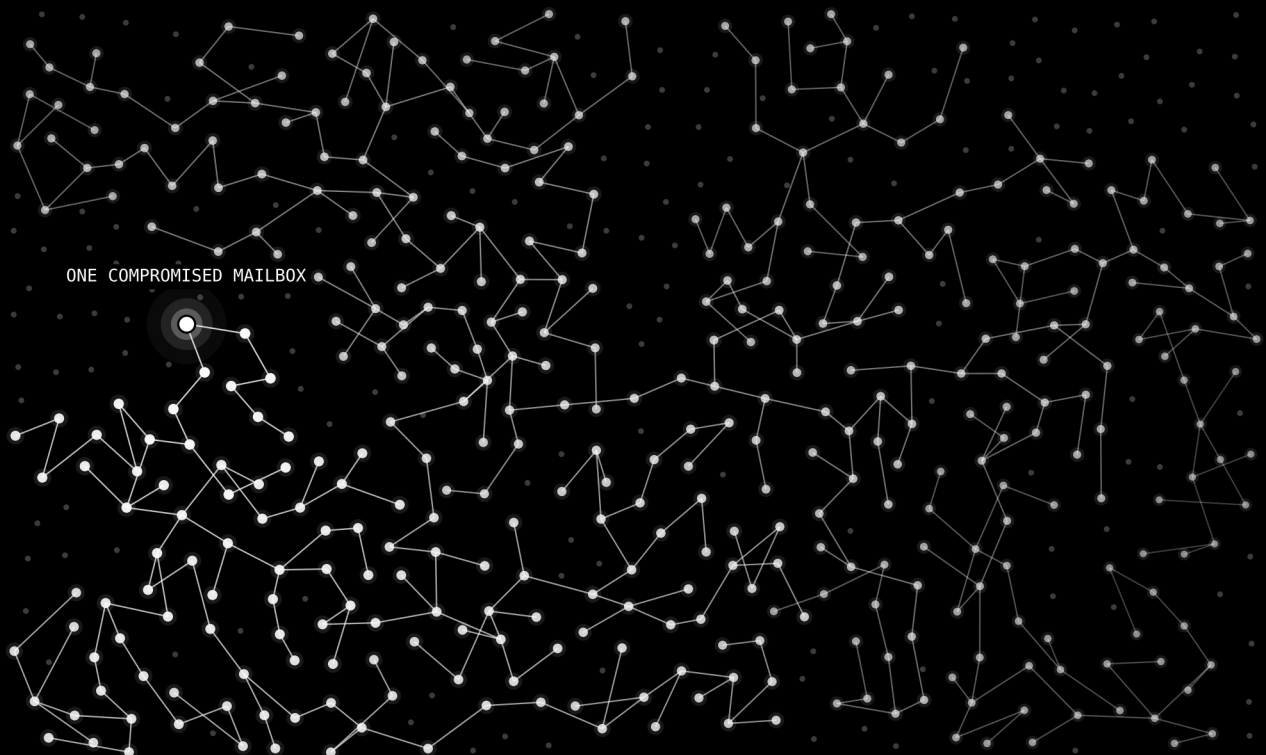




THREAT INTELLIGENCE REPORT · CONFIDENTIAL

AegisAI Q3 2026 State of the AI Attack Report

Based on more than 100 million emails across over 1 million protected mailboxes. The AegisAI threat research team confirmed every phishing verdict in this report.



73%

of tracked organizations reached by a single self-replicating calendar-invite worm, sent mailbox to mailbox from hijacked accounts.

AEGIS-IS-REDTIDE-002 · illustrative: each point is an organization, lit points were reached

AT A GLANCE

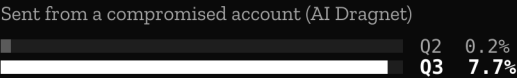
Q3 2026 at a Glance

Phishing rose 49%, driven by compromised accounts and new attack channels like the calendar.

THE TWO KEY ISSUES BEHIND THIS QUARTER'S MISSES AND INTRUSIONS

01 · COMPROMISED ACCOUNTS

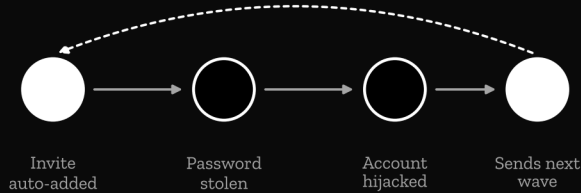
38x rise in AI Dragnet attacks sent from hijacked, trusted accounts



Passes SPF, DKIM and DMARC. Clean sending history.
Often someone already in your contacts.

02 · CALENDAR ATTACKS

73% of tracked organizations hit by one self-spreading calendar-invite worm repeats with each new victim



No click needed. Under 19 minutes from victim to next sender.
Up to 80% of malicious invites land despite native controls.

WHAT'S DRIVING THEM: AI AT BOTH ENDS OF THE PHISHING SPECTRUM

+88%

AI Spearphishing volume
targeted, AI-written lures, since Q1

11.6x

AI Dragnet share of phishing
one template, thousands of inboxes

3 in 10

phishing messages get past
provider filters, in every attack type

INSIDE THE REPORT

6

attack trends,
each with a real campaign

4

tracked intrusion sets,
incl. the REDTIDE worm

7

actions security teams
should take now

100M+

emails across 1M+ mailboxes,
confirmed by our researchers

Executive Summary

Phishing rose 49% this quarter, relative to overall email volume. Two issues drove most of what slipped past email defenses and most of the intrusions that followed: attackers sending from compromised accounts, and attackers delivering through the calendar. Both turn the signals defenders trust (a known sender, a clean authentication result, an email they can quarantine) into cover. AI speeds up both: it writes the lures, and it pushes a single template to thousands of organizations in minutes.

THE TWO KEY ISSUES

1 · Compromised accounts: attackers send as the people you trust (38x)

Attackers have always impersonated trusted brands and people. Now they take over real accounts and send from them. Their messages pass every authentication check, carry a clean sending history and often come from someone already in the recipient's contacts. AI Dragnet attacks from hijacked accounts jumped 38x (0.2% to 7.7%), and human-written phishing from compromised accounts climbed from 4.3% to 5.4%. In the same quarter, AI Dragnet attacks slipped past email providers' own filters more often (22.2% to 30.6%) and passed DMARC more often when they did (61.9% to 73.2%). Every hijacked account is an intrusion in its own right, and attackers use it to launch the next attack.

2 · Calendar attacks: phishing that lands with no click and outlives the email (73%)

Calendar platforms add a malicious invite to the recipient's calendar automatically, before they do anything. Quarantining the email leaves the event in place, so the link sits next to real meetings and its reminder can fire days later. In AegisAI's tracked campaigns, up to 80% of malicious invites still reached calendars with native controls switched on, because the "known senders only" setting trusts compromised accounts. The quarter's largest campaign combined both issues: AEGIS-IS-REDTIDE-002, a calendar-invite worm sent from hijacked Google Workspace accounts, reached 73% of AegisAI's tracked organizations. Each victim who entered a password became the next sender, in one case in under 19 minutes.

What's driving them: AI at both ends of the phishing spectrum

AI Spearphishing (+88%) aims AI-written phishing at one person or a small group, built from real context about their role, vendors and email threads. It grew faster than any other category: volume rose 88% since Q1 2026, and its share of all confirmed phishing climbed from 9.7% to 12.3%.

AI Dragnet attacks (11.6x) blast one AI-written template at as many organizations as possible, often within minutes. Their share of all confirmed phishing grew 11.6x (0.5% to 5.8%), and they show the sharpest shift to compromised accounts and calendar delivery.

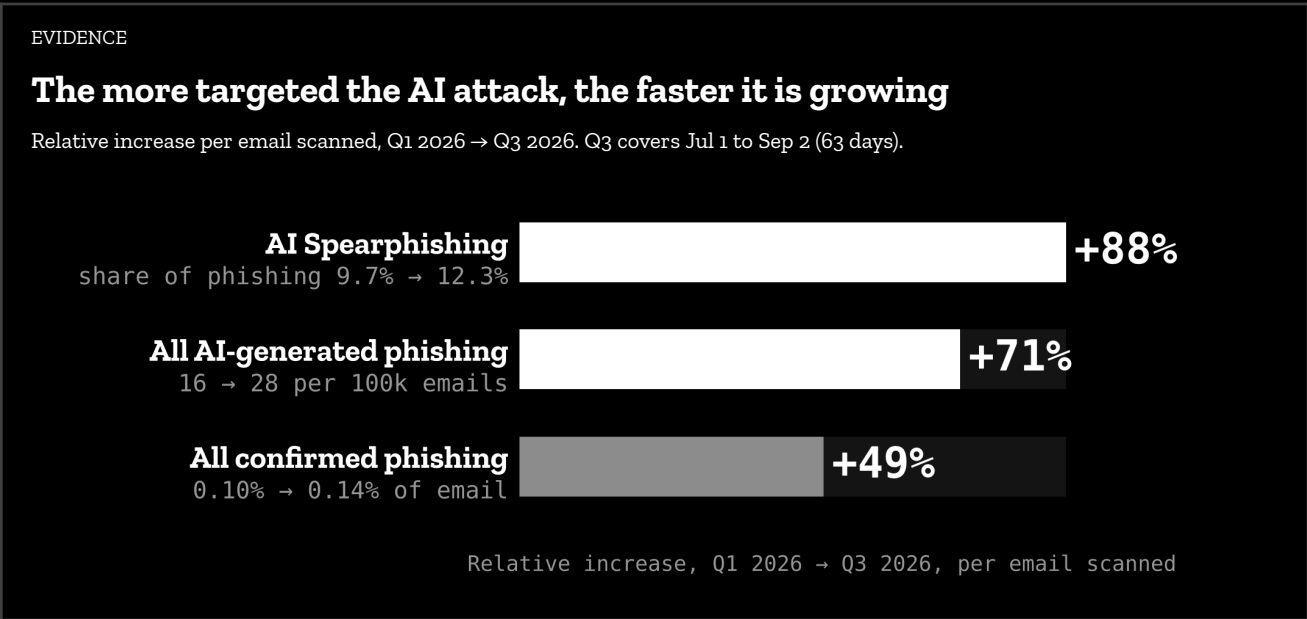
Why speed and investigation matter

These campaigns move in minutes: one reached 18% of tracked organizations in three minutes, and the worm turned victims into senders in under 19. Defenses that act after delivery start behind. Stopping these attacks takes a verdict before delivery, built on active investigation of what each message claims rather than on who sent it. Section 06 shows how Proteus, AegisAI's own foundational model trained on email, does this: on a benchmark of 3,020 unseen attacks, it identified more than 99%, against 81.3% for the leading general-purpose model, and reached its median verdict 7.2x faster.

