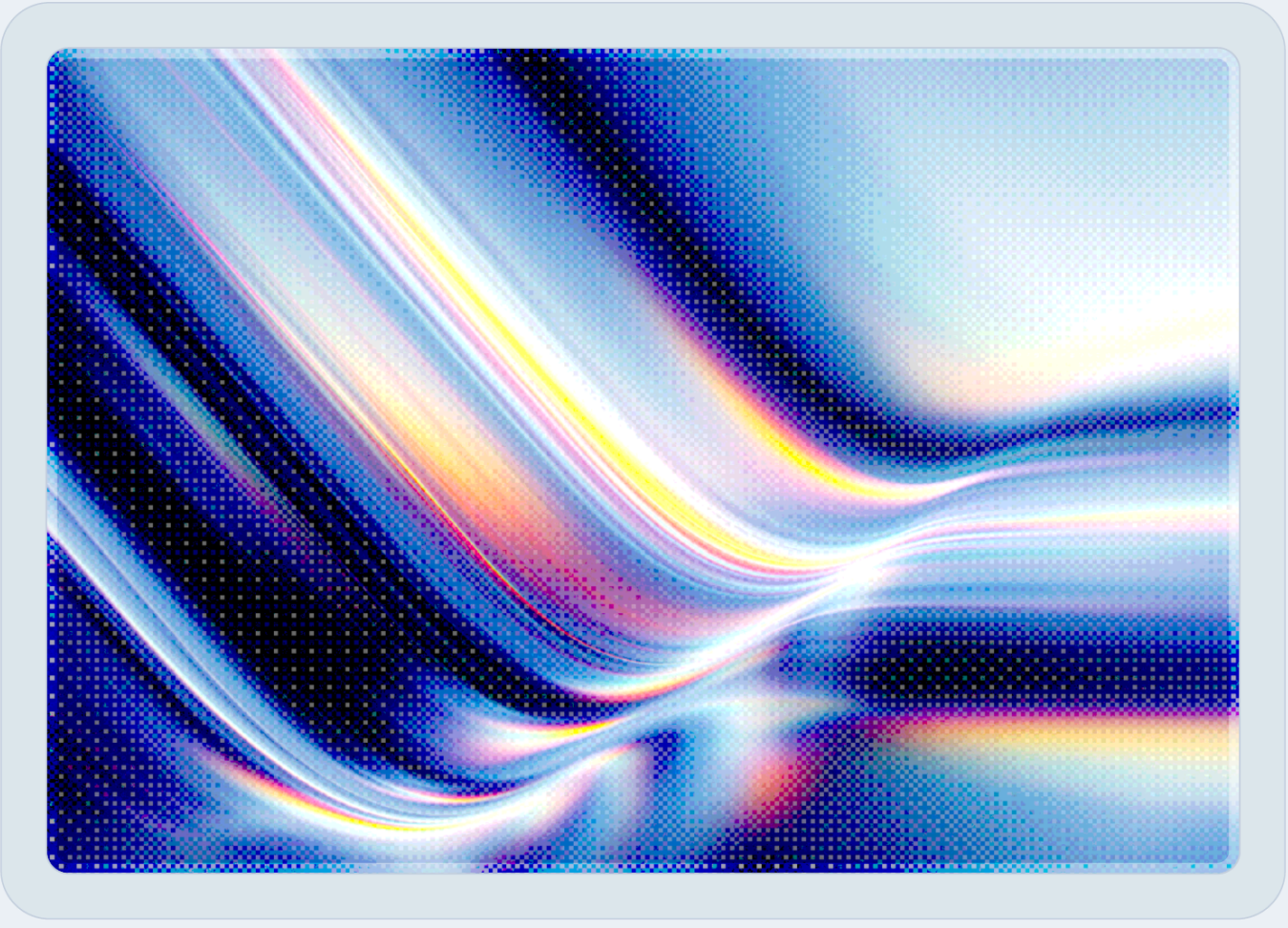
CSAI Foundation | Cloud Security Alliance

AI-Generated Lures Behind Microsoft Cloud Account Takeovers

Passkey-Themed Social Engineering and Generative AI Converge Against Microsoft 365

2026-09-14

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

Microsoft disclosed two related but operationally distinct intrusion patterns against its cloud customers in September 2026, and together they illustrate a broader pattern in which identity-focused social engineering and, in at least one of the two campaigns, generative AI-assisted content generation are converging inside the Microsoft 365 threat landscape. Security teams should read the two disclosures as connected data points rather than isolated incidents, because both point toward the same underlying shift: attackers are spending less effort on technical exploitation and more on personalized, persuasion-based tactics that get a human to hand over access voluntarily.

