Island

AI Playbook for Security Teams

Your guide to evaluating AI risk, closing enterprise security gaps, and confidently saying yes to AI at scale.

AI adoption is accelerating. The security tools you rely on were not built for it. This playbook gives you the framework to close the gaps.

THE CHALLENGE FOR SECURITY TEAMS

AI sprawl is outpacing the controls built to stop it.

Your organization has a sanctioned AI provider. Maybe two. Claude, Copilot, ChatGPT, or Gemini. But that only covers a sliver of what's actually happening.

Business units are buying their own AI tools. Employees are signing into personal AI accounts and installing browser extensions. Executives are pushing AI into every function without a unified framework. The sprawl is unprecedented.

Legacy DLP, endpoint protection, and network monitoring were built for a world where applications didn't read content, observe behavior, retain context, connect to internal systems, or take autonomous action. AI does all of these things.

Three gaps security teams are dealing with right now

Visibility gaps
No single view into which AI tools, agents, models, or tenants employees are actually using across web, desktop, and extensions.

Control gaps
Existing point solutions cover fragmented entry points. Getting to full coverage requires patching together 7 or more tools.

Data exposure
Sensitive data enters AI prompts. Corporate data moves to personal AI accounts. Model training on employee inputs goes unchecked.

This playbook gives security teams a practical framework for understanding the AI threat landscape, evaluating where gaps exist, and building the controls needed to stop saying no to AI and start saying yes with confidence.

WHAT'S INSIDE

Your roadmap to enterprise AI security

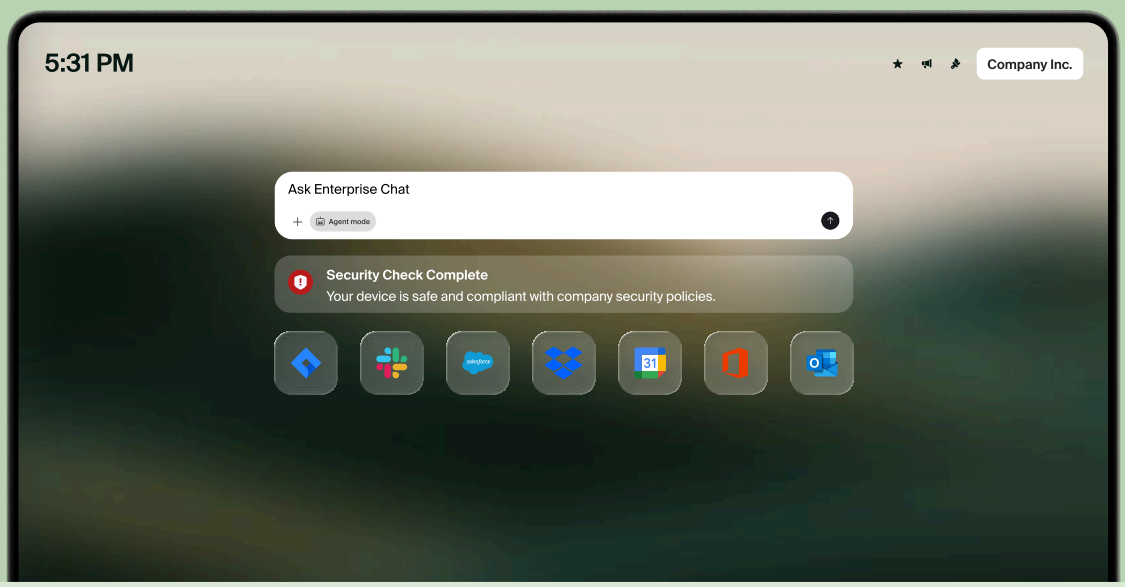
1

The New AI Threat Surface

AI is not arriving through a single door. It enters through five, six, eight simultaneous entry points, and most of those entry points were not on your radar when you last updated your security architecture.

