE

Generative AI lets attackers fashion sophisticated variations on classic phishing campaigns quickly. Darktrace researchers observed a significant increase in novel social engineering attacks across thousands of active Darktrace/Email customers corresponding with the widespread adoption of ChatGPT: [10]

135%

Darktrace detected increase in novel social engineering attacks from January to February 2023 [11]

24.4%

The proportion of phishing emails using obfuscation techniques increased in 2023 [12]

70%

of global employees notice increase in scam emails and texts in the last 6 motnhs [9]

With complex, novel threats making up a higher percentage of phishing attacks, enterprises are effectively less protected for longer periods of time. Darktrace research shows traditional email security solutions — including native, cloud, and 'static AI' tools — take an average of 13 days from an attack being launched to detect and remove it.[13] That's too long. Allowing threats to remain in employees' inboxes for nearly two weeks cannot help but increase the odds of attacks' finding success.

'CLOUD PLATFORM ABUSE' IMPERSONATES GO-TO SERVICES

Darktrace detected a steady rise in the use of legitimate cloud services like MailChimp and SharePoint to perpetrate attacks. Emails sent from these legitimate sites avoid typical domain filtering and may contain malicious payloads that only reveal themselves after users click a link. Everything appears to be legitimate until recipients download a file.

9. Darktrace, "Generative AI: Impact on Email Cyber-Attacks," https://darktrace.com/resources/generative-ai-impact-on-email-cyber-attacks
10. Based on the average change in email attacks between January and February 2023 detected across Darktrace/Email deployments with control of outliers.
11. Internal Darktrace Research (https://darktrace.com/news/darktrace-email-defends-organizations-against-evolving-cyber-threat-landscape)
12. Silicon Angle, 2023, "Report: Over half of phishing emails now use obfuscation tactics to avoid detection" by Maria Deutscher, https://siliconangle.com/2023/10/02/report-half-phishing-emails-now-use-obfuscation-tactics-avoid-detection/
13. Darktrace, "Major Upgrade to Darktrace/Email," https://darktrace.com/news/darktrace-email-defends-organizations-against-evolving-cyber-threat-landscape

Why the security gap keeps widening

Despite investments in Multi-Factor Authentication (MFA), education, and email security, companies keep getting phished. As attacks continue to push traditional, siloed defenses beyond their limits, research shows the percentage of emails eluding Secure Email Gateways (SEGs) increased by 29% from 2022 to 2023.[14]

The gap keeps widening as the traditional approach to email security:

- Relies on rules and deny lists, keeping defenders constantly playing catch-up
- Fails to detect novel attacks or sophisticated threats
- Requires a "first victim" and subsequent time spent analyzing the incident in order to extract the indicators of compromise (IoCs) needed to establish rules and signatures for future threat detection
- Inundates security teams with false positives
- Cannot take appropriate actions to stop threats without disrupting business operations
- Constantly requires maintenance and policy updates
- Requires teams to duplicate workflows from native email provider

Overreliance on historical data

The majority of tools used by organizations today rely on historical attack data to identify and stop known email threats from reentering inboxes. But as noted above, checking emails against lists of known-bad IPs and domains won't catch or prompt systems to hold back new phishing emails. More often than not, solutions based on checking rules and signatures race to keep pace with attacks that have already happened and fall farther and farther behind. These solutions are adept at catching low-impact and generic attacks but can't catch more sophisticated or AI-assisted phishing.

In addition, native solutions like Microsoft or Google have made significant investments in security in recent years, becoming competitive with top SEGs. Consequently, in operating gateways, many security teams are left with duplicate workflows and added expense for similar capabilities. Some newer Integrated Cloud Email Security (ICES) solutions employ AI to improve upon this flawed approach by using data augmentation to watch for similar-looking emails instead of only for direct matches. AI adds valuable horsepower, but the basic approach remains flawed in that it still requires a 'patient zero.'

Both SEG and ICES solutions lack visibility across the digital estate, failing to correlate attacks between email and network, cloud, or endpoint, let alone allowing security teams to get ahead.

