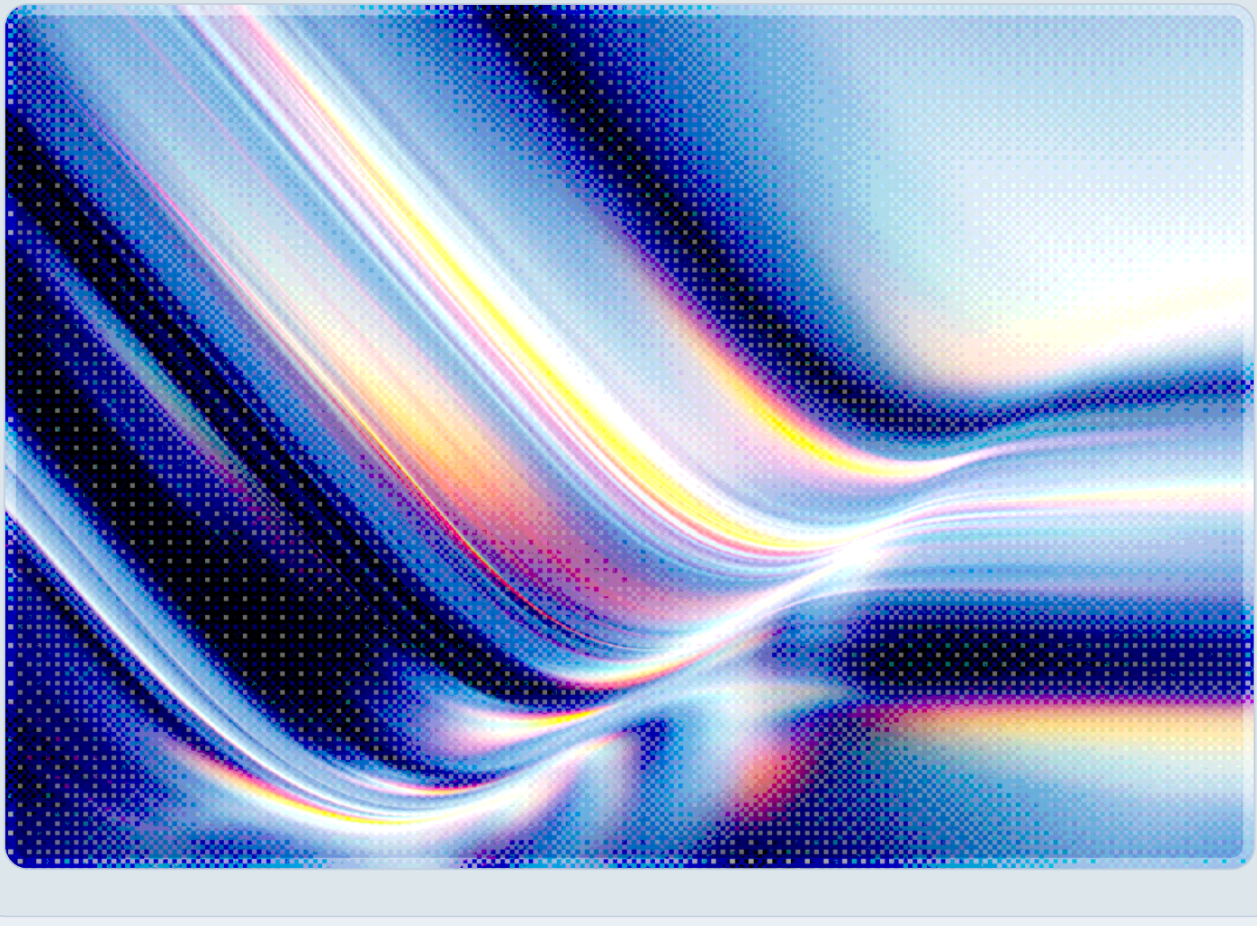
CSAI Foundation | Cloud Security Alliance