Where AI enters the enterprise today

- **Browser-based AI (web destinations)** Consumer AI applications accessed through any browser, including personal accounts for tools like Claude, ChatGPT, and Gemini. Most organizations have no way to distinguish corporate tenant usage from personal usage.
- **Browser extensions** AI-powered extensions that can read page content, inject into prompts, and exfiltrate data. Risk profiles change dynamically as extension behavior evolves.
- **Agentic AI browsers** Emerging browsers like Atlas and Comet ship with AI natively embedded and operate outside corporate policy by default.
- **Desktop AI applications** Tools like the ChatGPT desktop app and Claude Cowork operate at the OS level, outside browser visibility and standard DLP coverage.
- **AI-powered IDEs** Development environments with AI built in, like Cursor and GitHub Copilot, operate at the code level with access to sensitive repositories, internal tooling, & proprietary logic.
- **AI connections to enterprise systems** MCP integrations and AI connectors that move data between AI tools and internal applications like Slack, Salesforce, and internal databases.
- **Network-level AI traffic** AI activity traversing the network that point security tools may not capture or attribute to specific users, tools, or tenants. This includes AI coding tools like Claude Code, and productivity tools like Claude Cowork, that operate outside the browser entirely.
- **AI agents** Autonomous software that can read application content, take actions across systems, and retain context across sessions. This includes vibe coding apps like Lovable, Base44, and Replit that generate and deploy code autonomously on behalf of users.



Why legacy controls fall short

Legacy security tools were designed to protect a perimeter. AI collapses the perimeter. Traditional DLP cannot see what employees are typing into AI prompts. Endpoint tools do not understand model training implications. Network monitoring cannot distinguish between corporate and personal AI tenant usage. And no single legacy tool covers all eight entry points listed above.

The result is a forced choice most security leaders are already confronting: block AI usage and slow down the business, or accept risk and hope the gaps do not get exploited. Neither is a real strategy.

The next wave of AI raises the stakes further.

AI-native browsers, agentic assistants, and AI agent software can read application content, observe behavior, retain context, connect to internal applications, and take action across sensitive systems. This expands the attack surface and introduces new threats like prompt injection and automated misuse at scale, precisely where traditional security controls are least effective.

The Hidden Costs of *Blocking AI*

For many security teams, the instinct is to block first and evaluate later. It feels like the cautious move. But blocking AI carries real costs that often go unaccounted for in the security calculus.

What blocking actually costs the organization

Innovation slowdown

When employees cannot access AI tools, they lose competitive productivity. Teams in peer organizations are moving faster. Yours are not.

Shadow AI growth

Blocking sanctioned tools does not eliminate AI usage. It drives it underground. Employees use personal accounts, unmonitored devices, and workarounds that create greater exposure than the sanctioned tools would have.

Missed return on AI investment

AI promises real productivity gains. Most organizations are still struggling to see a return. Three obstacles keep getting in the way: getting the right AI to the right users, deploying AI apps at company scale, and scaling AI agents across business functions.

The goal for security teams is not to stop AI. It is to make AI safe to use. That requires a fundamentally different approach than what most security stacks are currently capable of delivering.

Eight Entry Points *Security Teams Must Cover*

A complete AI security strategy requires coverage across every entry point where AI enters or operates in your environment. Most organizations have partial coverage at best. Here is a map of the eight critical areas and the key questions to ask for each.

8 AI entry points security teams must cover

1 Browser-Based AI

- Can you distinguish corporate AI tenant usage from personal accounts?
- Do you have visibility into what data employees are sending to each tool?
- Can you enforce policy on prompt content before it leaves the browser?

2 AI Browser Extensions

- Do you have a real-time inventory of every AI extension installed across the organization?
- Can you assess extension risk dynamically as extension behavior changes?
- Can you block or modify extensions based on risk profile?

3 Emerging AI Browsers

- Does your security posture extend to employees who switch to AI-native browsers like Atlas or Comet?
- Can you maintain visibility and policy enforcement outside your managed browser environment?

4 Desktop AI Applications

- Does your endpoint coverage extend to the data flowing through tools like the ChatGPT desktop app?
- Can you apply the same DLP policies that govern browser-based AI to desktop AI applications?

5 AI-Powered IDEs

- Do you have visibility into what code and internal tooling AI-powered IDEs like Cursor and GitHub Copilot can access?
- Can you govern what sensitive repositories and proprietary logic these tools are exposed to?

6 MCP Integrations and AI Connectors

- Can you see MCP tool calls that move data between applications like Slack and ChatGPT?
- Can you govern which MCP integrations agents are permitted to use?
- Do you have visibility into what company knowledge users are accessing via AI connectors?

7 Network-Level AI Traffic

- Is AI activity traversing your network captured and attributed to specific users, tools, and tenants?
- Does your coverage extend to tools that operate entirely outside the browser, like Claude Code and Claude Cowork?
- Can you maintain visibility across roaming users and off-network devices?