DMARC fails to draw the line

Typical enterprise email security relies on the Domain Based Message Authentication, Reporting, and Conformance (DMARC) security protocol to verify the identity of email senders using the Domain Name Server (DNS), Sender Policy Framework (SPF), and DomainKeys Identified Mail (DKIM) protocols – an approach that requires lists of known-bad senders.

Now required by major providers like Yahoo and Google for sending bulk emails, DMARC aims to help email administrators identify and prevent spoofing of the company's own domains, a common phishing technique used to trick employees into revealing sensitive data. DMARC is often positioned as a way for organizations to 'solve' their email security problems, but Darktrace research shows that, without additional intelligence, a significant number of potential threats can manipulate email security and authentication systems to evade detection.

Darktrace/Email-DMARC overcomes weaknesses in conventional approaches to stop spoofing and phishing. This new capability helps upskill businesses with step-by-step guidance and a clear, efficient path to enforcement that makes sure legitimate emails continue to reach users' mailboxes.

Darktrace/Email-DMARC helps to reduce the overall attack surface by providing visibility over shadow IT and third-party vendors' sending on behalf of an organization's brand, while informing recipients when emails from their domains are sent from un-authenticated DMARC sources. The solution integrates with the wider Darktrace product platform, sharing insights to help further secure your business across email attack path and attack surface management.

14. Global Newswire, "New Report Reveals that Nearly Three Quarters (71%) of AI Detectors Can't Tell If a Phishing Email Has Been Written By a Chatbot," https://www.globenewswire.com/news-release/2023/10/02/2752530/0/en/New-report-reveals-that-nearly-three-quarters-71-of-AI-detectors-can-t-tell-if-a-phishing-email-has-been-written-by-a-chatbot.html

DARKTRACE'S AI:

A necessity to strengthen native defenses in the age of AI

Security professionals need to fight AI with AI. Taking a multi-layered AI approach, the Darktrace ActiveAI Security Platform equips security teams to outrun and preempt sophisticated social engineering, phishing, and ransomware attacks led by AI-powered tools.

Darktrace/Email detects subtle symptoms of phishing across companies' entire digital estates to stop threat actors from taking over legitimate user accounts, stealing data, and mounting outbound phishing campaigns against supply chains. A platform-driven approach, personalized anomaly-based detection, and precise autonomous response bring several advantages across the incident lifecycle:

- Faster detection and containment of advanced and novel phishing threats
- Enhanced protection across your entire mailflow and messaging – including Microsoft Teams
- Autonomous response to stop attacks before they escalate
- Complete brand protection while enhancing your existing security investments
- Optimized user experience and security team workflows

58%

of malicious threats stopped by Darktrace/Email pass through other email solutions

Darktrace End of Year Threat Report 2023 [7]

Only Darktrace/Email stops threats

13 days

earlier than traditional solutions [13]

A business- vs. attack-centric approach

The native email security safeguards built into platforms like Microsoft and Google have come a long way and have become foundational component of email security but still take the same backward-facing, attack-centric approach as SEGs. Darktrace takes a business-centric approach that automatically recognizes what doesn't belong in your unique environment and each individual user's inbox. This is what makes it the perfect complement to native email security – covering the full spectrum of generic to advanced attacks.

AI LEARNS YOUR USERS

Instead of studying previous attacks and racing to catch up with threat actors, Darktrace/Email trains itself on your business data. Darktrace AI observes how your employees interact with their inboxes and builds a profile of what constitutes normal for every user – their relationships, tone and sentiment, content, when and how they follow or share links. The platform maintains context and a deep understanding of what normal looks like for your organization and the way individuals work and communicate to recognize suspicious activity that may indicate an attack.

