Island

Solution Brief

# Enterprise AI

## Govern a workforce that's both human and agentic.

## Introduction

Agents are joining your workforce, working alongside your people at a breakneck pace. They didn't go through onboarding. They don't know your rules. And they don't fear consequences.

That combination, scale plus autonomy plus no guardrails, is what makes ungoverned AI genuinely dangerous.

Most organizations are already past the point of deciding whether to adopt AI. They're in the middle of it, with sprawl they didn't plan for, costs they can't fully see, and an attack surface that grew faster than their controls did.

The numbers show how far the problem has already traveled.

*53%*

of AI interactions are autonomous actions

*93%*

of prompts are approved without the human in the loop actually being in the loop

*1 in 3*

MCP servers carries a high or critical severity finding

*92%*

Of MCP server owners have no verifiable organizational affiliation

Source: Island original research, 2026

## What changed

Before agents, AI governance mostly meant keeping sensitive data out of a chatbot. Agents changed the shape of the problem entirely. They operate on their own, run in a loop until the task is done, and will find a way to finish the mission, including ways nobody intended.

Three things happened at once:

### Cost moved outside anyone's control.

Every tool call, retry, and long-running task adds tokens automatically, with nobody approving it. AI spend is rising nearly fourfold as organizations move from isolated use cases to enterprise-wide adoption.

### The attack surface grew past what policy covered.

MCP servers get installed from anywhere. A single bad install can read every other tool on the host, along with its credentials, and quietly reach everything connected to that agent.

### Governance can't keep up.

No inventory of which agents exist. No policy over what they can do. No audit trail of what they did. When the board asks what agents are doing across the company, most security teams can't answer.

## Why existing tools only see part of it

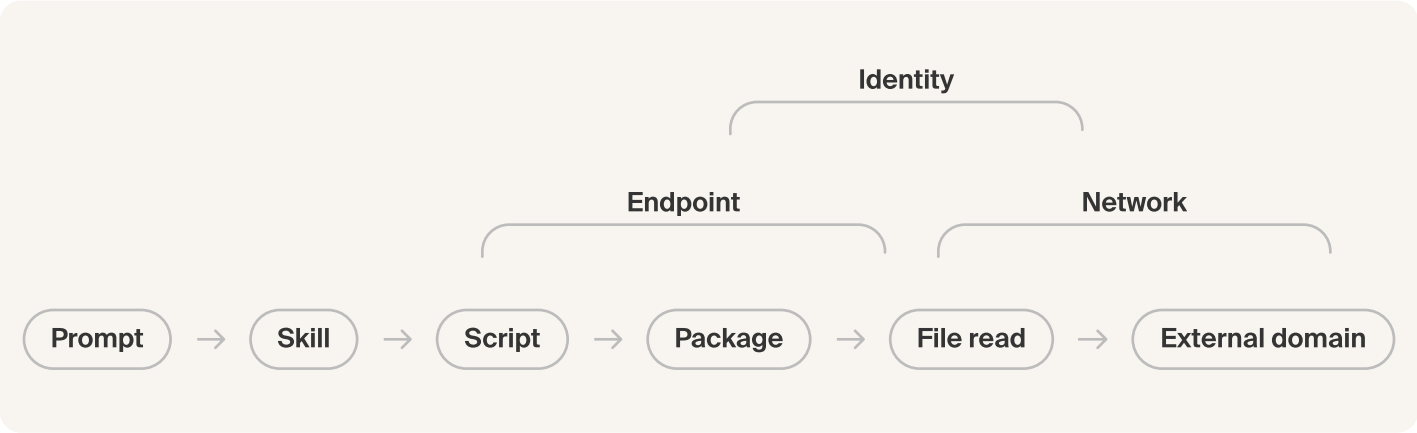
Every vendor approached this problem through the lens of what they already had:

**Network tools** see the wire, not what's happening on the endpoint.

**EDR** sees the endpoint, not the prompt or the intent behind it, and nothing inside a browser.

**Identity tools** see which credential touched a resource, not why.

**CASB** sees the service-side action, not the context behind it.

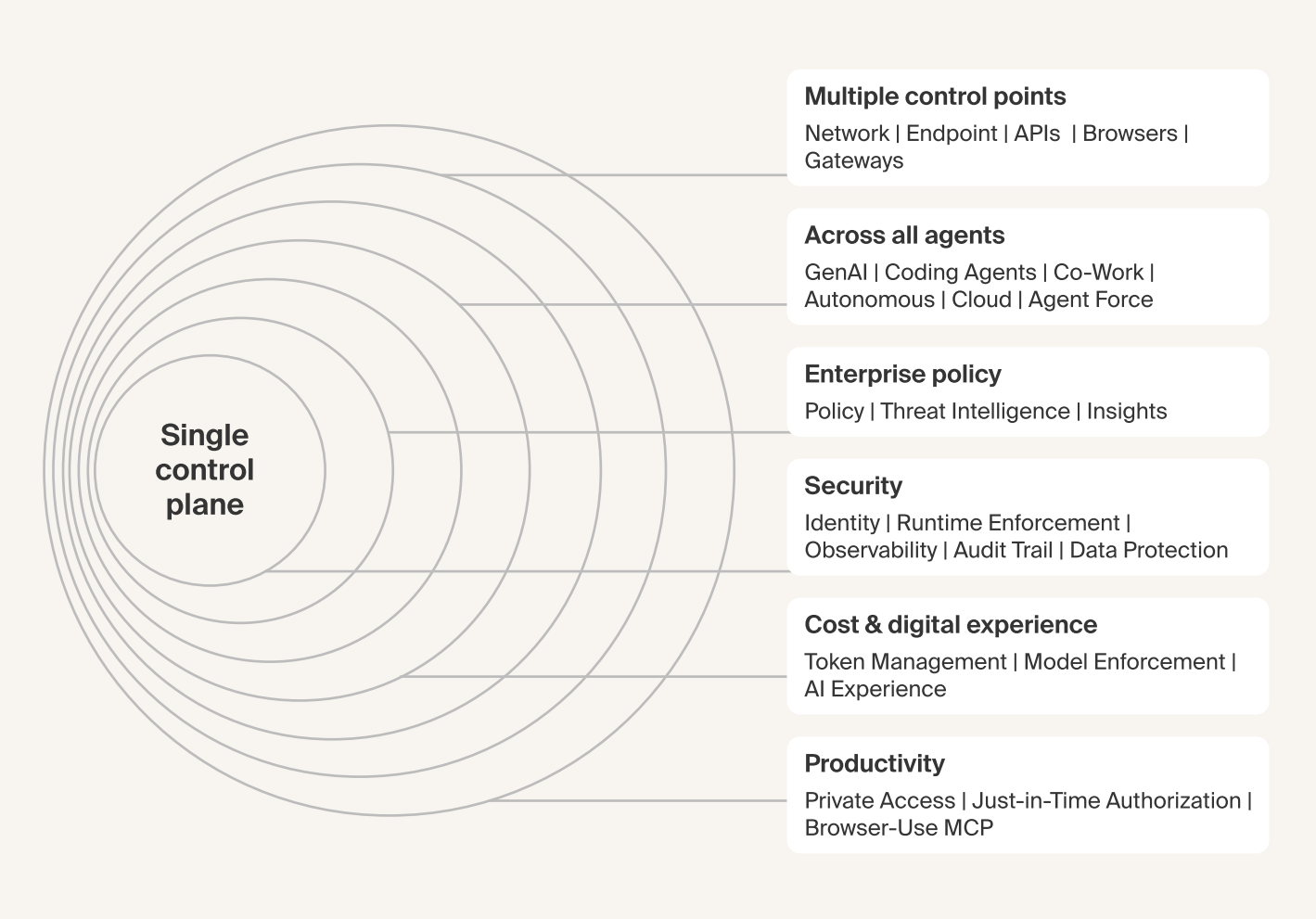



That's why incidents get discovered weeks after the fact, by someone piecing together logs from four different tools, instead of at the moment something went wrong. In an agentic world, that delay isn't acceptable. Agents don't wait. They act, they compound, and by the time a fragmented stack catches up, the blast radius is already set.