The bottom line: attackers now exploit the signals defenders lean on most: a known sender, a passing authentication check, a quarantined email. Roughly three in ten confirmed phishing messages of every attack type slip past the recipient's email provider entirely. Section 03 breaks this into six trends, each with a real campaign and what to do about it.

TOTAL CONFIRMED PHISHING	AI-GENERATED VOLUME	AI SPEARPHISHING VOLUME
+49%	+71%	+88%
from 0.10% to 0.14%	from 0.016% to 0.028%	share of phishing 9.7% → 12.3%

Rates show confirmed phishing as a share of all email scanned. We calculate percent changes from unrounded rates, so they can differ slightly from the rounded figures shown.



+88%

growth in AI Spearphishing volume since Q1 2026, the fastest of any category.

WHAT THAT MEANS

Phishing overall grew 49%. AI-generated phishing grew faster, and targeted AI phishing faster still. A larger share of what reaches inboxes is now AI-written and aimed at a specific person.

AI Dragnet attacks are the clearest sign that attacks are getting harder to stop. They used to arrive from throwaway infrastructure that sender-reputation filters caught easily. This quarter they moved into compromised, legitimate mailboxes, and the most striking example spreads itself: a Google Calendar invite worm that adds itself to each recipient's calendar with no click, harvests their password, then sends the next wave from their account (AEGIS-IS-REDTIDE-002, section 05).

AI DRAGNET SENDER COMPROMISE

7.7%

+7.5pt vs. prior quarter

AI DRAGNET EVASION RATE

30.6%

+8.5pt vs. prior quarter

AI DRAGNET SHARE OF CONFIRMED PHISHING

5.8%

+5.3pt vs. prior quarter

AI Dragnet: how much each measure multiplied since Q2 2026

Sent from hijacked accounts **38x**
sender compromise 0.2% → 7.7%

Share of all phishing **11.6x**
0.5% → 5.8%

Got past provider filters **1.4x**
evasion rate 22.2% → 30.6%

Multiple of prior-quarter value, Q2 2026 → Q3 2026

02 · UNDERSTANDING THE ATTACKS

Understanding the Attacks

This report sorts every confirmed phishing message into one of three attack types. The split answers two questions: was the message written by AI, and was it aimed at a few people or blasted at many? A third pattern, trust hijacking, cuts across all three: it describes where attackers send from, not how they write.

AI Spearphishing

What it is. Phishing written by AI for one specific person or a small group. Think of a sniper rather than a shotgun. The attacker, or an AI agent working for them, gathers what's public about the target (their role, their manager, their vendors, a deal or project in progress) and has a language model write a message that fits. The result reads like a real colleague or supplier, because it's built from real context.

Why it matters. Spearphishing used to be expensive: an attacker had to research each target and write each message by hand, so attackers saved it for high-value people. AI removes that cost, so tailored attacks can now be sent at volume. That's why AI Spearphishing is the fastest-growing category this quarter (+88% volume).

What it looks like. A reply inside an existing email thread (thread hijacking is the top AI Spearphishing lure, at 20.3% of volume), a vendor asking to update banking details, or an urgent request that appears to come from an executive. Business email compromise (BEC) phrasing is more common here than in any other attack type.

AI Dragnet attacks

What it is. One AI-written phishing template sent to as many organizations as possible at once, like a fishing trawler's dragnet that catches whatever is in its path. There's no research on individual targets; the attacker relies on reach. The same message, sometimes byte-for-byte, can land in hundreds of organizations within minutes.

Why it matters. Dragnet phishing used to be the easiest to catch: generic wording, throwaway domains, quickly blocklisted. AI changes the first part, because the template now reads fluently and even sounds personal (AI Dragnet scores highest of the three attack types for personal-sounding language). Trust hijacking changes the rest: this quarter, AI Dragnet attacks started sending from real, compromised accounts and now get past provider filters more often than any other attack type.

What it looks like. AEGIS-IS-REDTIDE-002, a calendar-invite worm, reached 73% of AegisAI's tracked organizations by sending each new wave from its previous victims' accounts. AEGIS-IS-FILLTHEBILL-004 reached 18% of tracked organizations in three minutes with one identical "billing document" message.

Human-written phishing

What it is. Confirmed phishing that does not score as AI-generated. We measure the two AI attack types against it.

Why it matters. Human-written phishing is not going away, and it can still be highly sophisticated. AEGIS-IS-CAPTCHATHEBILL-005, the most technically advanced campaign in this report, needed no AI at all.

Trust hijacking (applies to all three)

What it is. Sending phishing from a real, legitimate account the attacker has taken over, instead of from an address or domain they created. Impersonation pretends to be a trusted sender; trust hijacking is the trusted sender.

Why it matters. Most email defenses, and most people, judge a message partly by who sent it. A hijacked account passes authentication checks (SPF, DKIM, DMARC), has a clean sending history, and may be someone the recipient has emailed before. Every one of those signals says “safe.” Each compromised account also becomes a launch point for the next attack, which is how REDTIDE spread on its own.

How common it is. Sender compromise rose in two of three attack types this quarter: AI Dragnet from 0.2% to 7.7%, and human-written from 4.3% to 5.4%. Among phishing that evaded provider filters, AI-generated attacks passed DMARC more often than last quarter.

Key terms used in this report

Confirmed phishing. A message the AegisAI threat research team verified as phishing. We express volumes as a share of all email scanned, so they reflect attack activity rather than changes in total mail.

Evasion rate. The share of confirmed phishing that the recipient’s own email provider let through to the inbox, caught only by AegisAI. A higher rate means more attacks are slipping past standard filtering.

Sender compromise. The share of phishing sent from a real, legitimate account an attacker had taken over, rather than from an address or domain the attacker set up.

DMARC pass rate. The share of messages that passed DMARC, the standard check that a message really came from the domain it claims. Phishing that passes DMARC looks authentic to filters.

Intrusion set. A cluster of related phishing activity AegisAI tracks as a single operation, linked by shared infrastructure, templates and timing. We name each one AEGIS-IS-**<NAME>**-**<number>**.

03 · TOP ATTACK TRENDS

Top Attack Trends

Attacks are increasing, and the signals people and filters use to catch them are weakening. Each tile gives the headline number, a real campaign that shows it, and what to do about it.

03 Trust hijacking: attacks come from compromised accounts

38x

rise in AI Dragnet attacks sent from hijacked accounts
(0.2% → 7.7%)

EXAMPLE REDTIDE, a worm sent from hijacked accounts, reached 73% of tracked organizations.

WHAT TO DO Treat a known sender as a starting point, not proof.

04 Cheap attacks slip past filters as often as sophisticated ones

3 in 10

messages in every attack type get past provider filters, cheap or targeted

EXAMPLE FILLTHEBILL reached 18% of tracked organizations in 3 minutes.

WHAT TO DO Add a layer that judges intent, not just reputation.

03 **Trust hijacking: attacks come from compromised accounts**

38x

rise in AI Dragnet attacks sent from hijacked accounts
(0.2% → 7.7%)

EXAMPLE REDTIDE, a worm sent from hijacked accounts, reached 73% of tracked organizations.

WHAT TO DO Treat a known sender as a starting point, not proof.

04 **Cheap attacks slip past filters as often as sophisticated ones**

3 in 10

messages in every attack type get past provider filters, cheap or targeted

EXAMPLE FILLTHEBILL reached 18% of tracked organizations in 3 minutes.

WHAT TO DO Add a layer that judges intent, not just reputation.

05 **Attackers borrow trust from real brands and services**

44%

of credential-theft links land on trusted platforms; Google Workspace is the top impersonated brand

EXAMPLE FAKESERVICE sent through a two-decade-old domain with a 100/100 reputation score.

WHAT TO DO Judge a message by what it asks for, not where it seems to come from.

06 **The old red flags are disappearing**

+570

words longer, on average, than human-written phishing, and more personal-sounding

EXAMPLE CAPTCHATHEBILL shows security scanners a harmless page and real targets the phish.

WHAT TO DO Retire training built on spotting typos and generic greetings.

04 · TECHNICAL DEEP DIVE

Technical Deep Dive

AI-generated content reads more human across the board, and AI Dragnet attacks, the cheapest kind, moved from disposable infrastructure into compromised, legitimate mailboxes.

Three very different kinds of phishing ended the quarter almost equally hard to catch.

Unless a chart states a comparison, figures cover Q3 to date, Jul 1 to Sep 2 (63 days). Each subsection explains how we calculate its measures; section 08 covers the full methodology.

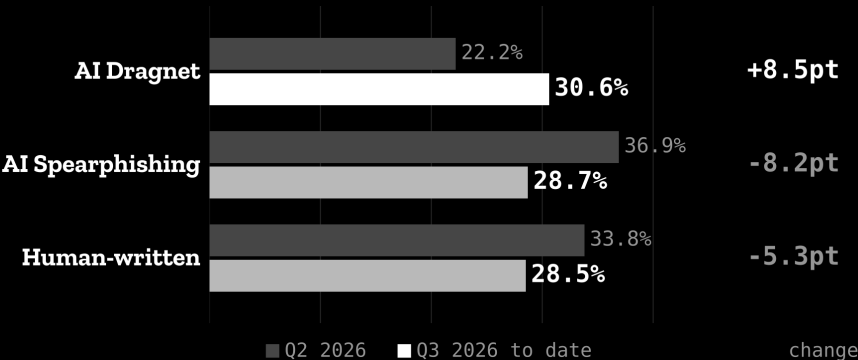
Evasion Mechanics

Evasion fell for AI Spearphishing and human-written phishing, and rose for AI Dragnet attacks, the cheapest kind.

HOW WE MEASURE A message counts as evaded when the recipient's own email provider delivered it to the inbox and AegisAI was the only control that flagged it. We compare rates against confidence intervals on the prior quarter to separate real change from sampling noise.