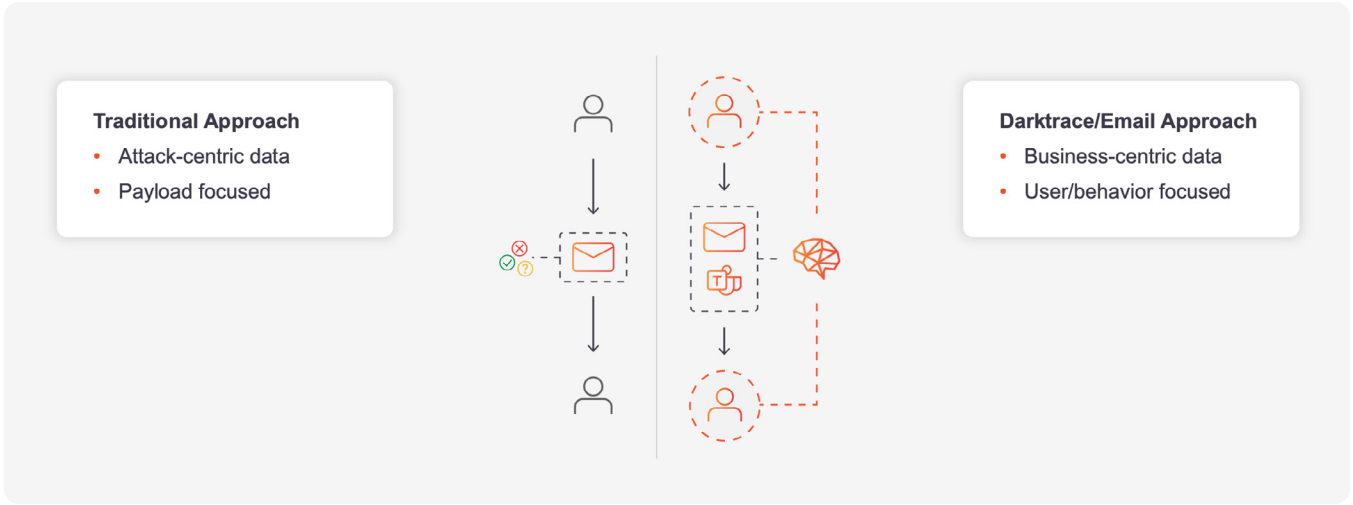
THE RIGHT REAL-TIME DECISIONS

Darktrace's AI instantly determines whether an email should be delivered to a particular inbox. By getting to know each user personally, Darktrace recognizes when internal or external employees have deviated from their typical behavior and potentially have been impersonated — and explains why.

NO SIGNATURE REQUIRED

Darktrace/Email improves upon historical and supervised learning approaches by:

- Not assuming the next attack will look like the last one, or the one before that
- Understanding normal to quickly identify what is and isn't suspicious
- Precisely responding to incidents faster than humans can, by taking the minimal action that reduces risk without causing business disruption



Darktrace AI takes the right action at the right time

While other tools typically offer just two options – "allow" or "deny" an email delivery — Darktrace mounts an automated response tailored to fit the situation. Understanding threats in context and what makes an email suspicious equips the platform to take the least aggressive action necessary to neutralize only the risky components of emails, without disrupting business continuity.

For example, if a link couched within an email is all that seems suspicious, Darktrace might block the link and allow the rest of the message to be delivered. Other appropriate autonomous response actions Darktrace/Email might take include:

- Moving an email to junk or a specific folder
- Rewriting links
- Removing or converting attachments
- Withholding and escalating messages with high threat level to the security team

Darktrace reduces benign user-reported emails by

60%

Internal Darktrace Research

> Much better than our previous system which allowed for users to click a dodgy link and required us to shutdown many accounts and manually unlock, Darktrace seems to eliminate all issues related to cybersecurity.

