sonet.io

White Paper

AI Without APIs

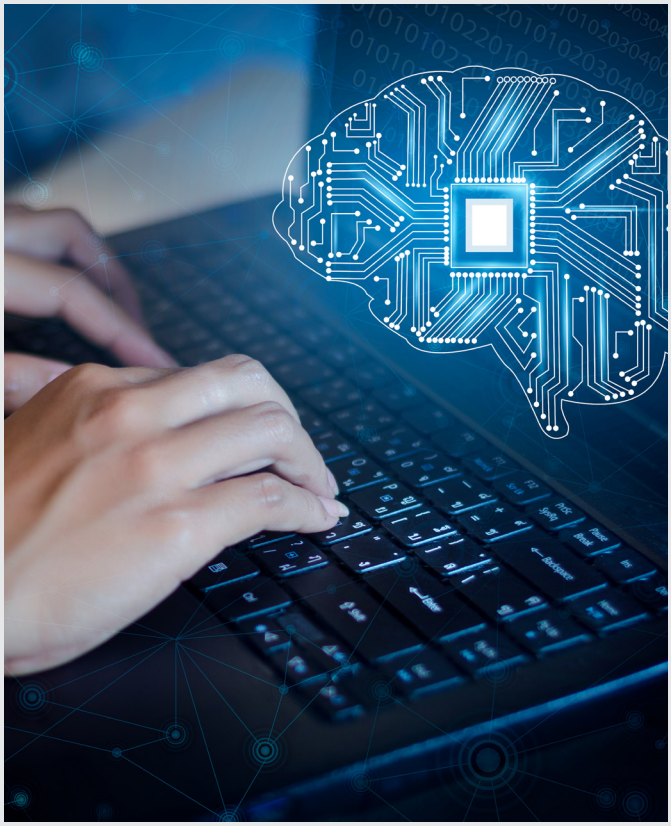
How Vision Agents Unlock Secure Automation for Enterprises App

The Pressure to Deliver on AI

AI has quickly moved from an experiment to an expectation.

Across industries, technology leaders are being asked to “bring AI into the business” to prove it can save time, reduce costs, and make teams more efficient. Boards want progress, employees want better tools, and customers expect faster results.

For IT leaders, that pressure feels familiar. It is the same challenge that has defined every technology shift: make it secure, make it work with what already exists, and make it measurable.



As a result, AI pilots stall before they scale.

Security teams raise concerns about governance and visibility. Operations leaders hesitate to introduce unproven tools or workflows that could disrupt stability or slow delivery. Developers struggle to integrate systems that were never designed to communicate with one another.

This paper explores a practical alternative. Instead of forcing automation through APIs or connectors, what if AI could work the way people do by seeing, understanding, and interacting with applications directly inside a secure environment?

That is the promise of Vision AI Agents. More importantly, it is a model that allows IT leaders to show measurable progress on AI while maintaining the same security, control, and compliance standards that already protect the enterprise.

The Enterprise AI Gap

The path from AI interest to AI impact is rarely straightforward.

Most organizations start with small proof-of-concept projects that automate limited tasks. A few scripts or bots demonstrate potential, but scaling those ideas across critical systems becomes a different story.

The reason is simple: enterprise software was designed for humans, not for agents.

Many of the applications that run supply chains, financial operations, or HR workflows have no APIs or inconsistent integrations. They were built for a person sitting at a screen, clicking, typing, and making decisions in real time.

When automation depends on APIs, those applications are left out. When it depends on RPA, teams inherit new layers of infrastructure to maintain, each with its own cost and fragility. When AI tools operate outside enterprise boundaries, risk and compliance quickly become the focus instead of innovation.

The challenge is now compounded by the rise of AI browsers and agentic interfaces that allow models to act on behalf of users inside web sessions.

Recent research from Brave, LayerX, and Palo Alto Networks (Brave Comet Prompt Injection; LayerX CometJacking Analysis; Palo Alto Networks: Agentic Browsers and Security Risks) has shown how these tools can be exploited through:

- Prompt injection attacks that alter or hijack agent instructions
- Unencrypted credential storage and session replay vulnerabilities
- Unauthorized access and lateral movement between browser sessions

For IT and security leaders, this reinforces an important truth: every time automation leaves the enterprise boundary, new risk follows.

Closing that gap requires a new approach, one that recognizes the realities of the enterprise environment and helps IT leaders bring automation to where the work already happens.

Why APIs and MCPs Are Not the Complete Answer

APIs were designed to make automation simple. Connect one system to another, move data between them, and let software handle the rest. For modern SaaS applications, that model works well. But most enterprise environments are not made up of SaaS tools alone. They include legacy systems, on-prem applications, and specialized tools that were never designed for external access or automation.