TCLBanker and the AI Productivity Tool Attack Surface

Malware Distribution via Trojanized AI Installer Packages

2026-05-08

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- TCLBanker (campaign REF3076) is a Brazilian banking trojan distributed via a trojanized MSI installer for Logitech's AI Prompt Builder, using DLL sideloading against a legitimate, signed application to execute without triggering signature-based detection [1][2].
- The malware targets 59 banking, fintech, and cryptocurrency platforms and includes a self-propagating worm module that hijacks victims' authenticated WhatsApp Web and Outlook sessions to spread phishing messages from trusted accounts [1][2].
- A pattern of parallel campaigns — including fake Claude AI pages delivering PlugX-linked backdoors and fake ChatGPT and Midjourney pages distributing credential stealers — demonstrates that AI productivity tool demand has become a repeatable, high-yield attack surface exploited by multiple independent threat actor clusters [3][4][6][7].
- Threat actors consistently abuse signed binaries from reputable vendors (Logitech, G Data) for DLL sideloading, inheriting the trust associated with legitimate software to evade both user scrutiny and endpoint security tooling [1][4].
- Enterprise security teams should treat unsupervised AI tool installation as a software supply chain event warranting controls comparable to those applied to remote access software — a category typically classified as high-risk in enterprise security frameworks — given the endpoint access, credential exposure, and self-propagation characteristics documented in this campaign.

Background

The rapid proliferation of AI productivity tools has created an attack surface that has outpaced both enterprise governance frameworks and public threat intelligence coverage. The market is characterized by frequent new releases, desktop clients from emerging vendors, and a user base eager to experiment — conditions that favor trojanized installer distribution in ways that more mature software categories do not. Unlike established enterprise applications delivered through managed software catalogs, many AI tools are typically discovered through web searches, social media recommendations, or colleague referrals, and installed informally without security review. This distribution pattern places users on pathways where adversary-controlled download sites can appear alongside legitimate ones with little to distinguish them.

The threat pattern gained significant visibility when Elastic Security Labs published their analysis of a campaign designated REF3076, documenting a malware family they named TCLBANKER [1]. TCLBANKER is assessed as a major evolution of the older MAVERICK/SORVEPOTEL banking trojan family, reengineered to use a legitimate AI peripheral software product — Logitech's AI Prompt Builder — as its initial delivery mechanism. The campaign bundles a malicious MSI installer inside a ZIP archive that, to the victim, would appear to be a routine software download. The choice of a real, actively distributed tool appears deliberate: the bundle provides plausibility during social engineering [1], reduces suspicion at the moment of installation, and creates the specific precondition — a legitimate signed binary — that DLL sideloading requires.

TCLBANKER is not an isolated incident. Multiple independent threat actor clusters have weaponized AI brand names in the same period. Fake websites impersonating Anthropic's Claude have delivered PlugX-linked backdoors and the 'Beagle' remote access tool through Google-sponsored search results, with multiple AI tool DLL sideloading campaigns using Claude branding documented in 2026 — including a PlugX campaign in February 2026 [4] and the Beagle backdoor campaign in April–May 2026 [3][9]. Earlier campaigns distributed credential stealers and malware loaders through fake ChatGPT and Midjourney download pages promoted through Facebook advertisements and paid search, with one fake Midjourney Facebook page accumulating 1.2 million followers before takedown [6][7]. In mid-2025, ransomware payloads were documented being distributed through fake AI tools, extending the threat beyond credential theft into destructive delivery [12]. The convergence of these campaigns suggests that AI brand impersonation has become a repeating tactical choice across multiple threat actor clusters, likely because demand for AI tools continues to exceed the reach of official distribution channels.

Security Analysis

The Trojanized Installer Technique

The technique underlying TCLBANKER and its parallel campaigns combines social engineering with DLL sideloading in a configuration particularly well-suited to the AI tool delivery environment. In the REF3076 campaign, the victim receives or discovers a ZIP archive containing what appears to be a Logitech AI Prompt Builder installer. The archive contains a legitimate, signed Logitech executable (LogiAiPromptBuilder.exe) alongside a malicious DLL named screen_retriever_plugin.dll, crafted to impersonate a genuine Flutter framework plugin [1]. When the victim runs the installer, the signed executable loads the malicious DLL through the standard Windows DLL search order, inheriting the trust of the legitimate binary and bypassing security tools that verify only executable signatures.