Desktop Support Engineer, Healthcare and Biotech

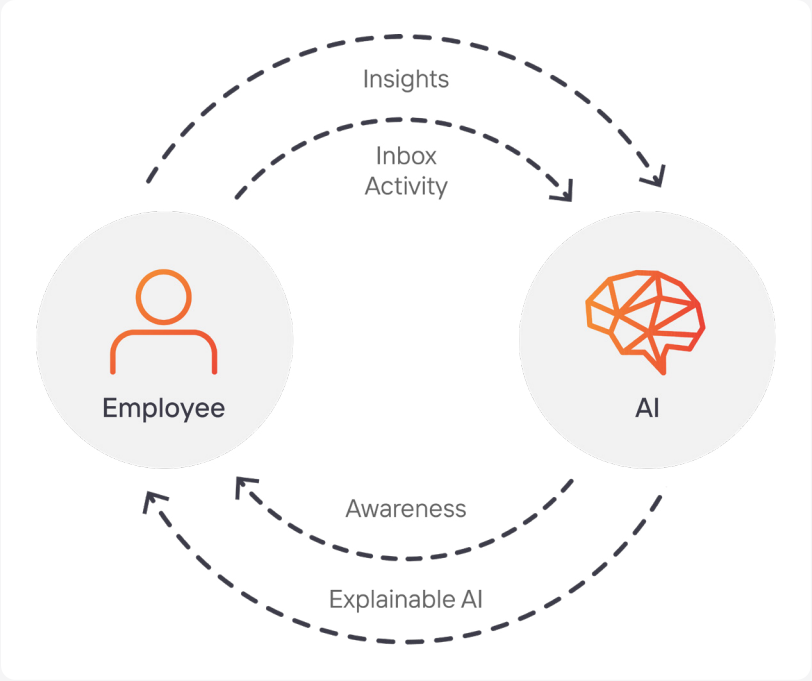
Gartner Peer Reviews

PLATFORM KEEPS USERS IN THE KNOW

Natural Language Processing (NLP) translates millions of data points and micro-decisions into clear digests that explain why the platform flagged an email and actions taken by AI in response.

Using Explainable AI, Darktrace/Email generates and sends actionable narratives to users so they can make informed decisions about opening or escalating messages for closer inspection by the security team.

Instead of seeing the same static banner at the top of every email that gets flagged, multi-layered AI can tailor warnings and available options to fit the individual threat and level of risk. Users might see a different notice every time, which reduces banner fatigue and keeps them actively engaged in the process as new threats appear.



AI improves user reporting from the ground up.

While other email security solutions assume most user reporting is false, Darktrace equips individuals with the intelligence needed to make the right call.

The detailed analysis and continuous coverage provided by Darktrace/Email creates an Employee-AI feedback loop that improves the detection algorithm and decision-making over time.

Platform approach up-levels operations

Darktrace/Email builds on the benefits of your native email security to stop unknown threats while eliminating mail latency, ongoing configuration maintenance, and duplicate costs across your IT estate. Faster detection and autonomous response free security teams from having to waste so much time investigating and releasing emails.

Instead of multiple dashboards, Darktrace's intuitive UI captures a user's activity across email, Microsoft, and Google accounts in a single pane of glass. Details include real-time snapshots of user identities and actioned emails, segmented by type of attack, and visibility into what happens after attackers get hold of an account that might point to account takeover.