Evasion rose for AI Dragnet and fell for the other two attack types.

Evasion rate: the share of confirmed phishing that got past the recipient’s own email provider spam filter entirely, caught only by AegisAI. Prior quarter → this quarter, by attack type.



Baseline: Q2 2026 (91 days). Current: Q3 to date, Jul 1 to Sep 2 (63 days).

Average evasion across the three attack types: 29%, with a 2.1-point spread between them, down from 14.7 points in Q2.

AI Dragnet evasion rose from 22.2% to 30.6%, above the 26.0% upper bound of its Q2 confidence interval, so the smaller Q2 sample does not explain the jump. AI Spearphishing fell from 36.9% to 28.7% and Human-Written from 33.8% to 28.5%. The three attack types now sit within 2.1 points of each other (28.5% to 30.6%), against a 14.7-point spread in Q2 (22.2% to 36.9%).

Content-Based Personalization

The oldest advice for spotting a fake email, that it reads stiff, generic and impersonal, no longer holds. AI-generated phishing now reads more personalized than human-written phishing, and the gap is widening.

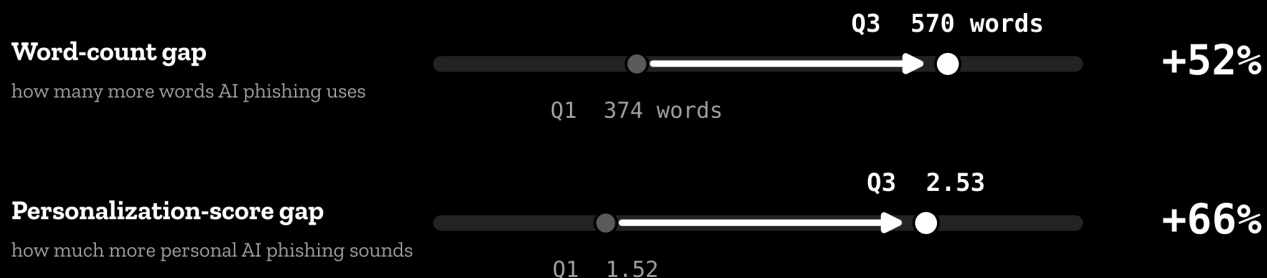
HOW WE MEASURE We score each message for how strongly it reads as written for a specific recipient: references to their name, role, organization or a prior conversation. Higher scores mean more personal-sounding language, not necessarily more individual targeting. The gap figures compare the average AI-generated message with the average human-written one.

Avg. content-personalization confidence by attack type

HUMAN-WRITTEN	AI DRAGNET	AI SPEARPHISHING
2	6	5

Average confidence score that a message is written for a specific recipient. Higher means more personal-sounding language.

AI Dragnet’s personalization-confidence score fell from 8.4 in Q2 to 5.8, still above AI Spearphishing’s 4.9 and the highest of the three attack types. As in Q2, personalization cues do not track fan-out: AI Dragnet has the least individually targeted structure and the most personalized-sounding language.

Word count and personalization gap: AI-generated vs. human-written, Q1 2026 → this quarter

Gap = average AI-generated phishing minus average human-written phishing. Both differences are statistically significant.

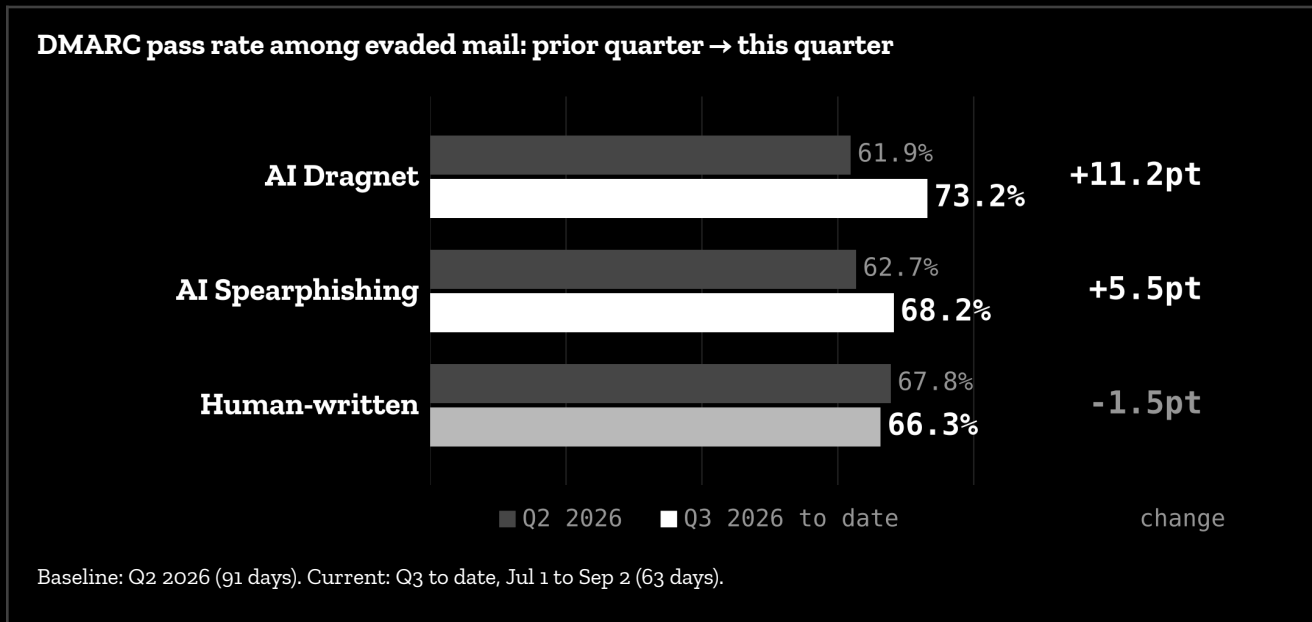
AI-generated phishing runs measurably longer and more personalized than human-written phishing, and both gaps widened this quarter. Both differences are statistically significant. "AI-written phishing is short, generic and impersonal" is now backwards as a rule of thumb. Length and a personal touch no longer indicate a genuine message.

SO WHAT The two heuristics people use to spot AI phishing at a glance, "it reads short" and "it feels generic," are losing reliability. Both gaps widened again this quarter, as AI volume rose fastest.

Trust Hijacking: From Impersonation to Compromise

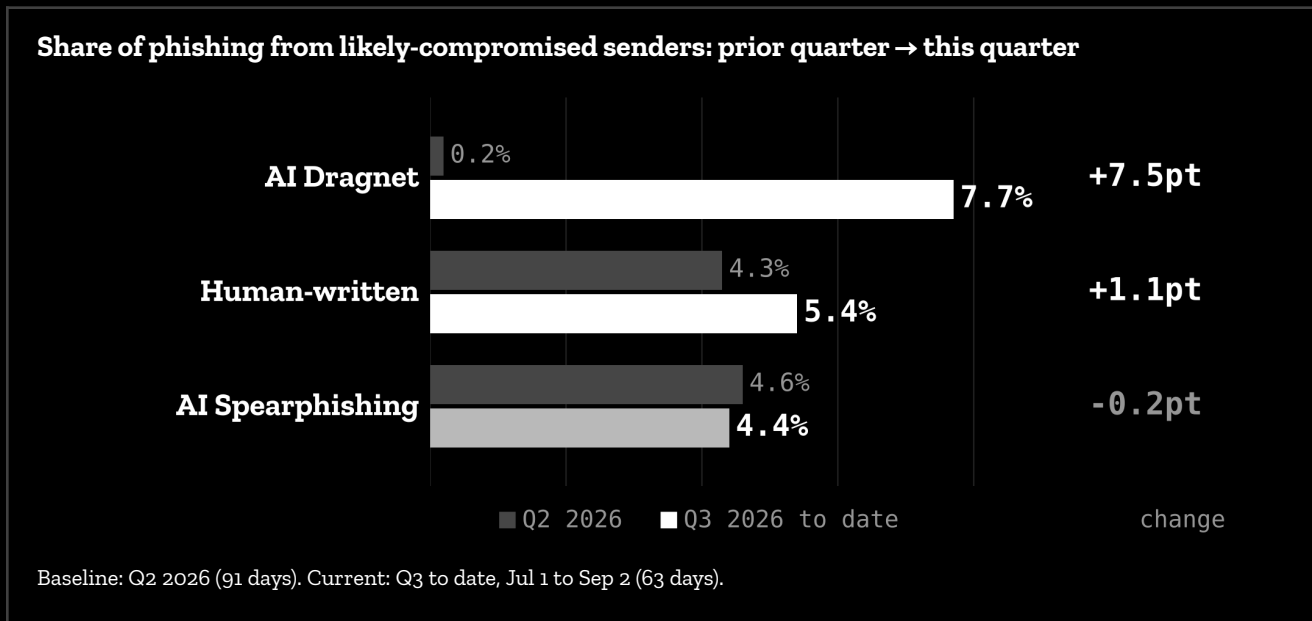
Attackers are no longer just pretending to be trusted senders. They are taking over trusted accounts and sending from them. We call this trust hijacking. It showed up most sharply in AI Dragnet attacks, and four measures line up with it: DMARC pass rates on evaded mail rose, the sender-compromise rate rose from 0.2% to 7.7%, domains are reused, and sends concentrate in narrow, scripted windows.

HOW WE MEASURE DMARC results come from the authentication results on each evaded message. A sender counts as likely compromised when phishing arrives from a real, previously legitimate account that authenticates correctly, rather than from attacker-created infrastructure. Domain reuse is the share of an attack type's unique sending domains seen two or more times in the quarter.



Among evaded mail, DMARC pass rates diverged by attack type. AI Dragnet rose from 61.9% to 73.2% (+11.2pt) and AI Spearphishing from 62.7% to 68.2% (+5.5pt), while Human-Written slipped from 67.8% to 66.3% (-1.5pt). Both AI-generated attack types moved toward infrastructure that authenticates more cleanly; Human-Written did not.

AI Dragnet's sender-compromise rate rose from 0.2% to 7.7% this quarter, the largest single behavior change this report tracks.



