

DATA SHEET

Satori Abuse Mailbox Agent

Eliminate phishing triage toil with autonomous abuse mailbox processing

Key benefits

- Eliminates L1 security toil by automating repetitive review of reported messages
- Reduces phishing threat response times to quickly close the windows of exposure to sophisticated threats
- Handles alert volumes consistently without the need for additional human staff
- Combines autonomous execution with guardrails, human supervision, explainability, and audit.

Modern security teams are overwhelmed. User-reported suspicious emails, data loss prevention (DLP) alerts, and policy violations generate thousands of repetitive, low-value decisions every day. Even with strong detection, security operations center (SOC) and data security teams remain buried in level 1 (L1) triage. This slows response, increases burnout, and limits the value of upstream investments.

Proofpoint Satori addresses these problems. Satori is a set of solutions for autonomous security operations. Powered by agentic AI, Satori safely automates high-volume security workflows. It removes repetitive triage while preserving trust, control, and accountability. Teams can scale operations without sacrificing governance.

Autonomous, mission-driven

Satori uses multi-step workflows called missions that are designed to achieve specific security outcomes. Missions reason using Proofpoint's rich telemetry, apply guardrails, act where permitted, and log every decision.

Satori is built on Proofpoint's human and AI-centric security platform. It can access first-party signals and control points competitors can't match. This enables deeper context, safer automation, and tighter write-back into existing workflows.

Satori Abuse Mailbox Agent

User-reported emails create a continuous stream of "Needs Manual Review" (NMR) submissions that demand analyst attention. Even with strong detection, triage remains a significant operational bottleneck. The Proofpoint Satori Abuse Mailbox Agent autonomously analyzes and classifies all abuse mailbox submissions routed through CLEAR. It evaluates each submission, removes noise, and ensures that only true threats require human investigation, labeling reported messages with a disposition such as phish, malware, spam, or clean.

Satori can trigger remediation workflows established by organizational configurations to quarantine or tag messages.

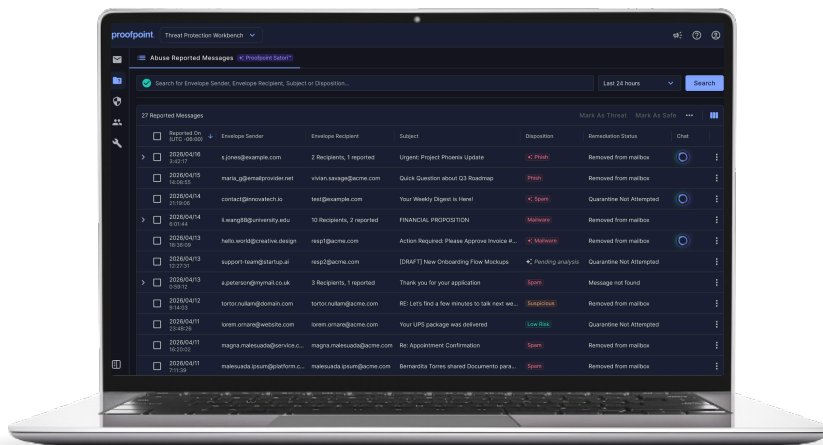


Figure 1. Satori Abuse Mailbox Agent operates in the Threat Protection Workbench. Satori decisions appear on the right in the Disposition column, denoted by the Satori “sparkle” and Satori icon in the Chat column.

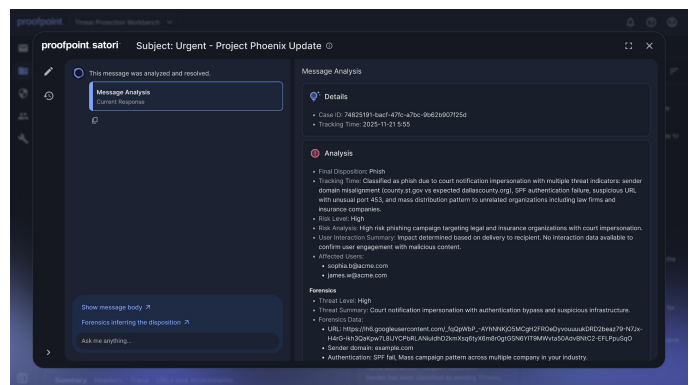
Execution with Full Explainability

The Satori Abuse Mailbox Agent operates directly within the Threat Protection Workbench, processing all incoming Needs Manual Review (NMR) submissions as an autonomous analyst. Satori provides a chat-based explainability interface, allowing analysts to ask questions about decisions and understand the reasoning behind classifications. This transparency builds trust and enables efficient validation of automated outcomes.

