bonfy.ai

SOLUTION BRIEF

Bonfy ACS™ For Agent Security

AI agents are rapidly moving from experiments to production, orchestrating complex workflows across SaaS apps, data stores, and external services. They read, write, and act on behalf of users and systems, often outside traditional security controls. Bonfy ACS™ provides adaptive content security for AI agents, protecting enterprise data across agent grounding, reasoning, and outputs without slowing innovation.

The AI Agent Risk Surface

As organizations adopt system-level agents, agentic coworkers, and browser-based agents, data now moves through multi-hop chains of LLMs, MCP servers, APIs, and SaaS systems. This creates new demands for data-surface visibility and posture management that traditional DLP, DSPM, and endpoint tools were never built to handle.

Key AI agent risks include:

- **Over-broad data access:** Agents are often grounded on entire repositories (SharePoint, Google Drive, CRM, file stores), exposing sensitive and regulated content to LLMs.
- **Unmonitored tool/API calls:** Agents freely send data to MCP servers or external services that may mishandle PII, client information, or IP.
- **Hallucinated or remixed content:** LLMs can generate outputs that unintentionally blend confidential, customer-specific, or regulated data.
- **Shadow agents and Shadow AI:** Employees adopt unsanctioned agent frameworks and browser-based agents that impersonate users and access corporate data without oversight.
- **No visibility or attribution:** Security teams lack insight into which agents accessed what data, how it was used, or which human or agent triggered the risk.

The result: organizations are “flying blind” as agents access, transform, and share sensitive content across systems the security stack does not monitor effectively.

Outcomes

With Bonfy ACS, organizations can:

- Adopt AI agents faster by enabling automated AI data governance
- Maintain strong privacy, regulatory, and contractual controls across agent workflows.
- Reduce tool sprawl by consolidating AI data security into a single, high-accuracy platform.
- Provide executives and boards with clear evidence that AI-related data risks are understood and controlled.

Bonfy ACS™: Data-Centric Security for AI Agents

Bonfy ACS™ is an AI data security platform that protects unstructured data across email, files, SaaS apps, collaboration tools, AI systems, and AI agents. Its entity-aware Knowledge Graph engine understands the business, people, customers, and consumers behind the data to deliver high-accuracy detection and true policy enforcement across both human and agentic workflows.

For AI agents, Bonfy ACS delivers three core protections:

- **Control what agents can see:** Bonfy limits which sensitive content agents can access so prompts and grounding data stay within business and trust boundaries.
- **Control what agents can do with data:** Bonfy lets agents check content safely during reasoning and execution to prevent misuse of PII, regulated data, and customer-specific information.
- **Control what leaves the organization:** Bonfy inspects agent outputs across email, SaaS, and collaboration tools to block risky messages and files before they leak sensitive data.

Key Capabilities for AI Agent Security

Bonfy ACS extends its multi-channel platform to AI agents with the following capabilities:

- **Unified multi-channel coverage:** One platform monitors data in motion, at rest, and in use across email, SaaS, collaboration tools, AI systems, and agents—eliminating blind spots.
- **Context-aware intelligence:** A knowledge graph maps your organizational entities and relationships for precise detection and fewer false positives.
- **Centralized policy and automation:** A single control plane streamlines discovery, labeling, remediation, and enforcement across all workflows.
- **Entity-level risk management:** Agents are modeled as first-class entities with cross-channel correlation and risk scoring.
- **Flexible deployment:** Delivered as cloud-native SaaS (multi-tenant, single-tenant, or BYOC) and integrates with Microsoft, Google, Salesforce, Slack, and more.

About Bonfy

Bonfy Adaptive Content Security™ is an AI Data Security platform that uses contextual intelligence, entity-aware analysis, and adaptive remediation to protect unstructured data across human and AI-driven workflows. Organizations gain unified visibility, accurate enforcement, and safe, scalable AI adoption without adding operational friction.