AI Dragnet's sender-compromise rate rose from 0.2% to 7.7%, far above the 1.1% upper bound of its Q2 confidence interval. AI Spearphishing was essentially flat (4.6% to 4.4%) and Human-Written rose modestly (4.3% to 5.4%). AI Dragnet, previously the attack type least reliant on compromised legitimate infrastructure, moved the most.

SO WHAT A jump from 0.2% to 7.7% in one quarter flips the ordering. AI Dragnet went from the attack type most reliably caught on sender reputation alone to the attack type least caught that way. Its sender-compromise rate (7.7%) is now the highest of the three, ahead of Human-Written (5.4%) and AI Spearphishing (4.4%).

AI Dragnet increasingly runs on a small set of reused domains, not disposable one-time infrastructure.

Attackers who register their own domains tend to leave fingerprints: hyphens and digits to build lookalikes (service-nowinc.com, login-secure24.net), long names, and extra subdomains stacked in front of the real domain (accounts.example-login.com). Established, legitimate domains tend to be shorter and cleaner. AI Dragnet’s sending domains score lower than human-written phishing on every one of these measures, which fits a shift toward sending from real, compromised domains rather than freshly registered lookalikes.

What sending domains look like, by attack type

	HUMAN-WRITTEN	AI DRAGNET	AI SPEARPHISHING
Domains containing a hyphen	8.1%	7.3%	—
Domains containing a digit	3.5%	2.9%	—
Avg. domain length (characters)	19.2	16.3	—
Avg. subdomains in front of the main domain	0.103	0.057	—

— = not reported for AI Spearphishing this quarter.

Domain reuse shows whether attackers are burning a fresh domain for every send or rotating a small, stable set. High reuse fits steady, automated operations running on infrastructure that already works.

Domain reuse: share of each attack type’s sending domains used more than once

HUMAN-WRITTEN	AI DRAGNET	AI SPEARPHISHING
45%	54%	43%

We report domain-diversity and domain-reuse figures for this quarter only. Q3 to date is a 63-day window against Q2’s 91 days, and a shorter window mechanically inflates apparent diversity, so we draw no Q2 comparison. AI Dragnet has the highest reuse concentration of the three attack types: it sent from 53.8% of its 926 unique domains two or more times, and those repeat domains carried 91.6% of its volume. That pattern fits a small set of domains in rotation, not one-time infrastructure.

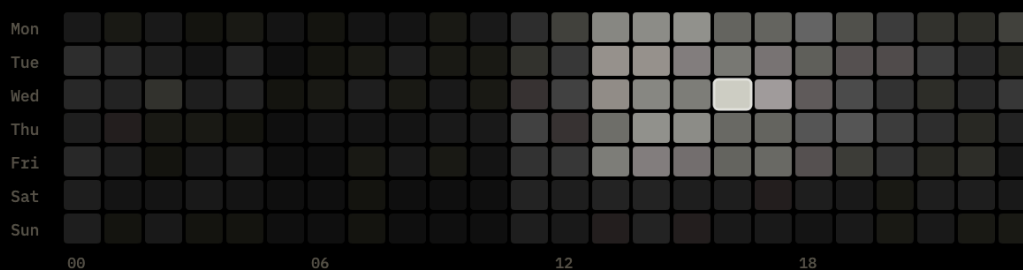
AI Dragnet sends in narrow, scripted windows, while the other two attack types send on a broad, organic schedule. The pattern is consistent with automated sending.

HOW TO READ THESE GRIDS Each grid shows when one attack type is sent: days down the side, hours (UTC) across the top. Brighter squares mean a larger share of that attack type’s own messages.

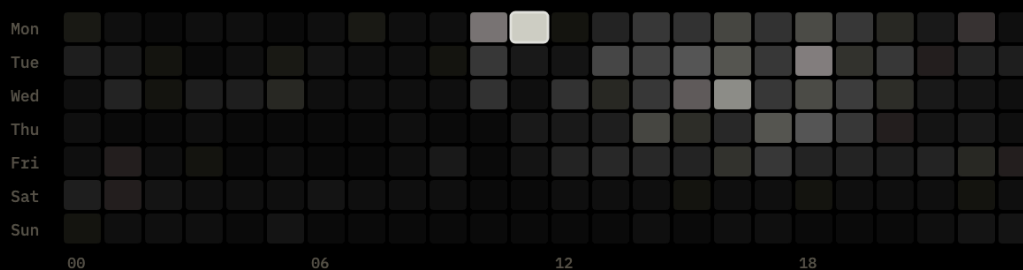
A single bright spot means sends fire on a schedule, the signature of a script. A broad bright band during business hours looks like people working. Compare the shape of each grid, not brightness between grids.

Human-Written: send timing (day × hour, UTC)

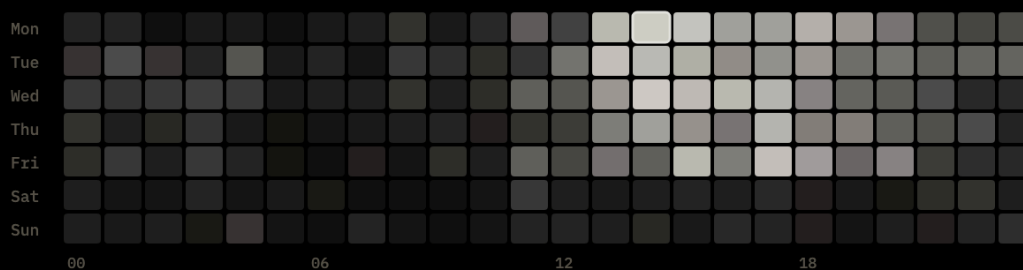
Peak: Wednesday 16:00 UTC · 2.9% of Human-Written's volume.

**AI Dragnet: send timing (day × hour, UTC)**

Peak: Monday 11:00 UTC · 5.5% of AI Dragnet's volume.

**AI Spearphishing: send timing (day × hour, UTC)**

Peak: Monday 14:00 UTC · 2.0% of AI Spearphishing's volume.



Human-Written phishing peaks midweek, on Wednesday around 16:00 UTC, with just under 3% of its volume in that single hour. Weekday business hours (09:00-17:00 UTC) account for 54.3% of its volume and weekends for 10.9%. AI Spearphishing has a similarly broad shape. Its peak falls on Monday around 14:00 UTC, with a flatter single-hour concentration (2.0%) and a comparable business-hours share (49.7%).

AI Dragnet breaks the pattern, as it did in Q2. Its peak falls on Monday at 11:00 UTC and concentrates 5.5% of its entire volume into that one hour, more than either other attack type, while carrying the lowest

weekend share of the three (7.8%). A narrow, scheduled spike that repeats across quarters is consistent with a scripted send process, not individually timed sending.

What Didn't Change: The Lures

None of this required a new attack type or new content. It required a different kind of account to send from.

Google Workspace is now the most impersonated brand in every attack type, most heavily in AI Dragnet.

Brand impersonation rate by attack type

	HUMAN-WRITTEN	AI DRAGNET	AI SPEARPHISHING
Google Workspace	12%	33%	25%
Social Media	6%	18%	11%
Microsoft 365	5%	9%	9%
Apple Ecosystem	3%	7%	4%
Retail/E-commerce	3%	6%	7%
DocuSign	1%	2%	2%

Google Workspace is the largest single impersonated brand this quarter in all three attack types: 11.6% of Human-Written volume, 25.2% of AI Spearphishing volume and 33.0% of AI Dragnet volume. Retail and e-commerce brands led Human-Written and AI Spearphishing in earlier quarters. We have not recomputed Q2 brand shares against this quarter's classification pipeline, so we offer no cause for the shift. Social media brands hold second place in all three attack types (6% to 18% depending on attack type).

Thread hijacking leads for Human-Written and AI Spearphishing this quarter. AI Dragnet splits almost evenly between thread hijacking and invoice/payment fraud.

Attack-type signals by attack type

	HUMAN-WRITTEN	AI DRAGNET	AI SPEARPHISHING
Thread Hijacking	18%	18%	20%
BEC	1%	2%	5%

The text below covers invoice and payment fraud, the narrow leader for AI Dragnet (19.4%).

Thread hijacking leads for Human-Written (17.5% of volume) and AI Spearphishing (20.3%). Invoice and payment fraud leads narrowly for AI Dragnet (19.4%, against 17.7% for thread hijacking), too close to call

within that attack type. We have not re-verified Q2 attack-type standings for Human-Written and AI Spearphishing, so we draw no quarter-over-quarter comparison.

Stylistic tells, not hard technical artifacts, still drive most AI-generated verdicts this quarter, as in Q2.

Stylistic LLM-artifact signals by attack type

	HUMAN-WRITTEN	AI DRAGNET	AI SPEARPHISHING
Absence of contractions	22%	82%	79%
Low lexical diversity	23%	62%	59%

Template/pretext LLM-artifact signals by attack type

	HUMAN-WRITTEN	AI DRAGNET	AI SPEARPHISHING
BEC finance-urgency phrasing	0.001%	5%	7%
False-context claim ("as we discussed")	0%	4%	1%
Unfilled template placeholder	0%	2%	2%

Stylistic signals (absence of contractions, low lexical diversity) remain the main drivers of the AI-generated verdict, as in Q2, and together account for the large majority of tallied hits in every attack type. Hard artifacts such as tool-name leaks and unfilled placeholders are a small minority. Nothing here is new.

That points to the account, not the lure, as what changed: the same attack types and brand targets, now sent from compromised accounts.

TRUST HIJACKING This quarter's defining shift. Impersonation fakes a trusted sender; trust hijacking uses a real one. A hijacked account passes authentication, carries a clean history and may already be in the recipient's contacts, so reputation-based filtering and "do I know this sender?" training both wave it through. Each compromised account then becomes the sender of the next attack. The intrusion sets that follow show it in practice, starting with a worm that spreads mailbox to mailbox.

05 · TRACKED INTRUSION SETS

Tracked Intrusion Sets

Four campaigns show this quarter's trends in real inboxes. Three are AI Dragnet. The fourth is human-written, included because it shows that the evasion techniques now spreading don't require AI.

- **AEGIS-IS-REDTIDE-002** AI Dragnet. A Google Calendar invite that adds itself to your calendar and makes you its next sender. Trend 03.