- Threat actors linked to the ShinyHunters and Helix extortion brands have run a passkey-themed vishing and smishing campaign against Microsoft cloud tenants since at least May 2026, using fake IT helpdesk calls to trigger adversary-in-the-middle and device-code phishing [1][2].
- A separate but concurrently disclosed campaign sent more than one million CEO-impersonation invoice fraud emails between August 3 and 5, 2026, with HTML artifacts consistent with generative AI-assisted drafting [3].
- Compromised Microsoft 365 accounts are being used to register attacker-controlled MFA methods, enabling persistent, user-independent access well after the initial phishing event [1].
- Post-compromise activity relies heavily on the Microsoft Graph API for reconnaissance and on high-volume SharePoint, OneDrive, and Exchange downloads for exfiltration [1][4].
- CSA's own July 2026 Forg365 analysis documented a related case of fake Entra passkey enrollment paired with vishing inside a phishing-as-a-service kit [6] — a single reported instance rather than an established trend, though it is consistent with the broader pattern of AI-generated phishing outperforming manually authored campaigns that CSA has separately tracked [5].

Background

On September 9, 2026, Microsoft Security Research published details of an active cloud intrusion pattern in which threat actors impersonate corporate IT help desks to compromise Microsoft 365 accounts [1]. The attackers call or text a target's personal phone number, claim that an urgent passkey,

multi-factor authentication, or single sign-on update is required to avoid a service disruption, and direct the victim to a look-alike Microsoft sign-in page. Despite the passkey framing, Microsoft notes the actors are not actually trying to enroll a passkey; the pretext exists to create urgency and route the victim into an adversary-in-the-middle proxy or an OAuth device-code flow that captures a valid, MFA-satisfied session. Microsoft has observed this activity since May 2026 and attributes the initial access work to two clusters, Storm-3121 and Storm-3032. Storm-3121 activity leads into the ShinyHunters and Falcon extortion operations, while Storm-3032 is described as a group that split from the BlackFile collective and now operates under the Helix extortion brand; Google's Mandiant tracks overlapping activity as UNC6671 [4].

A day later, on September 10, Microsoft disclosed a second, separately tracked campaign that it explicitly connects to generative AI [3]. Between August 3 and 5, 2026, attackers sent more than one million scam emails to corporate finance and accounts-payable staff, impersonating each target company's CEO, CFO, or president. The emails referenced a fabricated ServiceNow invoice and included a forged forwarded-message thread designed to look like the target's own executives had already discussed the purchase with ServiceNow's leadership, pressuring the recipient into authorizing an ACH transfer of roughly $50,000. Microsoft states that the campaign's HTML showed indicators consistent with generative AI-assisted template development — including excessive code comments, uniform section labeling, em-dashes, and delimiter banners that Microsoft itself describes as characteristic of large language model output — and that operators used these tools to produce recipient-tailored drafts at scale [3]. Microsoft has not confirmed that ServiceNow itself was compromised; the invoices and executive names were fabricated using publicly available corporate information.

Read together, coverage of the two campaigns from The Hacker News, BleepingComputer, and RH-ISAC points to a broader Microsoft cloud threat moment unfolding in the same week, with both campaigns targeting the identity and finance functions of U.S. enterprises in IT services, consumer goods, real estate, and discrete manufacturing [4][7]. Neither Microsoft's passkey advisory nor independent reporting has confirmed generative AI use in the SMS and voice lures of the passkey campaign specifically; the AI-assisted drafting evidence Microsoft published applies to the invoice fraud campaign. The distinction matters for defenders, because the two campaigns require different detection signals even though they share a target base and were disclosed in the same week.

Security Analysis

The Passkey Pretext and the Real Objective

The passkey-themed intrusion pattern is notable less for technical novelty and more for how effectively it repurposes a well-worn helpdesk-impersonation technique by attaching it to a trust signal — passkeys — that most employees now associate with stronger, phishing-resistant security. Microsoft's telemetry shows attackers researching a target organization and its staff before initiating contact, then using a phone call or SMS message to a personal device to bypass corporate email filtering entirely. The phishing infrastructure is disposable: domains such as passkeyhelpdesk[.]com, secure-passkey[.]com, and integratedsso[.]com, several registered through the Nicenic registrar, are stood up and pointed at victims within hours, following a pattern of prefixing the target company's name to a generic-sounding SSO or passkey domain [1][4]. Once a victim authenticates through the adversary-in-the-middle proxy or completes a device-code flow the attacker initiated, the session token is captured directly, and the attacker's next move is to register a new MFA method — a phone number, an authenticator app, or a software one-time-password token — under their own control. That step converts a single successful phishing attempt into durable, self-service access that survives a password reset and does not require the victim's continued participation [1].