8 AI Agents and Vibe Coding Apps

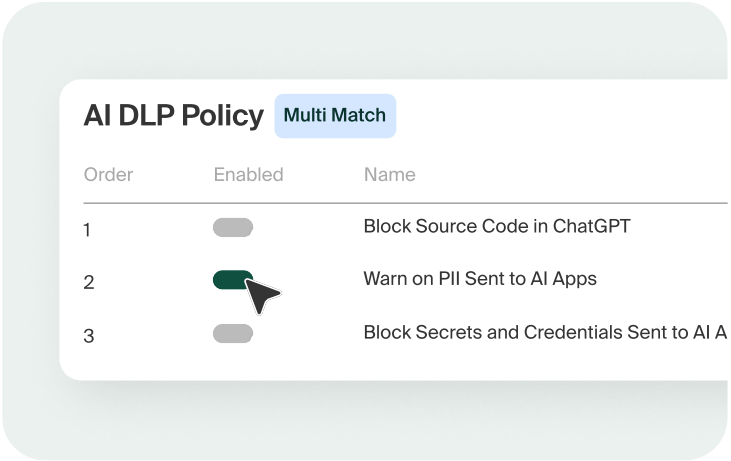
- Do your AI agents have assigned identities with defined permissions?
- Is every agent action fully audited?
- Do you have the ability to apply human-in-the-loop controls for sensitive operations?
- Can you govern vibe coding apps like Lovable, Base44, and Replit that generate and deploy code autonomously on behalf of users?

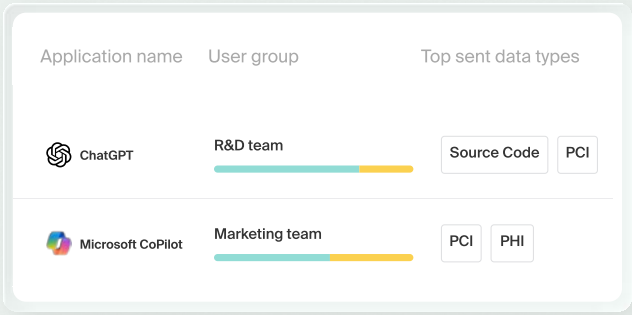
What Enterprise AI Controls *Actually Look Like*

When enterprises have full AI coverage in place, it looks nothing like patching together point solutions. Here is what mature AI security controls deliver across five key dimensions.

Visibility and Auditability

- Full visibility into AI usage across web and desktop, including which corporate vs. personal tenants are in use.
- Visibility into MCP tool calls that move data between applications.
- Deep audit logs capturing every AI interaction, including full chat conversations, to support compliance and incident review.
- AI data lineage showing the flow of corporate data across all AI interactions.



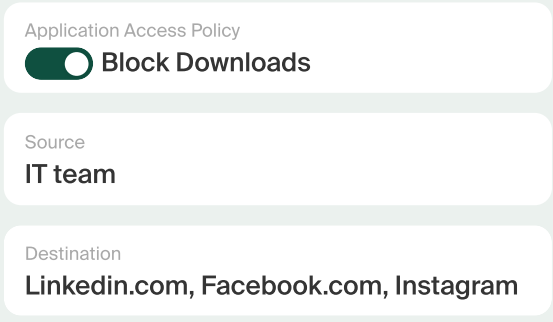


Data Protection

- Data loss prevention that redacts sensitive data before it reaches AI provider systems, at the prompt level.
- Data boundaries that keep corporate data from moving to personal AI accounts.
- AI-provider boundaries governing what data AI providers can see from application context.
- Built-in prompt injection defenses, including spotlighting and LLM classifiers.
- The ability to prevent AI vendors from training their models on employee prompts.