Precision with Satori Memories

Satori also leverages a “Memories” capability to tailor its outputs to each organization. It combines broad, global memories, based on universal patterns, with local memories established during a brief onboarding process and continuously refined over time through natural language instructions that team members share in the chat interface. Memories deliver precise, context aware results from day one, and also allow customers to continually refine the agent’s go-forward behavior.

Figure 2. Satori Abuse Mailbox Agent’s Explainability mode shows greater detail on Satori decisions. Its chat interface allows operators to ask any question about its decisions or share Memories to change future behavior.



Human-in-the-loop accountability

Agentic workflows are powerful and autonomous, but also benefit from multiple layers of human supervision for accountability. Humans-in-the-loop proactively sample and validate agent decisions for accuracy and policy adherence, while reviewing and even challenging Satori’s decisions to shape its go-forward behavior. Admins can escalate Satori decisions to the Proofpoint team for further review at any time.

Key Capabilities

Feature	What It Does
Abuse Mailbox Coverage (CLEAR NMR)	Automatically processes all CLEAR submissions labeled "Needs Manual Review."
Autonomous Email Classification	Evaluates and classifies user-reported emails using contextual and threat intelligence signals.
Execution Mode (Agentic Analyst)	Operates as an autonomous analyst, applying consistent decisions at scale.
Explainability Interface	Provides reasoning behind decisions via chat-based interaction.
Satori Memories	Memories are natural-language instructions shared with Satori to define go-forward behavior for ever-more precise outcomes mapped to organizational security, operational, or compliance preferences.
Business context onboarding	Satori offers immediate accuracy through an onboarding that analyzes historical alerts paired with a structured questionnaire to understand business context ensuring alerts are triaged correctly from day one.
Human-in-the-Loop Supervision	Enables escalation and controlled behavior refinement for future decisions.
Guardrails and Safe Actions	Mission guardrails bound agent actions; agents do not carry out enforcement (block/delete/policy changes); high risk alerts are escalated.
In-tool data privacy	Satori protects data in agentic workflows using strong cryptographic encryption and authenticated service-to-service communication. LLM inference occurs withing Proofpoint security-reviewed and contractually enforced system boundaries; data is processed ephemerally.
Workflow Triggering	Initiates downstream workflows that execute customer-defined actions (e.g., quarantine, tagging).
Audit and Logging	Captures classifications and interactions for auditability.
Seamless Workbench Integration	Operates directly within Threat Protection Workbench.
24x7 Elastic Operations	Scales to handle spikes in submission volume without performance impact.
Action-Based Pricing Model	Customers pay for automated work performed, not users, tokens, or seats.
Data Residency	Agent memories and chat history are stored at rest; alert data processed in memory and dropped after classification. Currently in the United States with EU data residency targeted for Q3 2026.



About Proofpoint, Inc. Proofpoint, Inc. is a global leader in human- and agent-centric cybersecurity, securing how people, data and AI agents connect across email, cloud and collaboration tools. Proofpoint is a trusted partner to over 80 of the Fortune 100, over 10,000 large enterprises, and millions of smaller organizations in stopping threats, preventing data loss, and building resilience across people and AI workflows. Proofpoint's collaboration and data security platform helps organizations of all sizes protect and empower their people while embracing AI securely and confidently. Learn more at www.proofpoint.com.

Connect with Proofpoint: [LinkedIn](#)

Proofpoint is a registered trademark or tradename of Proofpoint, Inc. in the U.S. and/or other countries. All other trademarks contained herein are the property of their respective owners.

DISCOVER THE PROOFPOINT PLATFORM →