New frameworks such as Model Context Protocols (MCPs) are emerging to help AI systems exchange data more effectively. These standards allow models to pull in context from specific applications or databases, improving coordination across AI agents. MCPs represent meaningful progress, but they share the same underlying limitations as APIs.

For IT teams, the challenges remain:

- **Incomplete coverage.** APIs and MCPs rarely expose the full functionality of enterprise applications. Many workflows still require manual steps or workarounds.
- **Ongoing maintenance.** Vendor updates, schema changes, and broken dependencies force constant rebuilds and testing.

- **Security exposure.** Each integration adds more tokens, credentials, and permissions to manage. Every new connection expands the potential attack surface.
- **Limited reach.** Most legacy Windows and desktop-based applications still have no APIs or structured endpoints at all.

Meanwhile, new studies have shown that even browser-based AI agents can introduce vulnerabilities if deployed without proper isolation. TechCrunch and Brave both reported how browser-integrated AI tools were manipulated through prompt injection and credential exposure, compromising user sessions and sensitive data (TechCrunch: AI Browser Security Risks; Brave Comet Injection Report).

For IT leaders, the takeaway is clear: even as APIs and MCPs evolve, the enterprise still needs a secure environment where automation can safely operate across all applications, including those that lack structured interfaces.

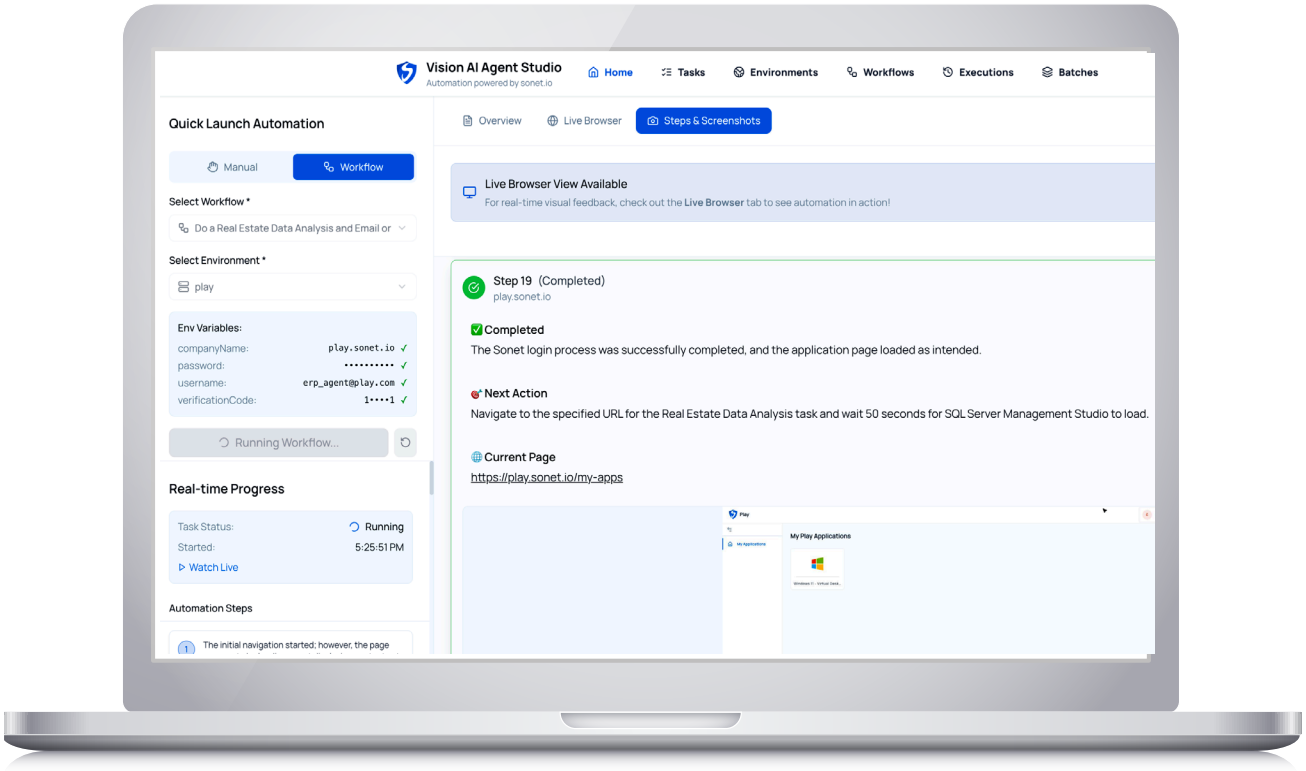
That is where Vision Agents begin to change the story.

A New Model: Vision AI Agents

Every technology shift brings a moment when the existing tools stop being enough. For enterprise automation, that moment has arrived. APIs and MCPs can connect data, but they cannot replace the way people use software. Most business workflows still depend on applications that require a person to sign in, navigate menus, and complete tasks that depend on judgment and context.