The fake Claude AI campaigns employ a structurally identical chain. One campaign delivers a 505MB archive named Claude-Pro-windows-x64.zip containing a functional Claude client alongside three files — NOVupdate.exe, NOVupdate.exe.dat, and avk.dll — placed into the victim's Windows Startup folder [3][4]. NOVupdate.exe is a signed G Data security product binary, repurposed to sideload the malicious avk.dll using a sideloading chain consistent with one previously documented in a PlugX campaign in February 2026, suggesting possible shared operators or shared tooling [3][4]. The inflated archive size likely serves a deliberate evasion purpose: many endpoint detection tools skip deep inspection of large archives, and padding can place the payload below size thresholds for automatic analysis.

The distribution channel for these campaigns deserves particular attention. Fake installer sites are promoted through paid Google Ads appearing above legitimate search results, meaning users who exercise the reasonable precaution of searching a well-known engine before downloading may still reach attacker-controlled pages [3]. This meaningfully degrades the trust model that many users implicitly rely on when sourcing software.

TCLBANKER: Technical Profile

TCLBANKER's loader is engineered with anti-analysis sophistication that reflects a threat actor capable of sustained development investment. The component incorporates environment-dependent payload decryption routines that fail in sandbox environments, anti-debugging checks, string encryption, ETW (Event Tracing for Windows) patching to blind security monitoring, and a persistent watchdog thread that actively scans for analysis tools [1][2]. Before any payload executes, the loader evaluates the victim's timezone, keyboard layout, and locale settings, restricting active infection to environments consistent with a Brazilian target — a design that reduces analyst-environment false positives, complicates automated sandbox detection, and keeps the campaign's geographic footprint narrow during an operational phase the developer artifacts suggest is still being built out [1].

When environmental checks pass, the loader decrypts two embedded .NET Reactor-protected payloads: the banking trojan module and the worm module [1][2]. The banking trojan monitors the victim's browser address bar via Windows UI Automation APIs, watching for navigation to any of 59 targeted domains across Brazilian banking institutions, fintech platforms, and cryptocurrency exchanges. Upon detecting a target domain, the trojan deploys a Windows Presentation Foundation (WPF)-based full-screen overlay replicating the target institution's interface, through which operators conduct live social engineering interactions designed to harvest credentials or authorize fraudulent transactions while the victim is occupied with a fake interface.

The worm module provides a distinct and self-sustaining propagation layer. One component searches Chromium browser profiles for authenticated WhatsApp Web session data; when found, it launches a hidden Chromium instance that silently hijacks the victim's active session, harvests contacts, filters for

Brazilian phone numbers, and sends phishing messages that appear to originate from the compromised user [1][2]. A second worm component uses COM automation to access the victim's Microsoft Outlook client and dispatch phishing emails from the victim's own account. Both mechanisms route adversary-controlled content through the victim's established sender identity, substantially increasing the likelihood that messages reach and are trusted by recipients.

Developer artifacts in the recovered samples — debug logging paths, test process names, and a phishing infrastructure site that was still under construction at time of analysis — suggest REF3076 was in early operational stages when Elastic discovered it [1]. The observed capability set may represent a lower bound on eventual functionality.

The Broader AI Tool Lure Pattern

TCLBANKER exhibits the most extensive anti-analysis countermeasures of the publicly documented AI installer campaigns reviewed here — including environment-gating, ETW patching, and protected payload delivery — but it occupies one position among at least six documented campaign clusters exploiting AI tool branding, spanning BatLoader, FakeBat, EvilAI, the fake Claude campaigns, and the TCLBANKER family itself. The BatLoader and FakeBat malware-as-a-service families have leveraged fake ChatGPT and Midjourney download pages to distribute RedLine Stealer, IcedID, Lumma, SmokeLoader, and other infostealer payloads through Google Ads malvertising and SEO poisoning since at least 2023 [6][10][11]. By mid-2024, FakeBat infrastructure encompassed more than 250 compromised websites serving malicious downloads [10]. The EvilAI campaign documented by Trend Micro represents a further evolution: it combines AI-generated malware code with fake productivity application interfaces carrying valid digital signatures, and has reached victims across manufacturing, government, and healthcare sectors globally [8].