DARKTRACE/EMAIL BY THE NUMBERS

Reduces the load on security teams by automating mailbox remediation that stops

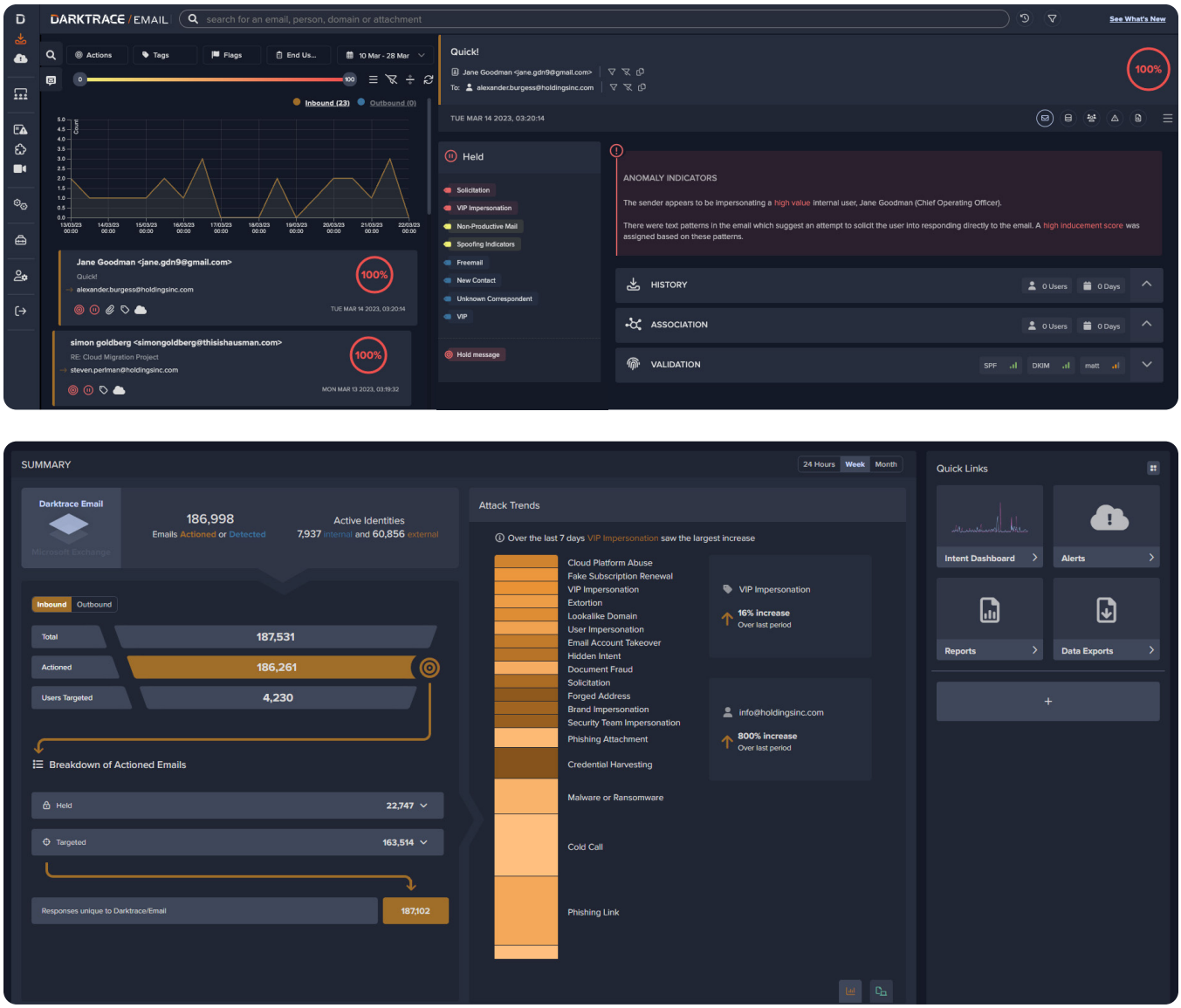
70%

more malicious links

Internal Darktrace Research

Leverages explainable AI to reduce triage time by

90%



The Darktrace/Email dashboard summarizes observations, detections, responses, and trends in a single pane of glass for the security team.

Other time-saving benefits include:

- Preemptively removing or sorting spam and other non-productive mail from user inboxes
- Previewing links and emails within the UI to spot indicators of phishing on a single screen
- Native install in minutes with Google, Microsoft 365, and On-Premise Exchange – no MX changes required
- Integrates with alerting tools and ticketing systems through flexible Darktrace APIs and syslog
- Supports multi-tenant and hybrid environments