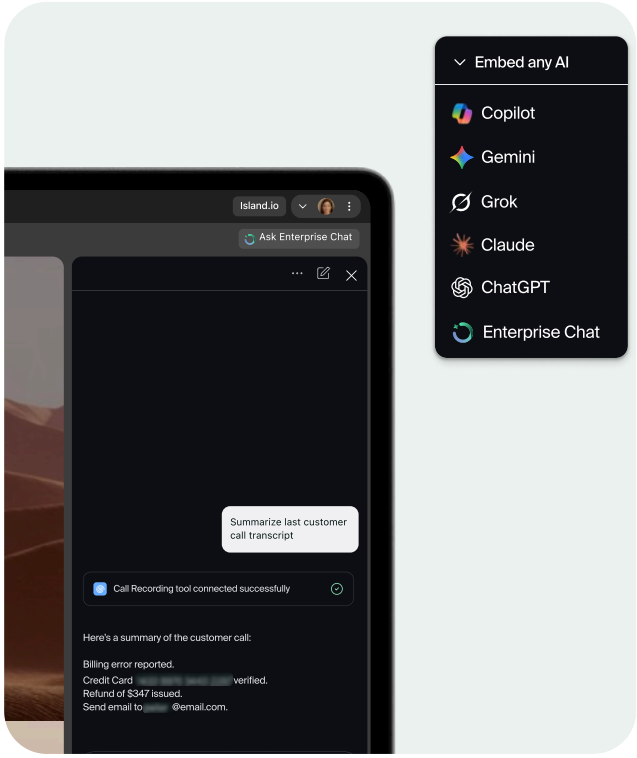
Access and Policy Control

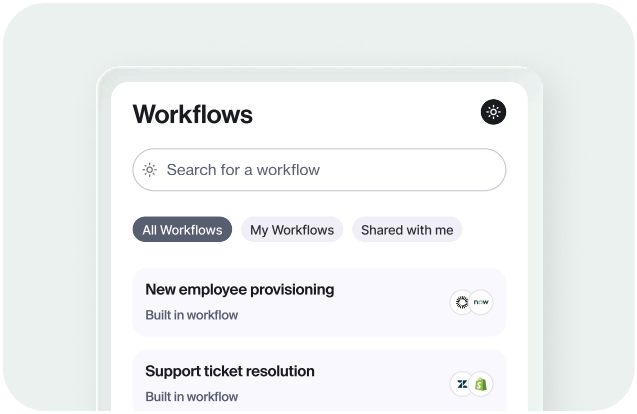
- Assigned policy for AI agents with defined permissions and scope.
- Context-aware policies governing prompt capture and conversation logging to meet regional compliance standards.
- Last-mile controls that allow administrators to modify AI tools without touching the underlying application: disable LLM tool connections, limit memory, enforce SSO, watermark content, and adjust interfaces.
- Extension management with risk scoring that updates dynamically when extensions go malicious.
- In-browser guidance that redirects employees to approved tools and explains policy in context.
- Private access controls that connect AI apps and MCP integrations to internal systems without requiring a VPN



AI Flexibility and Model Choice

- The ability to embed any AI provider the organization chooses, including Grok, ChatGPT, Gemini, Copilot, and Claude, so employees are not locked into a single model and organizations are not locked into a single vendor.
- Smart model routing that directs the right AI model to the right user based on role, task, and cost, so users always get the most relevant AI without having to think about it.
- Visibility into which models users are accessing, how frequently, and at what cost across teams and roles.
- A production-ready enterprise chat interface that supports multiple frontier models simultaneously, allowing organizations to optimize the best AI per task and per role.
- Support for an organization's own fine-tuned models alongside commercially available ones.





Agent Governance

- Full audit trails capturing every agent action and outcome.
- Human-in-the-loop controls for operations involving sensitive systems.
- MCP gateway governance controlling which integrations agents can access.
- A hardened execution environment that reduces exposure to prompt injection and misuse.

The "Say Yes to AI" *Checklist*

Before security teams can confidently greenlight AI for the enterprise, certain controls must be in place. Use this checklist to evaluate your current readiness. The more items you can check off, the more safely your organization can move from AI experimentation to AI at scale.

Checklist:

VISIBILITY

- [ ] We can see all AI tool usage across browser, desktop, and extensions in one place.
- [ ] We can distinguish corporate AI tenant usage from personal account usage.
- [ ] We have audit logs capturing AI interactions, including prompt content, for all users.
- [ ] We have visibility into MCP connections and AI connector usage.
- [ ] We can see AI agent actions and outcomes in real time.

DATA PROTECTION

- [ ] Sensitive data is redacted or blocked before it reaches AI provider systems.
- [ ] We have enforced boundaries that prevent corporate data from moving to personal AI accounts.
- [ ] We can control what application context AI providers are allowed to see.
- [ ] We have protections against prompt injection attacks.
- [ ] AI model training on employee prompts is disabled for all accounts.