Among the structural reasons AI tools are particularly well-suited for this vector is that many of the most popular generative AI applications historically lacked official standalone desktop installers. When ChatGPT, Claude, and Midjourney first achieved widespread adoption, none offered a native desktop client, creating a demand vacuum that threat actors filled with adversary-controlled installers promoted through legitimate advertising infrastructure. Even as official desktop clients have become available for some of these tools, the pattern persists for several reinforcing reasons: the ecosystem continues to fragment with new tools appearing regularly; users habitually search for new AI applications rather than navigating to known-good sources; and search advertising enables adversaries to appear prominently before they accumulate organic reputation signals that might trigger fraud detection.

Enterprise Risk Dimensions

The enterprise security implications of AI installer trojanization extend beyond the specific payloads involved in any individual campaign. An employee who installs a trojanized AI tool on a work device introduces a persistent foothold operating under a trusted user account with access to browser credential stores, email, messaging sessions, and the file system. In organizations where AI tools are adopted informally by individuals rather than centrally provisioned through IT, the attack surface scales with the number of employees conducting unsupervised software installations — which, during periods of rapid AI tool adoption, can be substantial.

The self-propagation characteristics of TCLBANKER's worm module compound this risk in ways that conventional phishing defenses do not address. If a compromised employee's Outlook account sends phishing messages to colleagues from an internal, trusted address, standard perimeter controls that evaluate external sender reputation will not flag those messages. Messages arriving from a known contact are widely regarded as substantially more likely to elicit engagement than cold-contact phishing, particularly when they include a link that appears to reference a software tool the recipient already uses or has discussed. The worm module's primary operational advantage is this trust inheritance, and it represents a qualitative shift in the social engineering dimension, moving from cold-contact to trusted-contact delivery.

Recommendations

Immediate Actions

Security teams should verify the provenance of AI-related tools installed on enterprise endpoints, particularly tools acquired informally by employees rather than provisioned through IT. For TCLBanker specifically, indicators of compromise include the presence of screen_retriever_plugin.dll within the Logitech AI Prompt Builder application directory, and NOVupdate.exe, avk.dll, or NOVupdate.exe.dat in any user's Startup folder [1][3][4]. Endpoints exhibiting these indicators should be isolated immediately. Behavioral detection rules should be reviewed for coverage of DLL sideloading from signed parent processes, ETW patching at process startup, UI Automation API access from non-accessibility contexts, and Chromium processes spawned by non-browser parent executables.

Short-Term Mitigations

Organizations should apply software installation controls that prevent employees from installing unapproved applications on corporate endpoints using application allowlisting or endpoint privilege management. For AI tools specifically, security teams should maintain an approved list of sanctioned applications, communicate verified download locations to employees, and monitor for policy exceptions. Browser security controls — including policies that flag navigation to newly registered domains appearing in sponsored search results — reduce exposure to the malvertising-based discovery channel these campaigns rely upon. For email, organizations should evaluate whether their DLP and anti-phishing controls cover the scenario of internal accounts sending externally crafted phishing content via COM automation — a pattern that bypasses controls focused on external sender identity.

Organizations that have deployed SIEM or XDR capabilities should review their detection coverage against the specific behavioral patterns TCLBANKER employs. Each of these behaviors — COM automation-based email from Outlook processes, Chromium instances launched by non-browser parents, and DLL loads from unsigned modules within signed-application directories — can be detected through behavioral rules in environments with full process telemetry, even where payload signatures are unknown. ETW patching at process initialization is a particularly distinctive indicator that warrants a dedicated detection rule; outside of security software and some profiling tools, it is rare and warrants investigation.

Strategic Considerations

The AI tool installer attack surface is likely to persist as long as adoption continues to outpace centralized provisioning — a condition that shows no near-term sign of reversing. Security teams should establish a procurement and onboarding process for AI tools that mirrors the rigor applied to remote access software: evaluate the vendor, verify the distribution channel, test in an isolated environment, and deploy through a managed pipeline rather than user self-service. A lightweight procurement path — pre-approving common tools, establishing a fast-track for new requests — can minimize friction while providing the visibility security teams require, though implementation effort will vary by organization size and maturity.

