



Riskified Launches Agent Identity Risk Intelligence: Know Who Is Behind AI Personal Assistants like Muse at Checkout and in Customer Service

October 8, 2026

As consumer AI assistants begin to shop, cancel, file claims and negotiate refunds on their owners' behalf, Riskified extends its identity risk intelligence to assess the person behind the agent and the legitimacy of what it does for them, for orders and the service interactions that follow.

NEW YORK--(BUSINESS WIRE)--Oct. 8, 2026-- [Riskified](#) (NYSE: RSKD), a leader in digital fraud and risk intelligence, today announced Agent Identity Risk Intelligence, a new set of capabilities designed to identify the AI personal assistant behind an order or a customer service request, resolve it to a real consumer identity, and return a risk signal in real time. Agent Identity Risk Intelligence extends Riskified's AI intelligence platform to a new actor in commerce: the consumer's own personal AI agent.

Consumer personal AI assistants such as [Meta's Muse](#), [Instinct](#), and [dots in ChatGPT](#) are moving from product discovery to action. Agents check out, keep working after the user closes the app, compare prices across sites by design, and proactively check for price-drop refunds and no-fee returns. Behaviors that were occasional for human shoppers, such as checking whether a price fell after purchase, canceling when a cheaper option appears, or reading a return policy to the letter, can become routine when an agent does them every day for every order. The question merchants face is no longer whether an interaction is automated, but who the automation acts for, and whether what it does on their behalf is behavior the merchant would accept from the person themselves.

Authenticated is not the same as trusted. New standards such as the [Personal Agent Protocol](#), introduced this month by Sierra and Meta with Shopify, Stripe, Walmart and others, may give merchants a secure way to confirm that an AI assistant is authorized to act on an account. That is necessary, and Riskified supports it. But authorization only answers the first risk question. An agent authorized by a stolen account is still an account takeover. An agent authorized by a throwaway fake account is still a first-order promotion abuser. An agent authorized by a refund-as-a-service ring, whose members hold real accounts and grant real permissions, is still fraud, now running at machine speed through the same support channels built to serve customers. The protocol proves the agent is authorized. Riskified proves whether the person who authorized it can be trusted.

The coming age of delegated fraud. The same capability that lets a legitimate shopper delegate a purchase or a return also lets a fraudster delegate a fraud scheme. Refund-as-a-service fraud rings already sell scripted claims of empty boxes and missing deliveries against retailer support teams. In April 2026, a U.S. federal court [sentenced eleven co-conspirators of the Artemis Refund Group](#), a ring that ran refund fraud as a subscription business: customers bought the goods, Artemis placed thousands of fraudulent orders against Amazon, Walmart, Target and Wayfair. Put a tireless, always-on AI agent in the hands of a fraud group like Artemis, and the scalability of refund fraud changes drastically.

"Agent protocols solve a real problem: merchants need a secure way to let a customer's assistant act for them. What they don't solve is whether that customer should be trusted. A fraud ring can authorize an agent as easily as a loyal shopper can," said Assaf Feldman, Chief Strategy Officer, Technology and Co-Founder of Riskified. "Riskified sits on top of those protocols and answers the harder question, using what our identity engine already knows about the person behind the agent, so merchants can welcome a good customer's assistant, decline a bad actor's, and notice the moment a real customer's agent starts doing something that the customer never would."

Announcing **Agent Identity Risk Intelligence**, part of Riskified's **AI Agent Intelligence**, first introduced in August 2025:

- 1. Agent Identity Risk Intelligence.** Riskified will recognize when an order or customer service request comes from a personal AI assistant. Where an assistant or commerce platform provides a verified agent credential, Riskified will accept it. Either way, Riskified will connect the request to the person behind it, drawing on what the network already knows from billions of orders, claims and chargebacks. This recognition will first be available to merchants on Shopify, where orders placed through Meta's Muse already arrive tagged as agent-originated, and will extend to other platforms as agent identification standards take hold. When an assistant acts on a shopper's behalf, much of the device and browser telemetry that merchants have relied on is absent. Where that telemetry is missing, Riskified leans on identity and network history, the behavior of the agent, which signals intent, and in post-checkout interactions, such as refund requests, it can engage the assistant directly, confirming details or applying step-up friction that a legitimate assistant can satisfy and a script cannot. The result is a risk decision that reflects both the identity behind the agent and the agent's own behavior. The same signals will feed AI Agent Intelligence reporting in Riskified Control Center, which will show merchants their share of orders and claims arriving through personal AI assistants, whether verified or inferred, broken out by assistant, where one can be identified. Agent Identity Risk Intelligence will be available through Riskified's existing APIs and the [AI Agent Approve](#) MCP server on AWS Marketplace.
- 2. Agent Identity Risk Intelligence for Decision Studio.** Riskified will support policies by agent type within Riskified Decision Studio, for example, auto-approving low-risk agent orders from verified identities while routing high-risk agent-originated refund claims for review. Control Center will report approval rates, fraud rates, and claim-abuse rates for agent-originated traffic alongside the rest of the business, so merchants can see whether that traffic behaves differently and tune their policies accordingly.
- 3. Agent Identity Risk Intelligence for Zendesk.** The same identity risk category will flow into Riskified's integration with Zendesk, so both human agents in Zendesk Agent Workspace and Zendesk's Specialized AI Agents can see when a

refund or return request arrives through a consumer AI assistant and act on the risk of the person behind it.

Since August 2025, Riskified has taken a consistent position on agentic commerce. Rather than inserting itself between merchants and the platforms building agentic checkout, it invests in the layer every protocol, assistant, and support channel ultimately depends on: knowing who is on the other side of the interaction.

Availability. Agent Identity Risk Intelligence opens to a limited beta for enterprise merchants in December 2026, with general availability expected to follow in 2027. Merchants interested in the beta can apply at riskified.com/lp/request-demo/.

About Riskified

Riskified (NYSE: RSKD) is a leader in digital fraud and risk intelligence. Many of the world's biggest brands and publicly traded companies rely on Riskified to stop fraud and abuse across the entire customer journey, from account creation to payments to disputes. Built and managed by a global team of risk analysts, data scientists, and researchers, Riskified's AI-powered platform analyzes the individual behind each interaction to provide real-time decisions and identity-based insights. Learn more at riskified.com.

View source version on businesswire.com: <https://www.businesswire.com/news/home/20261008934928/en/>

Riskified media contact

Corporate Communications: Julia Hallett, Director of Corporate Marketing and Communications press@riskified.com

Investor Relations: Stephen Shulstein, Head of Investor Relations ir@riskified.com

Source: Riskified Ltd.