- **AEGIS-IS-FAKESERVICE-003** AI Dragnet. A fake ServiceNow invoice that rebuilt its infrastructure three times to collect one payment. Trend 05.
- **AEGIS-IS-FILLTHEBILL-004** AI Dragnet. A billing-portal link that arrives with your email address already filled in. Trend 04.
- **AEGIS-IS-CAPTCHATHEBILL-005** Human-written. A document notice that shows a security scanner one page and an employee another. Trend 06.

The first, REDTIDE, puts AI Dragnet's jump in sender compromise (0.2% to 7.7%) inside a single campaign.

INTRUSION SET · AEGIS-IS-REDTIDE-002 · AI DRAGNET · TREND 03

A calendar-invite worm that turns each victim into the next sender

AT A GLANCE Reach: 73% of tracked organizations · Delivery: compromised Google Workspace mailboxes · Spread: self-replicating, mailbox to mailbox · Goal: Google credential theft · Active: late July to mid-September

REDTIDE is a worm, not a campaign in the usual sense, and it is AI Dragnet's jump in sender compromise (0.2% to 7.7%) seen as a single cluster. It reached 73% of AegisAI's tracked organizations. Every invitation is a genuine Google Calendar invite, generated by Google's own infrastructure from a real Google Workspace account, DKIM-signed and clean of any prior bad history. Google adds it to the recipient's calendar automatically, with no click and no acceptance required.

How the worm spreads

The invite links to a fake "verify you're human" gate and, behind it, a lookalike Google sign-in page. When a recipient enters their password, their mailbox becomes the next sender, and the next wave goes to their contacts. The target list grows from the worm's own victims rather than from an attacker's list. In the fastest replication AegisAI confirmed, under 19 minutes passed between an invite reaching a corporate mailbox and a different mailbox at the same company sending the next wave. A second branch of the same lure delivers a malware dropper instead of the sign-in page.

Why calendar attacks are so hard to clean up

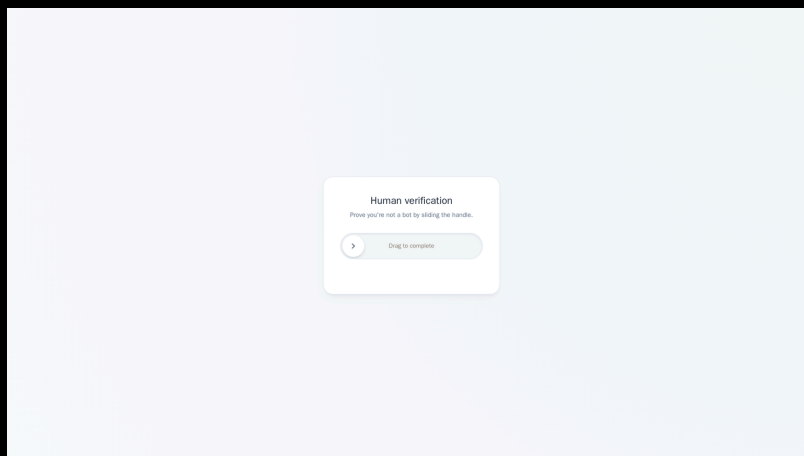
Calendar phishing outlives the email that delivered it. Quarantining the message does not delete the event it created, so the malicious link stays on the calendar next to real meetings, and its reminder can fire hours or days later when the original email is long forgotten. In AegisAI's tracked campaigns, up to 80% of malicious invites still landed on calendars with native controls switched on, because the "known senders only" setting trusts compromised accounts that have emailed the target before. Clean-up has to remove the calendar event along with the email, and sweep back through invites that already got through.

Each invitation posed as a bid, invoice-collection or voicemail notice and pointed to a CAPTCHA-style human-verification gate hiding a credential-harvesting page styled after Google's sign-in, on landing hosts that share an accounts.* naming pattern (see Indicators, below). Every message came from a real, previously

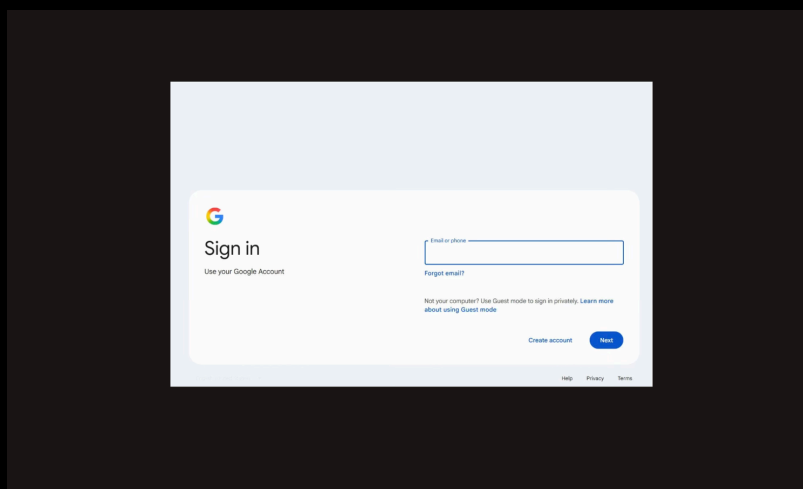
compromised Google Workspace mailbox, DKIM-signed and clean of any prior bad history. The sender genuinely was that account.

The invitation text carries the AI Dragnet markers described earlier: one templated bid notice, reused verbatim down to a fabricated reference number across unrelated targets, in the formal, contraction-free register of this report's AI-artifact tally. That places it in the AI Dragnet category, not in human-written or AI Spearphishing content.

Read more from AegisAI: [Google Calendar phishing and the RMM agent connection](#), [Red Tide: a calendar-invite worm that spreads mailbox to mailbox](#) and [Calendar phishing remediation](#).



The landing gate on one of this cluster's domains (accounts.tlcprroofings[.]com), as AegisAI's crawler captured it: a fake human-verification slider, not the Google-styled sign-in page. The same anti-bot mechanic appears on unrelated infrastructure elsewhere in this report.



The credential-harvesting page behind the gate: a spoofed Google sign-in, with no pre-filled recipient detail visible.

SO WHAT? Delivery needs no click and no reputation. The calendar entry exists before the recipient does anything, sent by an account with a real, clean sending history. The worm bypasses domain reputation, attachment scanning and “don’t click unknown links” training at the delivery step instead of defeating them one by one. Domain rotation defeats static blocklisting the same way: with 60+ distinct landing hosts observed, and new ones appearing through the full window, blocking any single domain has negligible effect on throughput. The layer that still catches it is the one that does not depend on the email or link: detecting that a compromised account’s sign-in and mailbox behavior has changed.

MITRE ATT&CK

T1586.002 Compromise Accounts: Email Accounts
 T1078 Valid Accounts
 T1566.002 Phishing: Spearphishing Link
 T1656 Impersonation

INDICATORS (REPRESENTATIVE, NOT EXHAUSTIVE)

INDICATOR	TYPE
accounts.ffzwi[.]icu	Landing domain, 7 tenants, 75 messages, Aug 21–26
accounts.gmggxcvrrqoqeshjdnhrxvqcr[.]icu	Landing domain, 5 tenants, 51 messages, Jul 22–Sep 8, longest-lived single host
accounts.tlcproroofigs[.]com	Landing domain, 7 tenants, 23 messages, Sep 14–17, compromised small-business site
accounts.wqrtysnfhrhrhppez[.]icu	Landing domain, 7 tenants, 23 messages, Aug 14–21
accounts.zojpyx[.]icu	Landing domain, 10 tenants, 16 messages, Aug 10–Sep 16, longest calendar span
"Invitation: ..." / "Canceled event: ..." / "New event: ..."	Subject format, Google Calendar’s own auto-generated notification prefixes

INTRUSION SET · AEGIS-IS-FAKESERVICE-003 · AI DRAGNET · TREND 05

A recurring wire-fraud persona that survived three rounds of its own infrastructure

AT A GLANCE Reach: over half of protected organizations · Delivery: aged, high-reputation domains via Amazon SES · Goal: wire fraud · Active: seven weeks

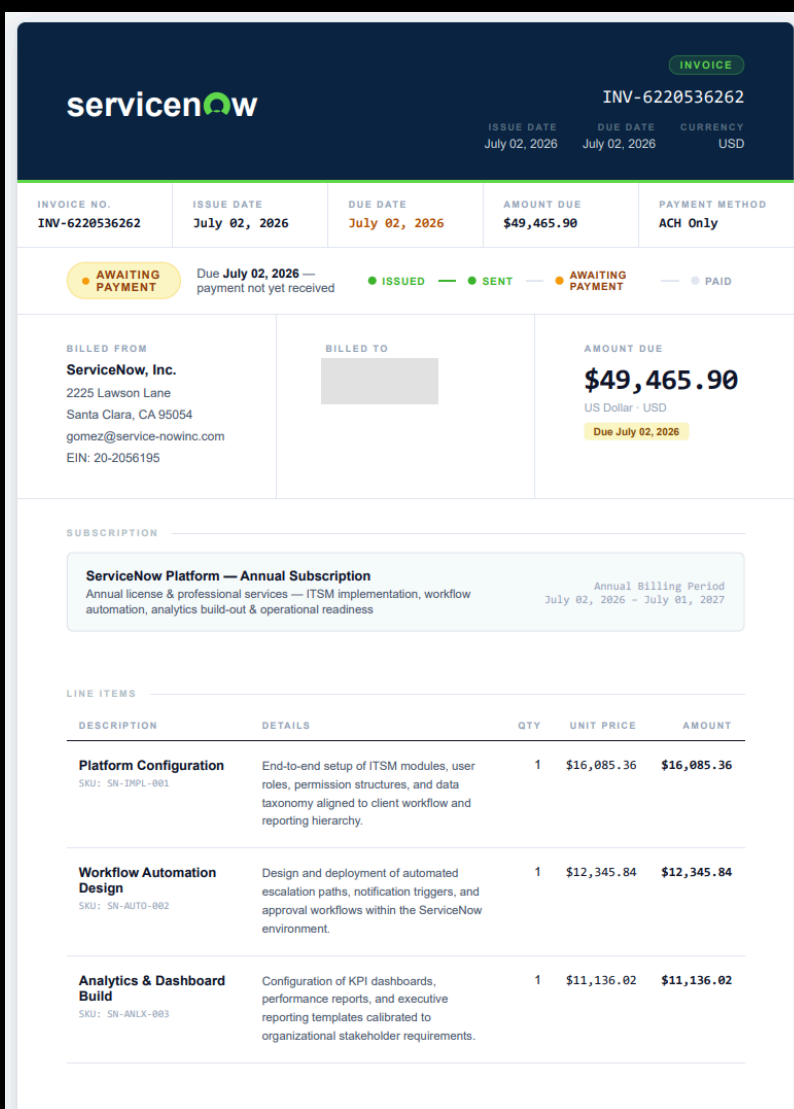
This is another AI Dragnet campaign: a fabricated ServiceNow invoice scam that reached over half of AegisAI’s protected fleet within days of its first wave, then reappeared for seven weeks.