# How Island sees the chain and can act on it

Island already sits in the browser, on the endpoint, and on the network, because that's where people do their work. Agents run in those same places.

The browser adds something no other position has. When an agent works inside the browser the way a person would, that traffic looks identical to user actions to network and endpoint tools. Only from inside the browser can you tell the two apart, and only from there can policy treat them differently.

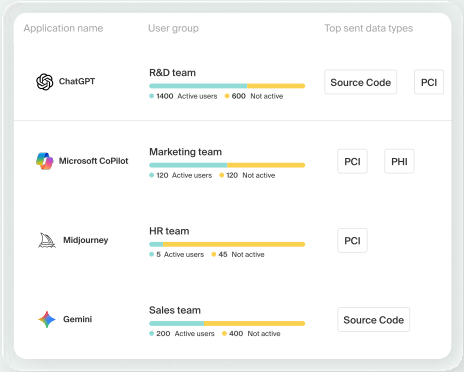



Island touches AI traffic through a set of control points: Island Enterprise Browser, Island Extension, Island Desktop, Island Network, the MCP Gateway, LLM gateway integrations, and OpenTelemetry. Together they feed one policy engine and one audit trail.

# What Island delivers

### Agentic Endpoint Posture

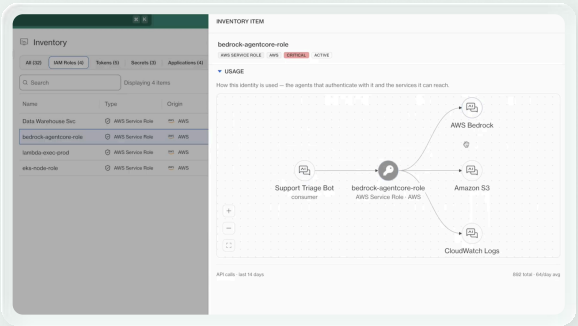
Island discovers every agent on the endpoint regardless of vendor and builds a full inventory: installed MCP servers, skills, code packages, browser extensions, development environment extensions, and hooks. It flags risky configurations, a malicious MCP server, an unapproved skill, an extension asking for far more permission than it needs. When something is flagged, Island removes it from the endpoint rather than only reporting on it.




### Agentic Identity

Every agent acting on a person's behalf needs its own non-human identity (NHI), and organizations are creating more of these than they can track. Island keeps a live inventory of every NHI, not a snapshot from provisioning but what's authenticated and running right now. Analyzing that activity surfaces what a static list never would: a hardcoded key in a repository, an identity dormant for months but still valid, an NHI whose access has grown past the task it was created for.

The MCP Gateway sits between every agent and the tools it calls. It issues just-in-time credentials scoped to one task that expire when the session ends. It holds keys itself, so the raw key never reaches the agent. It controls what an agent may do once it has access, including restricting it to read-only on a specific server or blocking a single destructive call.