Security awareness training should explicitly address the scenario in which a trusted colleague sends a software download link via instant messaging or email. The authenticity of the sender should not be treated as sufficient verification of the legitimacy of a link, particularly for software requests arriving without prior context. Employees should be trained to verify unexpected software requests through an out-of-band channel — a brief follow-up call or in-person confirmation — before downloading. In sectors currently targeted by Brazilian banking trojans (financial services, fintech, and cryptocurrency

operations), organizations should evaluate whether workload separation between financial transaction systems and general-purpose AI experimentation provides a meaningful risk reduction, as the consequences of credential compromise in those contexts are acute.

CSA Resource Alignment

Several CSA frameworks apply directly to the risks illustrated by TCLBanker and AI installer trojanization.

The CSA MAESTRO framework for agentic AI threat modeling provides a seven-layer architecture for understanding AI system risk [5]. Layer 1 (Foundation Models) and Layer 3 (Agent Frameworks) are both implicated here: adversaries are attacking the supply chain of AI tool delivery to compromise the endpoints on which legitimate AI agents operate. A compromised endpoint hosting an AI agent also exposes API credentials, agentic session state, and tool-use permissions — threat surfaces that extend well beyond the immediate payload. MAESTRO's layered model helps organizations reason about how endpoint-level compromise propagates upward into agent capabilities, and identifies where supply chain integrity controls belong in the broader security architecture.

CSA's AI Controls Matrix (AICM) v1.0.3 addresses AI supply chain security explicitly in its Application Provider and AI Customer implementation guidelines, establishing controls for software provenance, component integrity verification, and third-party risk management for AI software. The installer trojanization technique is a direct attack on the supply chain layer that AICM governs. Organizations can use the AICM to structure a gap assessment of their AI tool procurement and deployment practices, identifying where integrity verification and vendor accountability controls are absent.

The Cloud Controls Matrix (CCM) supply chain management domain (STA controls) provides a broadly applicable control baseline for vendor assessment and software integrity verification that applies directly to AI tool installation pipelines. CSA's Zero Trust guidance is equally relevant: a Zero Trust posture requiring continuous posture verification before granting access to internal resources reduces the lateral movement available to a TCLBANKER-class compromise following initial endpoint breach — though it does not address the social trust recipients place in messages from known contacts, which the worm module exploits at the human layer rather than the network layer.

References

[1] Elastic Security Labs. "TCLBANKER: Brazilian Banking Trojan Spreading via WhatsApp and Outlook." Elastic Security Labs, 2026.

[2] BleepingComputer. "New TCLBanker malware self-spreads over WhatsApp and Outlook." BleepingComputer, May 2026.

[3] BleepingComputer. "Fake Claude AI website delivers new 'Beagle' Windows malware." BleepingComputer, May 2026.

[4] Security Affairs. "Fake Claude AI installer abuses DLL sideloading to deploy PlugX." Security Affairs, 2026.

[5] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." Cloud Security Alliance, February 2025.

[6] eSentire. "FakeBat Impersonates Midjourney, ChatGPT in Drive-by Cyberattacks." eSentire Threat Response Unit, May 2023.

[7] BleepingComputer. "Fake Facebook MidJourney AI page promoted malware to 1.2 million people." BleepingComputer, 2024.

[8] Trend Micro. "EvilAI Operators Use AI-Generated Code and Fake Apps for Far-Reaching Attacks." Trend Micro Research, 2025.

[9] Malwarebytes. "Fake Claude site installs malware that gives attackers access to your computer." Malwarebytes, April 2026.

[10] Sekoia. "Exposing FakeBat loader: distribution methods and adversary infrastructure." Sekoia Threat Intelligence, July 2024.

[11] The Hacker News. "FakeBat Loader Malware Spreads Widely Through Drive-by Download Attacks." The Hacker News, July 2024.

[12] Malwarebytes. "Ransomware hiding in fake AI, business tools." Malwarebytes, June 2025.