ACCESS CONTROL

- [ ] Context-aware AI access policies are enforced consistently across all entry points.
- [ ] Employees are automatically authenticated to corporate AI tenants, not personal ones.
- [ ] We can approve, block, or modify AI tools at the last mile without changing the underlying application.
- [ ] AI browser extensions are continuously monitored with risk-based controls.
- [ ] Employees are guided toward approved AI tools with in-context policy communication.

AGENT GOVERNANCE

- [ ] AI agents have assigned AI policy with scoped permissions.
- [ ] Agent actions are fully audited and logged.
- [ ] Human-in-the-loop controls are available for sensitive operations.
- [ ] MCP integrations used by agents are governed and monitored.
- [ ] Agents operate in a hardened environment with prompt injection protections.

Checklist:

AI APP PUBLISHING

- [ ] We have a way to evaluate and onboard user-built AI apps without relying on a developer backlog.
- [ ] AI apps built by employees are delivered with identity, data protection, and access controls already applied.
- [ ] Published AI apps are visible to IT, with full audit trails for every user and every interaction.
- [ ] Employees can access approved AI apps directly alongside the tools they already use, without change management.
- [ ] We are not dependent on engineering teams to make a user-built AI app enterprise-ready.

ORGANIZATIONAL READINESS

- [ ] We have a clear policy framework that covers all AI entry points.
- [ ] Business units understand what AI tools are approved and why.
- [ ] We have a process for evaluating and onboarding new AI tools safely.
- [ ] AI costs and usage are visible and attributable to teams and roles.
- [ ] We are not relying on 7 or more separate tools to achieve AI coverage.

How to score yourself

Fewer than 18 checked: Your organization is in early-stage AI security. Focus first on visibility and data protection before expanding controls.

18 to 24 checked: You have partial coverage but meaningful gaps remain. Prioritize unified policy enforcement and agent governance.

25 to 30 checked: You have the controls in place to scale AI confidently. Focus on optimizing delivery and compounding productivity.

AI Maturity *Assessment*

Not every organization is at the same point in its AI security journey. Use this assessment to identify where your organization currently sits and what the most important next steps look like from there.

Stage 1 Reactive

You are aware AI usage is happening but have limited visibility into what, where, or how. Security posture is largely reactive: block when something surfaces, investigate after incidents. Point solutions may exist but coverage is inconsistent.

Key signals:

- No unified view of AI tool usage across the organization.
- Data protection relies on acceptable use policies rather than technical controls.
- No formal process for evaluating or approving AI tools.
- AI agents and MCP integrations are ungoverned.

Where to start:
Build your visibility foundation first. You cannot govern what you cannot see. Establish a single view of AI usage across web and desktop before expanding controls.

Stage 2 Developing

Your organization has some AI controls in place but coverage is fragmented. You likely have browser-level visibility or DLP for specific tools, but significant gaps remain across desktop, extensions, agents, and MCP integrations.

Key signals:

- Visibility exists for some AI tools but not others.
- DLP policies may cover web but not desktop AI applications.
- Extension management is ad hoc or manual.
- Agent governance is underdeveloped or absent.

Where to start:
Close coverage gaps across all entry points. Prioritize desktop AI protection and extension management. Establish agent identity and governance before autonomous AI use scales.

Stage 3 Defined

Your organization has consistent AI controls across most entry points. Policies are documented and enforced. You have meaningful visibility into AI usage and can audit interactions. Gaps may remain in agent governance and MCP oversight.

Key signals:

- Unified visibility across browser and desktop AI.
- Data protection policies are active and enforced.
- AI access policies are consistent across user roles.
- Agent use is emerging but governance is still maturing.

Where to start:
Mature your agent governance framework. Build out MCP integration controls. Begin measuring AI productivity impact alongside security posture.

Stage 4 Optimized

Your organization has full coverage across all AI entry points. Controls are consistent, automated, and integrated with your broader security stack. You can confidently say yes to AI tools, agents, and automation with full governance and oversight.

Key signals:

- Complete coverage across browser, desktop, extensions, agents, and MCP.
- Context-aware policy enforcement across all entry points.
- Full audit trails and data lineage for all AI interactions.
- AI costs are visible and attributable across teams and roles.