Vision Agents represent a new approach to automation that starts from that reality.

Rather than requiring applications to expose data through APIs or structured protocols, Vision Agents operate visually, just as a person would. They see



Vision AI Agent Studio

what appears on the screen, interpret it, and take action within the same interface that employees already use.

In many cases, they can also be taught the way work is already done. In Agent Studio, a subject matter expert can start recording, complete the task once, and then stop recording. This captures the on-screen steps and the decision points that matter. The system turns that walkthrough into a standard operating procedure (SOP) style set of instructions that becomes the agent's starting playbook. That reduces the need to handcraft "prompt knowledge" up front, or to ask experts to translate years of experience into perfect prompts. Teams start from the proven process, then refine it under governance by reviewing the steps, tightening instructions, and standardizing how the work should be done. As a side benefit, those recorded procedures help preserve institutional know-how about how things actually get done, including edge cases people tend to remember only when they are in the moment.

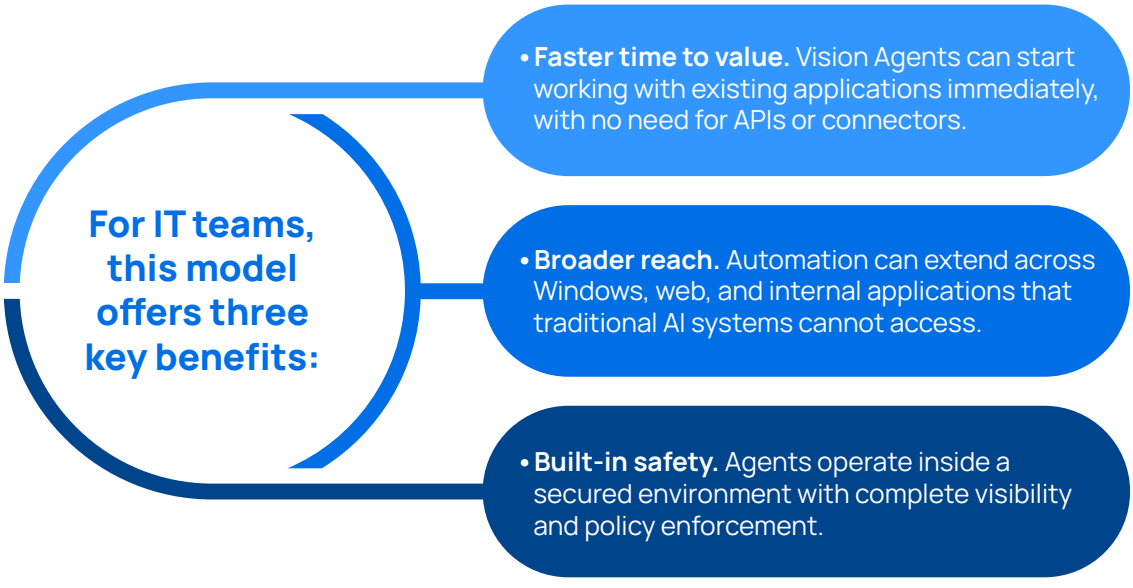
For IT leaders, this means automation can finally reach beyond modern SaaS tools and into the legacy and custom systems that keep the business running. A Vision Agent can open a Windows application, enter information, move between fields, and verify results all within a secure, monitored environment.

The advantage is not simply that Vision Agents can interact with more systems. It is that they can do so without rewriting code or changing infrastructure. They bridge the gap between today's applications and tomorrow's AI workflows by working where the data already lives.

When paired with a controlled browser workspace, Vision Agents can run safely inside enterprise boundaries. Policies for authentication, least privilege, and data protection can apply exactly as they do for human users. The result is automation that fits naturally within existing governance frameworks rather than requiring new ones.

For IT teams, this model offers three key benefits:

- **Faster time to value**. Vision Agents can start working with existing applications immediately, with no need for APIs or connectors.
- **Broader reach**. Automation can extend across Windows, web, and internal applications that traditional AI systems cannot access.



- **Built-in safety.** Vision Agents have no access to the Internet. They get access to only the secure virtualized apps based on their workflow needs.

Vision Agents do not replace people. They work alongside them, handling the repetitive or time-sensitive parts of a process while keeping IT in full control. For organizations struggling to turn AI ambition into measurable progress, they provide a practical way to start using the systems already in place.

Governance and Security: Keeping AI Under Control

Every conversation about AI in the enterprise eventually comes back to one question: **how do we keep it under control?**

Automation is powerful, but in regulated and security-conscious environments, it cannot come at the expense of visibility or compliance. For IT and security leaders, success depends on applying the same governance principles to AI agents that already protect human users.