Every version scored as AI-generated under AegisAI’s own classifier (23 to 29 depending on variant). Every message asked for payment by reply, with no malicious link, so there was nothing for a crawler to follow. A fabricated invoice PDF with real ServiceNow branding gave the ask a paper trail, but the con did not depend on it. The sender’s domain and a fabricated remittance contact, “Gomez,” did the work of legitimacy. The domain authenticated cleanly through a two-decade-old mail domain with no history of abuse, and the mail went out through Amazon SES.

This is the trust-hijacking principle from the Technical Deep Dive one layer down: an aged domain's reputation in place of a hijacked inbox. One target was re-approached four times over the seven weeks, each time with new subject-line phrasing, the clearest sign in this dataset of deliberate re-engagement.

The operator changed tactics between waves. Between an early wave and a second one five weeks later, the operator retired the original lookalike domain for two new ones, the fabricated invoice number began varying per target instead of repeating verbatim and at least one recipient received a different fabricated vendor identity in place of ServiceNow's. Only the "Gomez" persona carried over, the one constant across three rounds of otherwise-disposable infrastructure.

Every version places the campaign in the AI Dragnet category, not AI Spearphishing, even though the repeated approach to one recipient might suggest targeting. The classification keys on fleet-wide blast radius, not follow-up.



INVOICE NO.	ISSUE DATE	DUE DATE	AMOUNT DUE	PAYMENT METHOD
INV-6220536262	July 02, 2026	July 02, 2026	\$49,465.90	ACH Only

AWAITING PAYMENT Due July 02, 2026 — payment not yet received

● ISSUED — ● SENT — ● AWAITING PAYMENT — ● PAID

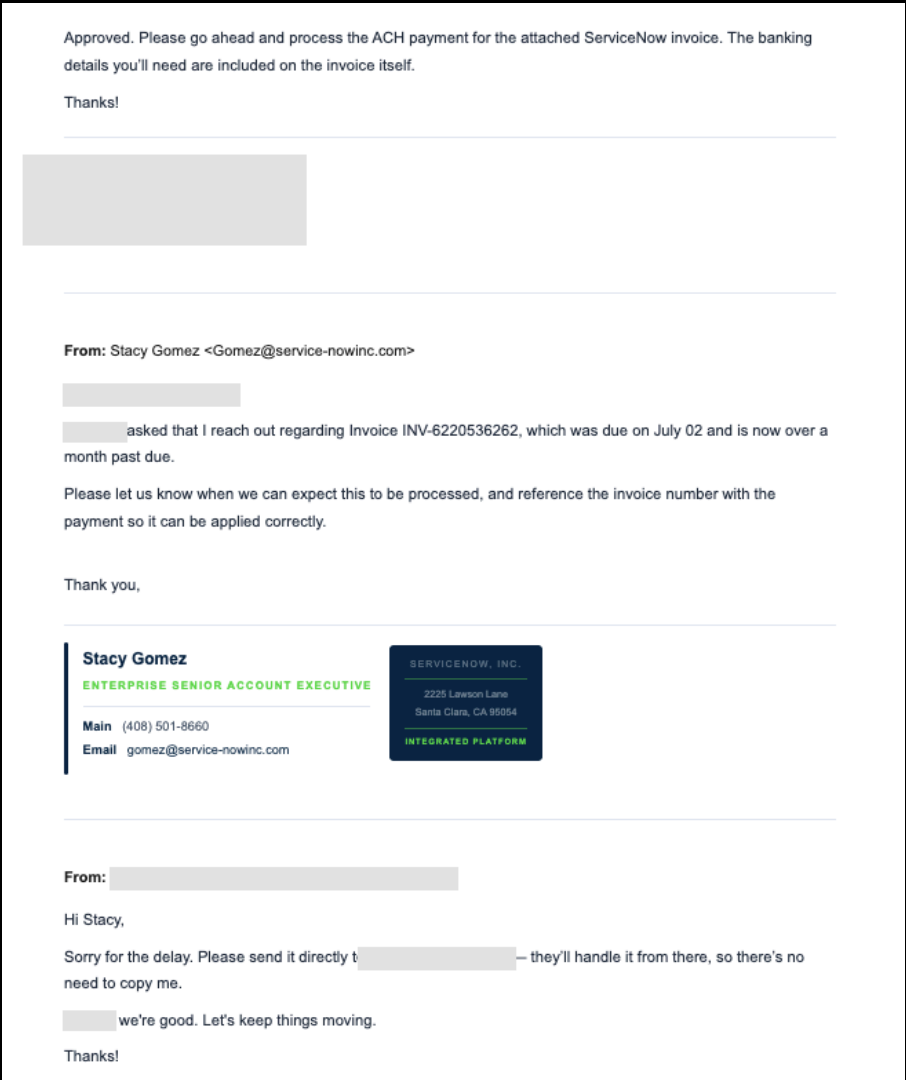
BILLED FROM	BILLED TO	AMOUNT DUE
ServiceNow, Inc. 2225 Lawson Lane Santa Clara, CA 95054 gomez@service-nowinc.com EIN: 20-2056195	[REDACTED]	\$49,465.90 US Dollar · USD Due July 02, 2026

SUBSCRIPTION

ServiceNow Platform — Annual Subscription
Annual license & professional services — ITSM implementation, workflow automation, analytics build-out & operational readiness
Annual Billing Period: July 02, 2026 - July 01, 2027

DESCRIPTION	DETAILS	QTY	UNIT PRICE	AMOUNT
Platform Configuration SKU: SN-IMPL-001	End-to-end setup of ITSM modules, user roles, permission structures, and data taxonomy aligned to client workflow and reporting hierarchy.	1	\$16,085.36	\$16,085.36
Workflow Automation Design SKU: SN-AUTO-002	Design and deployment of automated escalation paths, notification triggers, and approval workflows within the ServiceNow environment.	1	\$12,345.84	\$12,345.84
Analytics & Dashboard Build SKU: SN-ANLX-003	Configuration of KPI dashboards, performance reports, and executive reporting templates calibrated to organizational stakeholder requirements.	1	\$11,136.02	\$11,136.02

The fabricated invoice PDF, attached to the August-wave lure. It has real ServiceNow branding, a real ServiceNow corporate address and three professionally worded line items totaling the exact amount named in the email. Recipient details redacted.



The follow-up chase, forwarded internally by the recipient after approving payment. The fabricated persona has a full name, title and phone number: "Stacy Gomez, Enterprise Senior Account Executive." The recipient's reply references a second fabricated ServiceNow persona in the same thread, which suggests a multi-persona escalation chain, not a single sender. Recipient and individual names redacted throughout.

SO WHAT? This is the clearest evidence in this report that AegisAI's telemetry undercounts a campaign's reach. One recipient's internal security team runs an abuse inbox that AegisAI protects, and it surfaced this campaign forwarded from organizations with no relationship to this fleet, including a large, publicly recognizable consumer brand. One downstream vendor's housekeeping exposed a reach that this report's numbers cannot otherwise see.

MITRE ATT&CK

T1656 Impersonation T1585.002 Establish Accounts: Email Accounts T1583.001 Acquire Infrastructure: Domains

INDICATORS

INDICATOR	TYPE
service-nowinc[.]com	Lookalike sender domain, first wave

<real-executive-name>@service-nowinc[.]com	Fabricated sender persona, impersonating a real ServiceNow executive by name (name withheld here)
gomez@service-nowinc[.]com	Fabricated remittance contact, first wave: full fabricated identity “Stacy Gomez, Enterprise Senior Account Executive,” with a fabricated direct phone number and branded signature block
INV-6220536262	Fabricated invoice number, reused verbatim, first wave
rbtrees[.]com	DKIM-signing domain, registered 2004, reputation 100/100
servicenow.com[.]ai	Lookalike sender domain, second wave, registered via Marcaria (2023-08-14)
servicenow.me[.]uk	Lookalike sender domain, second wave
Gomez@servicenow.com[.]ai	Fabricated remittance contact, second wave
servicenow-uk@outlook[.]com	Fabricated remittance contact, free-mail variant
ARteam-execs@post[.]com	Fabricated remittance contact, alternate-vendor variant
INV-80044710620	Fabricated invoice number, second wave
linksderisar[.]com	DKIM-signing domain, registered 2012, reputation 92/100
mcmsj2htlf6vf2iedh7uch5fsdhimt2n	DKIM selector observed on linksderisar[.]com signatures
airlst-mail.linksderisar[.]com	SPF mail-from subdomain, second wave
54.240.98.237	Amazon SES sending IP, authorized by SPF for airstl-mail.linksderisar[.]com
emailfly[.]app	Co-occurring sender-infrastructure domain, registered 2026-08-09, reputation 40/100
visiontechsolutions[.]com	Co-occurring sender-infrastructure domain, registered 2014-03-24, reputation 96/100

INTRUSION SET · AEGIS-IS-FILLTHEBILL-004 · AI DRAGNET · TREND 04

A billing-portal phish that already knows the victim’s own email address

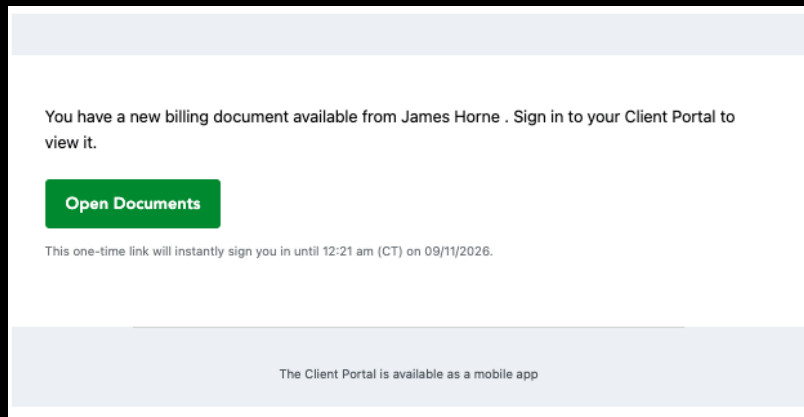
AT A GLANCE Reach: 18% of tracked organizations in three minutes · Delivery: single automated send, 2026-09-11 · Goal: credential theft

This is another AI Dragnet template: a byte-identical “new billing document” notification, sent from a fabricated Client Portal. It reached 18% of AegisAI’s tracked tenant base inside a single three-minute window on 2026-09-11, one new organization roughly every 12 seconds. That timing is consistent with a single automated send, not independently timed attempts.