Where to focus next:
Shift from security defense to AI enablement. Use your governance infrastructure to accelerate AI adoption, deliver AI in workflows, and scale automation safely.

A Platform Built *for the AI Era*

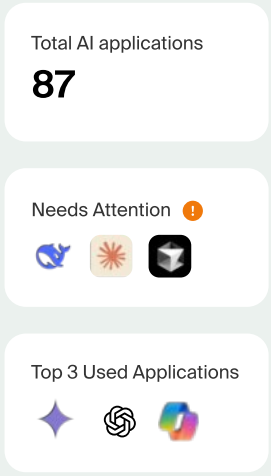
The controls described in this playbook are real. But most organizations trying to achieve them are doing so by assembling 7 or more point solutions across a fragmented entry point landscape, accepting coverage gaps and policy inconsistency as the cost of doing business.

Island AI Services was built to close those gaps from a single platform. Island takes the AI your teams already use, adds enterprise-grade controls, and delivers it through an environment designed for how AI actually enters and operates in the enterprise today.

AI Protect:
Visibility, control, and data protection across every AI entry point

Island AI Protect gives security teams complete visibility and control across all AI entry points: browser, desktop, extensions, and network. It replaces the fragmented point solution approach with one platform and unified policy.

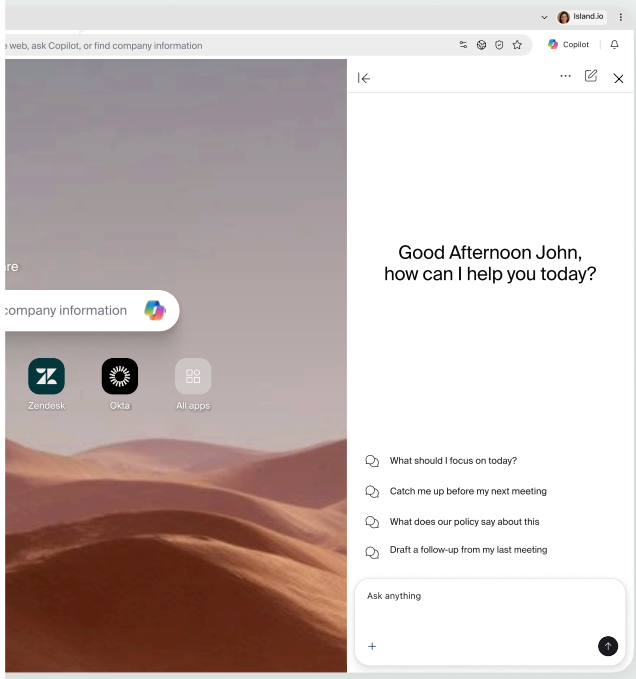
- Full visibility and data lineage into AI usage across web and desktop, including the breakdown of corporate vs. personal tenant usage.
- Data protection that redacts sensitive data before AI systems can access it and intercepts AI responses before they are rendered to the user.
- Data boundaries that keep corporate data from moving to personal AI tools.
- Built-in prompt injection defenses including spotlighting and LLM classifiers.
- Extension management with risk scoring across 200,000+ extensions that updates dynamically when extensions go malicious.
- Deep audit logs capturing every AI interaction to support compliance and incident review.



AI Browser:
AI embedded in the workflow, with enterprise context

The Island Enterprise AI Browser embeds any AI provider directly into every application page along with enterprise context. Users get the right AI for their job, in the flow of their work, without security needing to compromise on controls.

