Agentic Commerce Heads into Peak Season: Familiar Fraud Through New Traffic

Executive Summary

The RH-ISAC intel team assesses with moderate confidence that most agentic commerce risk this holiday season will come from existing fraud types, such as card testing, account takeover, and checkout-page card theft, delivered through a growing volume of AI-driven traffic. Attacks aimed at shopping agents are documented as researcher demonstrations, lab results, and attempted payloads. Public reporting does not document confirmed fraud losses from such attacks. Attackers are also running AI agents of their own against merchants. Gambit Security documented a campaign in which agent tools were used to compromise online retailers, with more than 600,000 payment card records taken from two companies. This assessment rests on vendor telemetry and a small number of primary research reports.

Adobe expects AI-referred traffic to U.S. retail sites to rise 130% year over year, with the largest increases on Thanksgiving (159%), Black Friday (95%), and Cyber Monday (88%). Companies should classify agent traffic and tighten login and checkout controls before Thanksgiving, including a review of checkout pages for injected scripts. Companies that operate agents should also review how those agents handle untrusted web content.

Community Impact

Commerce accounted for nearly half (48%) of AI bot traffic on Akamai's global network from July to December 2025. Of the commerce web attacks that targeted API endpoints from January 2024 to December 2025, hospitality accounted for 33% and travel for 9%. In HUMAN's 2025 data, travel and hospitality received 19.2% of agentic traffic, behind retail and e-commerce at 46.6%.

Adobe forecasts $275.1 billion in U.S. online holiday spending from November 1 to December 31, including $47.5 billion during Cyber Week.

Akamai reports that commerce organizations, like most other industries, place more than 90% of their AI bot activity in a "monitor" category and allow about three-quarters of the remaining activity to pass unrestricted, which Akamai describes as looser than in other industries.

In a February 2026 U.S. survey by HUMAN, only 11.7% of respondents were comfortable with an AI booking travel deals without asking for approval, while 42.6% were comfortable if it asked for final approval before payment. Based on that survey, many agent purchases this season will likely keep a human approval step, which puts the approval screen itself in scope for manipulation.

Technical Details

Fraud Arriving Through Agent Traffic

HUMAN reports carding rates on heavily-targeted travel and hospitality businesses reached 17.6% of checkout requests in 2025, up from 3.1% in 2022. The median in 2025 was 0.35%. One airline's rate of

©RH-ISAC

attempted carding attacks rose from under 1% of transactions in early 2025 to over 8% at the summer peak, then fell back below 1% by year end. HUMAN's research team also observed a carding-like pattern mediated by an AI agent. They state the threat actor rapidly cycled through multiple credit-card additions and repeated payment-completion attempts to see which card would authorize. The sequence included 11 card-add attempts and 6 payment attempts across two sessions, followed by a pivot to loyalty-point redemption after the card paths failed, mirroring established carding workflows, but executed through an AI browser agent. These figures measure attempts across HUMAN's own customers and come from a vendor report.

Attackers Using AI Agents

Gambit Security recovered the staging server of an operator that used three open-source AI tools, Hermes, Strix, and Cairn, to find and exploit online retailers. Skimmers were ordered against at least 27 named victims and confirmed in place on 19, and a fraud partner counted more than 600,000 card records taken from two companies. Gambit lists a hospitality company, an airline, and a Japanese travel booking site among the targets, but describes access to some victims only as "some level of access." The report is interim, and part of its evidence comes from AI-generated logs that Gambit says may be inaccurate.

NorForge separately found exposed Cairn operator panels through internet scanning, including one cluster configured with the Chinese language with more than 1,100 attack projects spread across 14 dashboards likely associated with this group. Targets were e-commerce sites mostly in the US, Japan, and Australia, and NorForge could not confirm most victim outcomes.

Attacks on Shopping Agents

Unit 42 documented indirect prompt injection on public websites in March 2026 and cataloged 22 payload techniques. Payment-related examples included hidden instructions that tried to force a purchase through a Stripe link and trigger a $5,000 transfer through PayPal. A third targeted an AI ad-review agent and tried to get a scam ad approved. Unit 42 states it is not aware of confirmed successes against deployed ad-review agents.

Researchers have demonstrated hijacking of agentic browsers. At Black Hat USA 2026, Zenity Labs showed that hidden instructions in the content an agent reads, such as an email, a calendar invite, or a web page, could redirect agents in five browsers. In one demonstration, an agent had Amazon's shopping assistant order an item on the victim's credit card for shipment to the attacker. Zenity's release does not claim exploitation in the wild.

Guardio Labs found earlier that Perplexity's Comet browser, asked to buy an Apple Watch, completed a purchase on a fake store Guardio built. In some runs it auto filled the saved card and address without asking for confirmation.

F-Secure tested a deliberately permissive shopping agent in a lab. A poisoned product review sent the agent to a phishing page in 12% of runs, and the agent submitted personal data there. The test does not measure how often this happens in the real world.

Protocols and Verification

Visa's Trusted Agent Protocol has merchants verify a signature (RFC 9421) on each agent request, using the agent's public key. Mastercard states that only registered agents can transact through Agent Pay.

©RH-ISAC

Google announced its Universal Commerce Protocol (UCP) in January 2026, compatible with its Agent Payments Protocol (AP2).

Visa's page describes how verification works. Neither Visa's nor Mastercard's public page states how many merchants or agents use its protocol, or which attacks it covers. A July 2026 preprint tested three platforms, including Google's AP2 reference implementation, and reported 33 structural weaknesses in lab conditions with no real funds moved.

Preparation and Mitigations

Security researchers and service providers have recommended a cautious posture and provided guidance on preliminary steps that can help reduce risks before peak shopping season:

1. Classify agent traffic and move it from monitor-only to enforced policy. Log it separately and apply rate limits. Where a protocol such as Visa's Trusted Agent Protocol is supported, verify agent signatures.
2. Tighten card-testing and account takeover controls ahead of Cyber Week. Set velocity limits on card additions and payment attempts and review use of stored payment credentials.
3. Show full line items on any screen where a user approves an agent purchase. Unit 42 identifies a total-only approval screen as a way to approve a hidden item unknowingly.
4. For any agent your organization operates or sponsors, treat retrieved web content as untrusted input and require human approval for purchases. Limit the agent's payment and data permissions. Unit 42 recommends design-level defenses because models cannot reliably separate instructions from data.
5. Review checkout pages for injected third-party scripts. Gambit published the skimmer domains and loader pattern it observed in the campaign, which members can use to hunt.
6. Share any observations of agent-driven fraud with the RH-ISAC.

©RH-ISAC