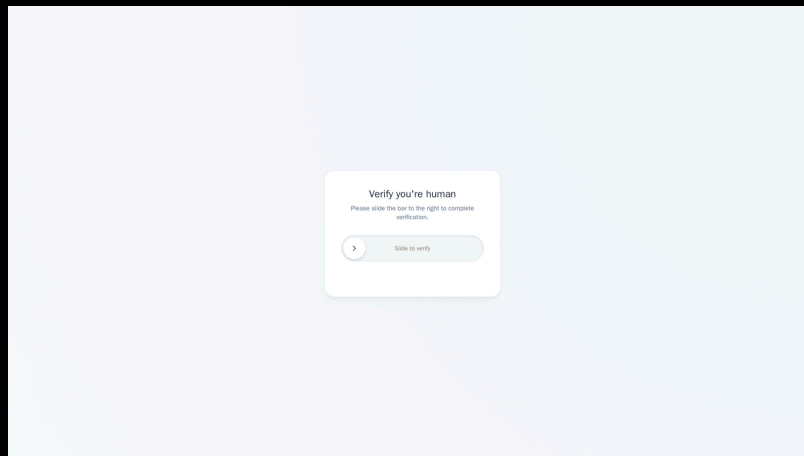
The template scored 15.0 under AegisAI’s production classifier. The message body carries zero personalization: same subject, same wording, same layout, sent from a domain registered years before the campaign. This blast-then-done fan-out places it in the AI Dragnet category, not the more targeted AI Spearphishing pattern.

The landing link is what varies per target. It pre-fills each recipient's real email address as a URL parameter, so the credential-harvesting page already shows the victim's address waiting for a password. The personalization is at the link level, though the body carries none. The link routes through MailChannels' legitimate transactional-email click-tracking service before landing on a domain registered roughly three months before the campaign.

Every AegisAI crawl of it succeeded and captured real evidence, not a decoy or a block: a fake human-verification gate, a slider styled to imitate a CAPTCHA, in front of packed and obfuscated JavaScript. The credential form sits behind the gate. The gate carries a recognized, reusable anti-bot kit signature and was not built for this campaign.



The lure as rendered for the recipient. Generic sender name, no recipient-specific detail in the body. The personalization lives in the link underneath "Open Documents," not in anything visible here.



The landing page as AegisAI's crawler captured it: a fake "Verify you're human" slider gate, not the credential form. The gate filters bots from real targets before the harvesting page renders.

SO WHAT? Pre-filling the recipient's email address removes the one step, typing an address, that might give a distracted target a moment to notice something is wrong. Only a password field stands between the click and a captured credential. Because the personalization lives in the URL and not the message body, bulk content-similarity detection still sees one template, while every link a target receives is unique to them.

MITRE ATT&CK

T1566.002 Phishing: Spearphishing Link T1204.001 User Execution: Malicious Link T1583.001 Acquire Infrastructure: Domains

INDICATORS

INDICATOR	TYPE
pdfx.qitbaq[.]icu	Landing domain, registered 2026-06-23, NameSilo, every crawl returned HTTP 200
plurastack[.]com	Sending domain, registered 2021-07-30, Name.com, Inc.
click.mailchannels[.]net	Real MailChannels click-tracking service, abused for link legitimacy
"Reminder: Complete your documents"	Subject line, byte-identical across every send
"smn=" URL parameter	Recipient's own email address embedded per-link, not in the message body
HUMAN_VERIFY_GATE_CFBEBACON_NOT_MANTH	Fake human-verification gate spoofing a device-check beacon, confirmed via crawl screenshot
OBFUSCATED_CHARCODE_PACKER	JavaScript packing/obfuscation signal on the landing page, same commodity kit signature

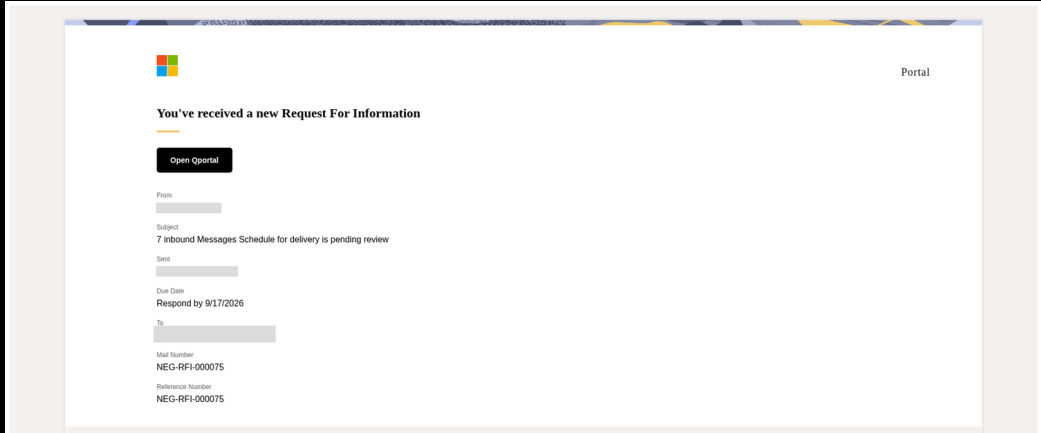
INTRUSION SET · AEGIS-IS-CAPTCHATHEBILL-005 · HUMAN-WRITTEN · TREND 06

A fake portal notification that hides its real payload behind two different disguises

AT A GLANCE Reach: nearly a quarter of protected organizations in 88 minutes · Delivery: compromised and purpose-registered redirect domains · Evasion: scanner-specific decoys

This campaign does not score as AI-generated, and it is the most technically sophisticated in this section, an example of the convergence between attack types described under Evasion Mechanics. A fake document-portal notification, with an identical embedded logo and a deliberately rotating subject line, reached nearly a quarter of AegisAI's protected fleet inside an 88-minute window, close to one new organization every four minutes. Randomizing only the subject while keeping the body untouched is a deliberate choice. It defeats any filter that keys on subject text, a common and cheap first pass, and leaves the same-origin signal intact for anything that fingerprints the message body.

Both first-hop domains redirect a second time to a shared final landing page, visible only by tracing AegisAI's crawl records. One is an aged domain showing signs of compromise; the other was purpose-registered about a month before the campaign. The final page uses two different tricks depending on who arrives. One first-hop domain gates behind a "prove you're human" challenge that blocks automated analysis outright, while the other silently redirects a non-victim to a real, unrelated software company's marketplace page, a decoy so ordinary it does not look like an evasion technique.



The portal notification as the recipient sees it. From, Sent and To fields redacted.

SO WHAT? Both tricks exploit what a security tool does differently from a human. A crawler resolves every link it is given and expects a payload at the end. A wall only a bot would hit, and a redirect to a page a bot has no reason to flag, both read as “nothing to see here” to a detection engine, while a targeted human still lands on the real page every time.

On the AI-generated question: the raw message initially scored as AI-generated, but that signal traced entirely to an unrelated real email thread pasted into the same body, not to the attacker’s text. Scored on the lure template alone, this campaign scores as not AI-generated: a sophisticated technique that needed no AI to build.

MITRE ATT&CK

T1566.002 Phishing: Spearphishing Link T1204.001 User Execution: Malicious Link T1656 Impersonation T1583.001 Acquire Infrastructure: Domains T1584.004 Compromise Infrastructure: Server

INDICATORS

INDICATOR	TYPE
kompactsystems[.]com	First-hop domain, registered 2005-12-27, Tucows, Cloudflare-fronted, likely compromised
caseforge-law[.]com	First-hop domain, registered 2026-08-17 (~1 month pre-campaign), Cloudflare-fronted, purpose-registered
sandrnansavrann[.]com	Shared final landing domain, found via crawl trace only, gated behind Cloudflare Turnstile
/wp-includes/sitemaps/...	URL path pattern, compactsystems[.]com
/wp-includes/theme-compat/...	URL path pattern, caseforge-law[.]com
"PENDING EXAMINATION" / "URGENT: Portal" / "HOLD PENDING" / "VERIFICATION ALERT"	Polymorphic subject-line variants (5+ observed) on identical body content
EMPTY_CAPTCHA_GATE	Crawl signal marking the Cloudflare Turnstile block at the final hop

Catching Faster Attacks: Speed, Investigation and Proteus

This quarter's campaigns moved faster than any human-led response. The window between first delivery and damage now runs in minutes, and the attacks that matter most look legitimate on every check a filter runs before delivery. Two things decide the outcome: how fast a defense reaches a verdict, and whether it investigates what a message claims or just checks who sent it. AegisAI built **Proteus**, its own foundational model trained on email, to do both. This section explains why that changes what defenders can catch.

Attacks now move in minutes, not days



Time for each tracked campaign to reach the stated scale, on a log scale. FAKESERVICE reached over half of protected organizations within days of its first wave.