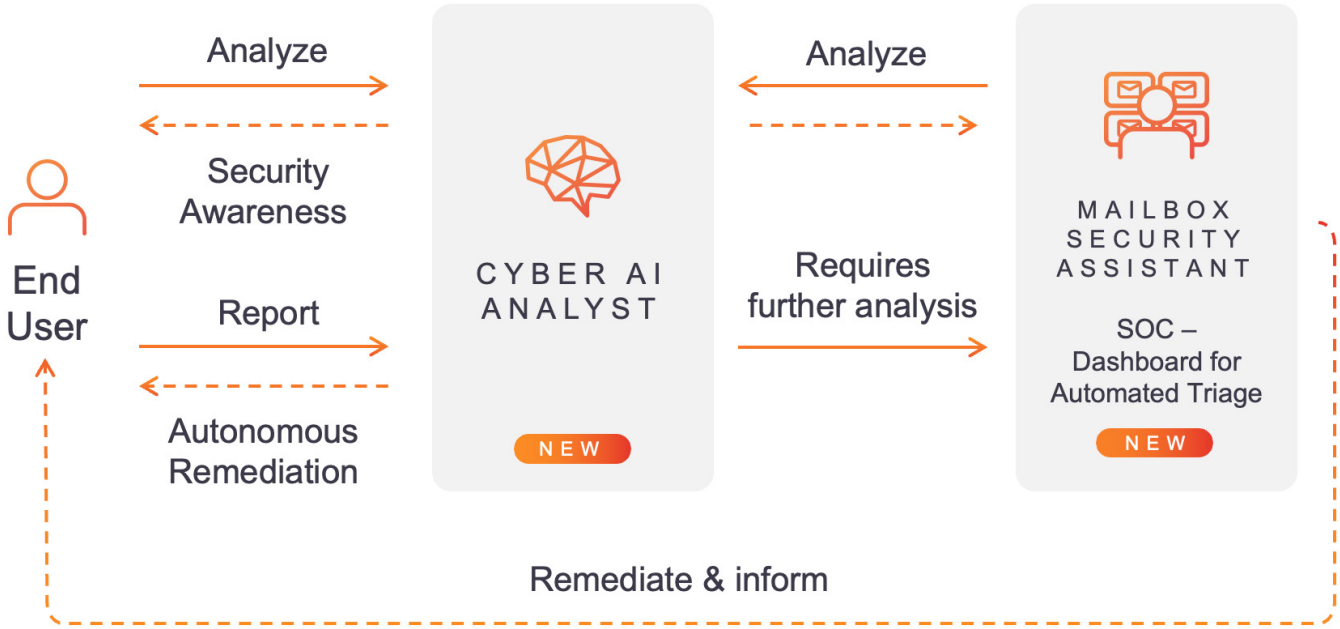
AI PIVOTS TOWARD PREVENTION

The insights generated by Darktrace's AI help security teams improve user education and fine-tune controls over time. For example, keeping track of when users move things in and out of email systems helps to refine policies for keeping inboxes clean. This in turn makes it easier to spot and remove phishing exploits before they cause harm.

Darktrace/Email Best-in-class Email Threat Protection

- Learns users' behavioral patterns to understand normal
- Recognizes when someone falls victim to phishing
- Takes the appropriate action in real time
- Integrates with the broader security ecosystem

Keep your defenders in control



Speed the mean time to respond with customizable AI investigations and automated remedation

Securing organizations in today's threat landscape and business environment necessitates a proactive approach that enhances the capabilities provided by native security vendors. AI can help defenders understand user behavior across your entire messaging attack surface, including inbound, outbound and lateral mail, and Microsoft Teams.

CONTACT DARKTRACE FOR A FREE TRIAL IN YOUR ENVIRONMENT.

About Darktrace

Darktrace (DARK.L), a global leader in cybersecurity Artificial Intelligence, delivers complete AI-powered solutions in its mission to free the world of cyber disruption. Its technology continuously learns and updates its knowledge of 'you' for an organization and applies that understanding to achieve an optimal state of cybersecurity. Breakthrough innovations from its R&D Centers have resulted more than 145 patent applications filed. Darktrace employs 2,200+ people around the world and protects over 9,000 organizations globally from advanced cyber-threats.

Scan to
LEARN MORE

DARKTRACE

Evolving threats call for evolved thinking™

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