From there, the attackers pivot to systematic reconnaissance using the Microsoft Graph API, enumerating users and groups, querying directory roles and privileged role assignments, and mapping SharePoint sites and OneDrive drives before mailbox contents are inspected. Microsoft observed a python-httpx user agent associated with the high-volume SharePoint and OneDrive access that follows, consistent with scripted, bulk-download behavior rather than manual browsing, and exfiltration windows that stretch from several hours to several days per compromised tenant [1][4]. This progression — quiet initial access, silent MFA persistence, broad API-driven reconnaissance, then bulk data collection — is the signature Microsoft asks defenders to correlate across events rather than evaluate as isolated alerts, since any single Graph call or sign-in in the chain can look unremarkable on its own.

Where Generative AI Enters the Picture

The clearest, Microsoft-confirmed use of generative AI in this news cycle sits in the invoice fraud campaign, not the passkey campaign. There, AI tooling appears to have been used to draft convincing, recipient-specific executive impersonation emails and to fabricate a supporting forwarded-thread narrative, at a volume — over one million messages in a three-day window — that would be difficult to sustain with manual copywriting [3]. This matters because it corroborates a pattern CSA's own AI Safety

Initiative research has already documented: generative AI systematically removes the writing-quality and localization barriers that used to make business email compromise detectable, and it does so at a scale that shifts email fraud from a boutique, hand-crafted operation into a commodity one [5].

The more directly relevant precedent from CSA's threat research is Forg365, a phishing-as-a-service platform analyzed by the CSA AI Safety Initiative in July 2026 that bundles AI-assisted, multilingual lure drafting with the same adversary-in-the-middle and OAuth device-code techniques Microsoft describes in the September passkey campaign [6]. That analysis specifically flagged a related extortion-affiliated actor combining vishing with fake Entra passkey enrollment to achieve real account takeover — the same operational pattern now showing up in Microsoft's Storm-3121 and Storm-3032 telemetry, months later and at larger scale, though Forg365 itself documented that pairing as a single reported case rather than an established trend. Read together, the two disclosures and CSA's prior research suggest that AI-assisted lure generation and passkey-themed social engineering may be converging into a repeatable criminal playbook that different actors are now running against the same Microsoft 365 attack surface, though Microsoft has not confirmed that the September passkey campaign's voice and SMS lures themselves used generative AI drafting, and CSA is not asserting that they did. The point of convergence is the surrounding ecosystem — the phishing-as-a-service tooling and the shared target base — rather than a single confirmed technical fact.

The table below summarizes the two campaigns side by side.

Dimension	**Passkey-themed social engineering**	**AI-assisted invoice fraud**
Primary vector	Phone call / SMS to personal device	Email to accounts payable staff
Pretext	Urgent passkey/MFA/SSO update from "IT helpdesk"	Executive-approved ServiceNow invoice
Confirmed AI role	Not confirmed by Microsoft	Confirmed: AI-assisted email template generation
Credential technique	AiTM proxy / OAuth device-code flow	None (payment fraud, not account takeover)
Persistence mechanism	Attacker-registered MFA method	N/A (single-transaction fraud)

Dimension	Passkey-themed social engineering	AI-assisted invoice fraud
Observed volume/timeline	Active since May 2026, multiple tenants	1M+ emails, August 3–5, 2026
Attribution	Storm-3121, Storm-3032 (ShinyHunters, Helix)	Not publicly attributed
Objective	Data exfiltration, extortion	Direct ACH funds transfer (~$50,000/target)

Recommendations

Immediate Actions

Organizations should treat any unsolicited call, text, or Teams message that asks an employee to update a passkey, MFA method, or SSO configuration as a probable social engineering attempt, and should route verification of such requests through a separate, pre-established channel rather than the number or link the requester provides. Security teams should also search Entra ID sign-in and audit logs for the pattern Microsoft describes: an unusual sign-in followed shortly by the enrollment of a new MFA method, particularly a newly added phone number or authenticator app on an account that has not recently changed its MFA configuration. Finance and accounts-payable teams should be briefed specifically on the ServiceNow-themed invoice lure and instructed to verify any executive-initiated ACH request through a known, independently obtained phone number before releasing funds, regardless of how internally consistent the supporting email thread appears.