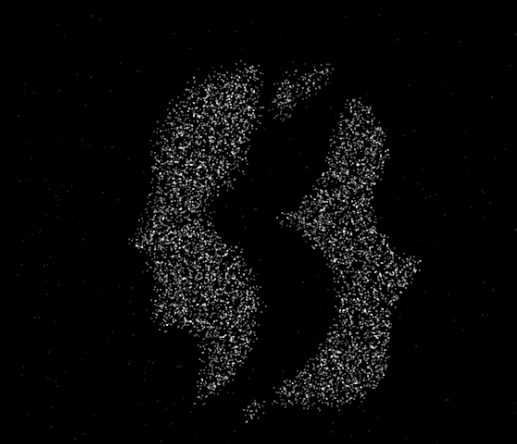
FILLTHEBILL reached a new organization every 12 seconds. REDTIDE turned a victim into its next sender in under 19 minutes. CAPTCHATHEBILL reached nearly a quarter of protected organizations in 88 minutes. At that speed, anything that acts after delivery (a user report, a SOC ticket, a retroactive sweep) starts behind the attack. Calendar attacks widen the gap: the invite lands on the calendar at delivery, with no click, so post-delivery cleanup can shorten exposure but can't prevent it.

Checking reputation can't keep up

Trust hijacking means the sender is real, authenticates cleanly and may have a history with the recipient. AI means every lure is new, so there's no signature to match. A filter that asks "is this message unusual?" sees nothing wrong. The question that still works is whether the evidence supports what the message claims about itself: where its link actually resolves, whether the two parties have really corresponded, and what the

headers and sending infrastructure show. Answering it takes an investigation, one in which each finding decides the next check, run on every message before it reaches a person.

Proteus: AegisAI's own foundational model for email



PROTEUS · AEGISAI LAB

Proteus is the first model from the AegisAI Lab. AegisAI trained it from the ground up on email; it is not a general-purpose language model with a security prompt on top. That choice matters for this quarter's attacks. General-purpose models are trained to cooperate with the text they receive, which is exactly what a convincing AI-written lure exploits, and anyone can probe a public model offline to tune a lure against it. Proteus instead reads each message the way an investigator would: it takes the body, headers, sender and thread history, link destinations, attachments and threat intelligence as one structured input, checks every claim the message makes against evidence the sender can't control, and lets each finding decide the next check. It does this inline, on every message, before delivery, and returns a verdict in a single forward pass.

ATTACKS IDENTIFIED · 3,020 UNSEEN ATTACKS

Proteus

>99%

Leading closed-source model

81.3%

Widely used open-weight model

80.0%

General-purpose models let through roughly 30x as many attacks.

VERDICT SPEED VS CLOSED-SOURCE MODEL

7.2x

faster median
verdict

26x

faster in the slowest
1 in 20 verdicts

Single forward pass, inline, before delivery.

On a benchmark of 3,020 human-reviewed phishing attacks from one week of production traffic, none seen in training, Proteus identified more than 99%. The leading closed-source general model identified 81.3% and a widely used open-weight model 80.0%, letting through roughly 30 times as many attacks. Proteus returned its median verdict 7.2x faster than the closed-source model, and 26x faster in the slowest 1 in 20

verdicts, the cases where a slow answer lets an attack through. In production, it runs at 99.996% precision, which typically means about one false positive every two years for most companies.

Benchmark limitations: the general-purpose baselines received a plain classification instruction and were not optimized; the sample is one week of AegisAI customer traffic; the benchmark measures missed attacks, not false alarms; and detection performance decays as attackers adapt, so AegisAI retrains continually and tests each version against nine attack families before release.

Why Proteus catches what reputation misses in these campaigns

Every campaign in this report was built to pass a reputation check. Each one still left evidence that contradicts what it claims to be, and that evidence is what Proteus investigates. Because the investigation happens before delivery, the verdict arrives before a calendar invite lands, before a second victim starts sending, and before the next organization in a three-minute burst is hit.

CAMPAIGN · SPEED	WHAT A REPUTATION CHECK SEES	EVIDENCE PROTEUS INVESTIGATES
REDTIDE <19 min victim to next sender	A real Google Workspace account. DKIM passes. Clean sending history. The invite is already on the calendar.	The same bid notice and fabricated reference number reused verbatim across unrelated senders. A link to an accounts.* lookalike domain behind a fake human check.
FAKESERVICE days to reach half of orgs	A two-decade-old domain with a 100/100 reputation score. No link or payload to scan.	A ServiceNow invoice sent from a domain ServiceNow doesn't own. Payment requested by reply to a fabricated remittance contact.
FILLTHEBILL 3 min to reach 18% of orgs	A sending domain registered years earlier. A link routed through a legitimate click-tracking service.	A byte-identical message to many organizations. A link pre-filled with the recipient's own address, landing on a domain registered three months earlier.
CAPTCHATHEBILL 88 min to reach ~25% of orgs	A human-verification wall, or a harmless software marketplace page. Nothing to flag.	Rotating subject lines on an identical body. Two redirect hops to a shared final page, one through a domain registered a month before the campaign.

Illustrative: the evidence present in each campaign, as documented in section 05. Not a trace of individual verdicts.

What this means for the findings in this report

The evasion figures in this report count attacks that got past the recipient's own email provider and that AegisAI still caught. Those are, by definition, the attacks a reputation check passes and an investigation catches. That is why the trends point where they do: as attackers move to hijacked accounts, aged domains and calendar delivery, more of what reaches people looks legitimate to every check except one that tests the message's claims. Proteus is how AegisAI sees these attacks at all, and why the share it catches after provider filtering (roughly three in ten confirmed phishing messages of every attack type) is a measure of what reputation-based defenses now miss.

Read more: [Introducing Proteus, the AegisAI foundational model.](#)

Implications and Recommendations

Attackers now spend less effort to look more legitimate. AI writes the lure, a hijacked account or aged domain delivers it, and a trusted platform hosts the payload. Three shifts follow for every organization.

Sender and domain reputation can no longer be a primary signal. Three of this quarter's four tracked intrusion sets reached inboxes without a suspicious-looking sender. AEGIS-IS-REDTIDE-002 sent from real, compromised Google Workspace mailboxes with clean histories. AEGIS-IS-FAKESERVICE-003 used aged, previously legitimate mail domains with high reputation scores. AEGIS-IS-FILLTHEBILL-004 routed its links through a legitimate email vendor's click-tracking service. As attackers keep using compromised accounts, legitimate SaaS services and aged domains to deliver attacks, a good reputation stops being evidence that a message is safe.

The impersonated brand changes; the playbook doesn't. Google Workspace is the most impersonated brand this quarter; retail and e-commerce brands led in earlier quarters, though classification changes mean the two aren't directly comparable. Distrusting one brand's name is a habit that expires. The durable habit is to treat any unexpected, urgent request to sign in, pay or share the same way, whichever company it claims to be from.

The cheapest AI phishing is catching up to the most sophisticated. AI Dragnet went from a rounding error (0.5% of confirmed phishing) to 5.8% in one quarter, and its sender-compromise rate rose from 0.2% to 7.7% of its own volume. If that trajectory continues, the assumption that mass-produced AI content stays crude and easy to catch will keep failing.

AI Spearphishing's growth and AI Dragnet's shift to compromised accounts moved in the same direction this quarter. The assumptions that made phishing easy to triage (clumsy language, disposable infrastructure, obvious targeting) are eroding at the same time.

What security teams should do now

Decide before delivery, at attack speed. Campaigns now spread in minutes, so a verdict after delivery comes too late. Investigate every message inline, checking what it claims against evidence the sender can't control, before it reaches an inbox or a calendar.

Verify requests out of band. Require a call-back or second channel for payment changes, new vendor banking details and credential prompts, including requests inside existing email threads.

Watch accounts, not just messages. Detect compromised-account behavior: new sign-in locations, new mailbox rules, sudden outbound bursts and calendar invites sent to many outside recipients.

Lock down calendar invites, and clean them up. Limit automatic calendar additions to known senders, but don't rely on it: compromised accounts count as known. When a phishing email is quarantined, remove the calendar event it created too, and sweep calendars for malicious invites that already got through.

Don't treat reputation as a verdict. Treat aged domains, trusted hosting platforms and legitimate click-tracking links as neutral, not safe.

Inspect links the way a person would. Link analysis needs to follow every redirect and get past human-verification gates, because attackers now show automated scanners a harmless page.

Update awareness training. Replace "look for typos and generic greetings" with "pause on any request to sign in, pay or share, however polished it looks."

08 · METHODOLOGY

Methodology

Data sources

This report analyzes more than 100 million emails sent to over 1 million mailboxes protected by AegisAI. Statistics cover Q3 2026 to date, July 1 to September 2 (63 days), compared with Q2 2026 (91 days) or Q1 2026 as stated on each chart. Intrusion-set case studies draw on activity tracked through mid-September, so some indicator dates fall after the statistical window.

Confirmed phishing

The AegisAI threat research team confirmed every message this report counts as phishing. We express volumes as a rate, confirmed phishing as a share of all email scanned in the period, so changes reflect attack activity rather than changes in overall mail volume or in the number of protected mailboxes.

Identifying AI-generated phishing

AegisAI's AI-generated content classifier scores each confirmed message. The classifier weighs stylistic signals common in language-model output, such as absence of contractions and low lexical diversity, alongside hard artifacts such as unfilled template placeholders, false-context claims ("as we discussed") and leaked tool names. We label messages that score above the classifier's threshold as AI-generated. The signal tables in section 04 show which signals drove verdicts this quarter. Where a message mixes attacker text with pasted legitimate content, analysts score the attacker's template on its own (see AEGIS-IS-CAPTCHATHEBILL-005).

Attack types

AI-generated messages are split into AI Spearphishing and AI Dragnet by fan-out: how widely the same message or template is sent across organizations. Low fan-out is AI Spearphishing; high fan-out is AI Dragnet. Messages that do not score as AI-generated form the human-written category.

Comparisons

We report changes in rates in percentage points (pt) and changes in volume and share as relative percentages. Because Q3 to date is shorter than Q2, all comparisons use rates, not raw counts. Where an

attack type was small in the prior quarter, the report states the upper bound of that quarter's confidence interval and notes when the new rate falls outside it. We report domain-diversity and reuse figures for this quarter only, because a shorter window mechanically inflates apparent diversity.

Limitations

This data reflects AegisAI's protected organizations, not all email, and the mix of customers, industries and email platforms shapes what we see. A single large campaign can move category-level figures; section 05 profiles the largest. Evasion rates count only phishing that AegisAI detected, so they describe what got past provider filtering, not what got past every control. Brand and attack-type classification changed between Q2 and Q3, so those figures are not compared quarter over quarter.