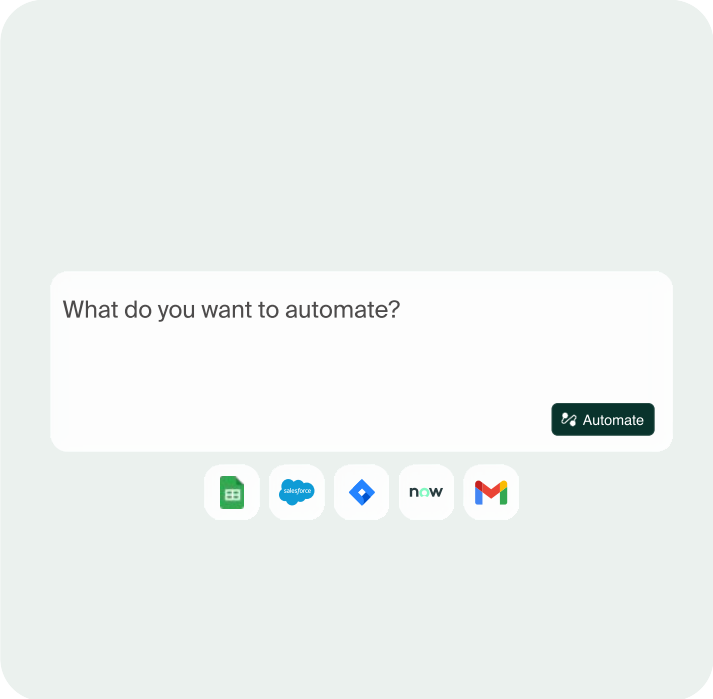
- Supports any AI provider: Grok, ChatGPT, Gemini, Copilot, Claude, and more.
- Enterprise context enriches AI interactions with knowledge of the user and role, application content, usage patterns, and company information.
- AI-provider data boundaries ensure that what AI providers can see is governed by policy, not left open by default.
- Enterprise Chat provides a production-ready, branded chat interface supporting multiple frontier models simultaneously, without engineering months.



AI Automation:
Governed agents that actually scale

Island AI Automation gives organizations the ability to build, share, and run AI agents inside a hardened enterprise environment, with full oversight and auditability.

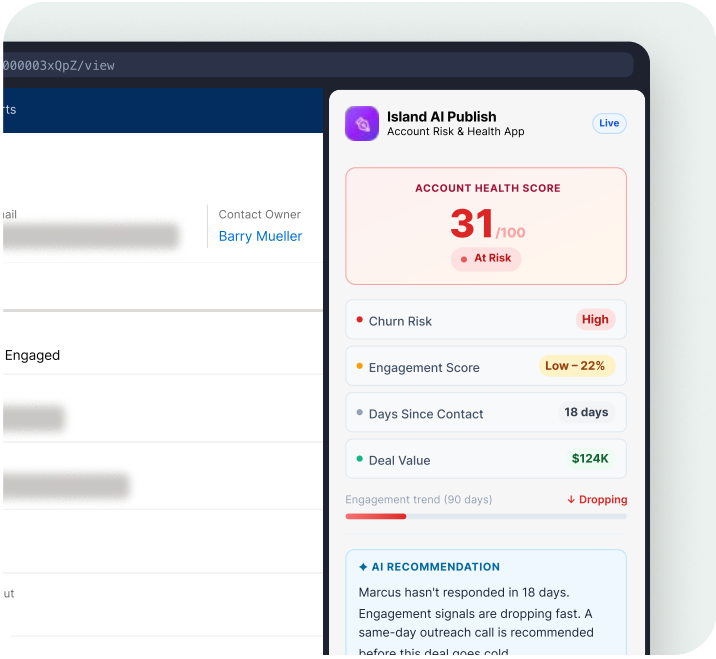
- On-demand agents for personal tasks invoked directly in the browser.
- Managed agents built in the Island AI Agent Studio with no-code configuration, defined workflows, and human-in-the-loop controls.
- MCP Gateway governed by Island, brokering agent access to 500+ integrations with flexible, policy-driven permissions.
- Full audit trails capturing every agent action and outcome.



AI Publish:
Enterprise-grade delivery for the apps teams are already building

Island AI Publish takes the AI apps your teams are already building and delivers them across the organization with identity, data protection, and policy controls already applied. No engineering backlog. No separate governance stack.

- Take any app built on tools like Lovable or Base44 and wrap it in a full enterprise envelope in minutes.
- Publish apps side-by-side with existing tools like Salesforce or ServiceNow without touching the underlying application.
- IT gets full visibility into every published app, every user, and every interaction.



One platform. Built for the AI era. So enterprises can move from AI chaos to AI advantage.

FINAL THOUGHT

AI security is not about saying no.

The organizations that will get the most out of AI are not the ones that blocked it longest. They are the ones that built the right controls to deploy it safely, deliver it to the right users, and scale it across the business with full visibility and governance.

Security teams are not standing in the way of AI. They are the ones who make it possible to go further, faster, and with confidence.

Island AI Services gives security teams complete visibility and control across every AI entry point, replacing fragmented point solutions with one platform that makes enterprise AI safe, productive, and governable at scale.

Ready to say yes to AI?

Schedule a demo ↪

Island