Short-Term Mitigations

Enforcing phishing-resistant, FIDO2-based MFA through Conditional Access removes the credential-harvesting payoff that makes adversary-in-the-middle proxies effective, and disabling the OAuth device-code authorization flow tenant-wide — unless a specific, documented business need requires it — closes the second access path both this campaign and the earlier Forg365 kit rely on [1][6]. Conditional Access policies that require a managed, compliant device for access to Exchange, SharePoint, and OneDrive limit how far a stolen session token can be used even if initial credential theft succeeds. Security teams should also enable and actively review Graph API activity logs and mailbox audit logs,

correlating directory enumeration, role queries, and bulk download events across a single identity rather than alerting on them individually, since Microsoft's telemetry shows the value of this attack chain is only visible in aggregate. On the email side, tightening DMARC enforcement to reject or quarantine, and evaluating whether existing email security tooling can flag generative AI stylistic artifacts such as repetitive formatting and delimiter banners, will help against the invoice fraud vector specifically.

Strategic Considerations

The pairing of AI-assisted content generation with identity-focused social engineering is likely to keep expanding rather than remain confined to these two campaigns, and CSA's prior research on AI-weaponized phishing found that AI-generated lures have already surpassed expert human red-teamers in measured effectiveness [5]. CSA's broader research on AI-enabled influence operations reaches a similar conclusion from a different angle, finding that frontier AI systems reliably outperform expert humans at persuasion and recommending that organizations shift security-awareness training away from feature-based detection and toward process-focused verification [9]. In CSA's assessment, security awareness programs built around spotting grammatical errors or generic phrasing are losing effectiveness against this threat, and organizations should shift training emphasis toward behavioral verification — confirming unusual requests through independent channels — rather than content inspection. Longer term, enterprises should treat the device-code OAuth flow and SMS-based MFA as legacy protocols to be phased out in favor of phishing-resistant authentication, and should build incident response playbooks that assume a compromised identity will be used for silent MFA persistence and API-driven reconnaissance before any data movement is detectable through conventional means.

CSA Resource Alignment

This incident connects most directly to three pieces of CSA AI Safety Initiative research published earlier in 2026. Forg365: AI Lure Generation in an M365 Phishing Kit examined a commercial phishing-as-a-service platform that bundles AI-assisted, multilingual lure drafting with the same adversary-in-the-middle and OAuth device-code techniques central to the September passkey campaign, and specifically documented an affiliated actor pairing vishing with fake Entra passkey enrollment — the closest known precedent to the Storm-3121/Storm-3032 pattern Microsoft has now confirmed at larger scale [6]. AI-Weaponized Phishing: Nation-State Quality at Commodity Scale provides the broader evidentiary basis for treating the invoice fraud campaign's AI-assisted drafting as part of a structural shift rather than an isolated data point, documenting the effectiveness inflection point at which AI-generated social engineering began outperforming human-authored attacks and recommending the same phishing-resistant MFA and out-of-band verification controls this note reiterates [5]. A third study, AI

Superpersuasion: Influence Operations and Enterprise Security Risk, published in June 2026, provides the evidentiary basis for this note's recommendation to de-emphasize content-based phishing detection in favor of process-focused, out-of-band verification, having found that frontier AI systems reliably outperform expert humans at persuasion [9]. Together, these findings map onto the identity and access management and threat-and-vulnerability-management domains of the AI Controls Matrix (AICM) v1.1, CSA's current control framework for AI-related risk, which organizations can use to structure conditional access, MFA, and API monitoring controls referenced above into an auditable governance program [8].

References

[1] Microsoft Security Research. "Passkey-themed social engineering leads to identity and cloud compromise." Microsoft Security Blog, September 9, 2026.

[2] RH-ISAC. "Passkey-Themed Social Engineering Lures Attempt to Compromise Identity and Cloud Platforms." RH-ISAC, September 2026.

[3] Microsoft Security Research. "Protecting organizations from AI-assisted executive impersonation and invoice fraud." Microsoft Security Blog, September 10, 2026.

[4] The Hacker News. "Attackers Use Passkey Phishing to Hijack Microsoft Cloud Accounts and Exfiltrate Data." The Hacker News, September 2026.

[5] Cloud Security Alliance. "AI-Weaponized Phishing: Nation-State Quality at Commodity Scale." CSA AI Safety Initiative, June 2026.

[6] Cloud Security Alliance. "Forg365: AI Lure Generation in an M365 Phishing Kit." CSA AI Safety Initiative, July 2026.

[7] Lawrence Abrams. "Passkey-themed phishing attacks lead to Microsoft 365 data theft." BleepingComputer, September 2026.

[8] Cloud Security Alliance. "AI Controls Matrix v1.1: Framework for Trustworthy AI." Cloud Security Alliance, 2026.

[9] Cloud Security Alliance. "AI Superpersuasion: Influence Operations and Enterprise Security Risk." CSA AI Safety Initiative, June 28, 2026.