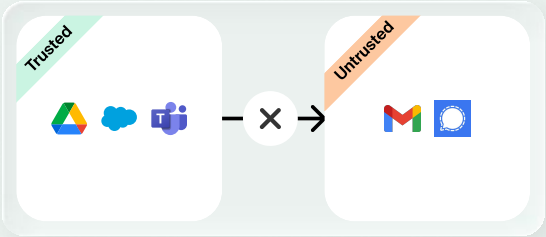


# What Island delivers

### AI Protect

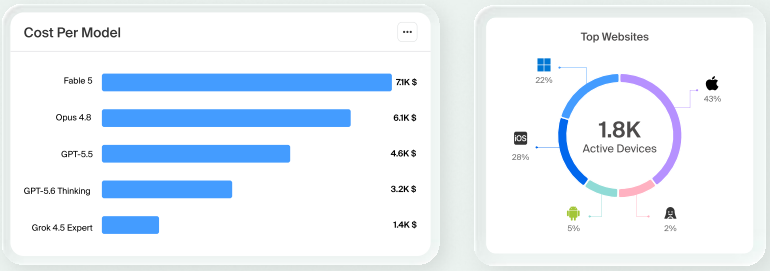
Real-time classifiers run inline and catch known threat types as they happen: prompt injection, jailbreak, agent drift, malicious URLs, malicious MCP servers, data leakage, and unwanted content. These are purpose-built models, not a general-purpose model handling every check, which is what lets them sit on the wire fast enough that the agent never feels it.

When Island prevents an agent's action, it tells the agent why. An agent drafting a customer reply tries to send a full account number in plain text. Island stops it and explains, and the agent rewrites the reply without the number instead of failing.




### AI Cost and Experience

Island tracks real usage, cost, and performance per person and per agent across browser, desktop, network, and tools, drawn from actual sessions and tool calls rather than estimates or login events. It catches a team routing every task through the priciest model when a cheaper one would do. Cost breaks down by session, model, and tool call. Because Island sits inline, it acts on what it finds, routing a task to a cheaper model when the premium one isn't needed and telling the user why in the moment.

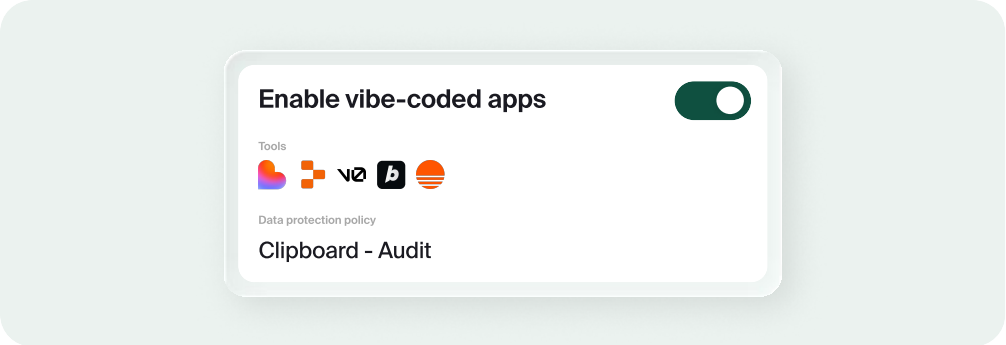


## What Island delivers

**Enterprise Vibe Publishing**

Anyone can describe what they want and have a working app before their coffee gets cold. The process built around a handful of vetted apps per quarter doesn't survive that volume. Sharing breaks because builders pass URLs around in chat. Governance breaks because internal apps hold sensitive information and reach sensitive systems.

Island Enterprise Vibe Publishing is the matching way to ship. A user builds in any common platform and submits the app through Island, where an admin attaches a standard policy covering access control, data protection, and device posture in a few clicks. The app publishes to a central library where employees search by category and owner. Published apps sit inside Island's application boundary with full data controls applied. Nobody reaches one outside Island.




## What Island doesn't replace

The existing security stack covers real ground. Network tools, EDR, identity products, and vendor admin consoles all stay in place. None of them was built to govern what an agent does at the moment it makes a tool call, reads a file, or moves data somewhere it shouldn't.

Island doesn't build agents and doesn't run them. It doesn't compete with the vendors whose models power them. Agents already have the run of your systems. They never got the handbook explaining what they're allowed to do with it. Island is what gives them one.

## Ready to govern your *agentic workforce?*

See how Fortune 500 companies enable AI at work with Island.

**Schedule a demo at island.io**