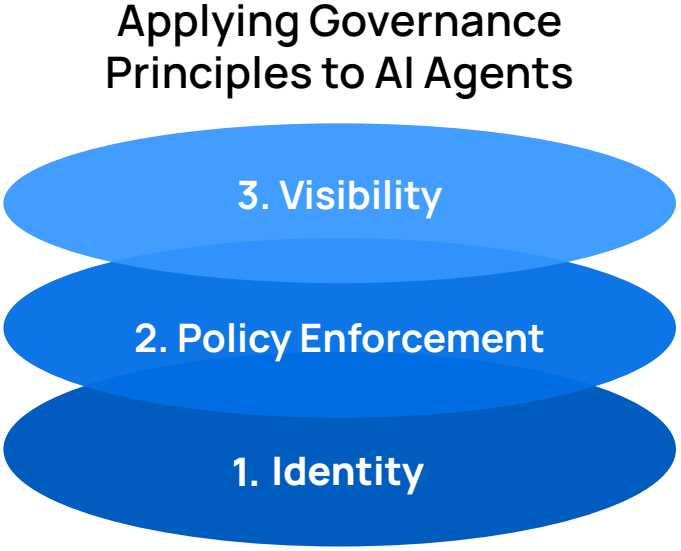
That starts with identity.

Each Vision Agent must operate as a defined entity with clear ownership, authentication, and privilege boundaries. An agent that can log into systems and manipulate data is, in effect, another user. It needs to be provisioned, monitored, and governed as one.

By assigning identity and access policies to every agent session, organizations can enforce least privilege and maintain accountability for every action taken.

The second layer is policy enforcement.

Whether the activity comes from a person or an agent, the same zero trust controls should apply.



This includes restrictions on data movement, clipboard use, downloads, and printing, as well as rules that govern which systems an agent can access. Policies must be consistent and observable so that no automation bypasses corporate safeguards.

Visibility is the third pillar.

Without full observability, even the best policies are incomplete. Every session, whether human or automated, should be recorded and logged. Real-time monitoring allows security teams to understand what occurred, when, and why.

This level of visibility goes beyond basic logging. It allows teams to review each step the agent took, in plain language, and understand the logic behind every action. That clarity is essential for diagnosing issues, validating behavior, and demonstrating to business stakeholders that AI-driven workflows remain compliant and predictable.

When automation operates inside a controlled browser workspace that is isolated from the public internet, these requirements become much easier to meet. A private execution environment prevents

untrusted web content from influencing agent behavior and keeps sensitive enterprise data from leaving the boundary. The browser then becomes a central enforcement layer where authentication, DLP, and session recording already exist.

By placing Vision Agents inside that workspace, IT leaders maintain full control without deploying new infrastructure or additional security products.

Strong governance does more than reduce risk. It builds trust in the system, enabling IT teams to expand automation confidently and demonstrate that AI can operate safely inside enterprise boundaries.

The Practical Path Forward

For most IT leaders, the biggest barrier to enterprise AI adoption is not vision but practicality. The opportunities are clear. The challenge lies in finding a secure, low-risk way to begin.

The good news is that bringing automation into production does not have to mean a full platform rebuild. By starting small and working within existing governance frameworks, teams can begin to demonstrate real value from AI in a matter of weeks.

Start with a focused use case

The first step is to identify a process that meets three criteria:

1. It relies on repetitive, screen-based work that consumes valuable time.
2. It touches systems that lack reliable APIs or are difficult to integrate.
3. It has clear, measurable outcomes such as time saved or errors reduced.

Examples might include onboarding contractors into multiple systems, extracting data from reports, or populating forms across different applications. These are the kinds of workflows where Vision Agents can make an immediate impact.

Capture the human baseline before you automate. Record the person who already performs the task end-to-end, especially the key decision points and common exceptions. That demonstration can be translated into an SOP-style set of instructions that becomes the agent's operating playbook, reducing prompt-heavy setup and ensuring the first automated workflow reflects how the work is actually done.

Integrate security from the start

Before any automation begins, establish policies that define how agents will operate.

- Authenticate agents through your existing identity provider.
- Limit permissions to only the systems and actions required.
- Enable session recording and activity logs to maintain visibility.

By treating each Vision Agent as a digital employee, you reinforce governance instead of creating exceptions.

Test, measure, and refine

Once an initial workflow is automated, evaluate both the outcomes and the oversight. How much time was saved? Were there any policy violations or visibility gaps?

The goal of the pilot is not just to prove that the agent can complete a task, but that it can do so securely and within your existing controls.

Scale intentionally

After validating early results, expand automation to additional workflows. Reuse the same security model, access policies, and observability framework. Each new deployment should be faster and safer than the last.

As adoption grows, the organization gains confidence that AI can operate safely at scale. This approach replaces the traditional "big-bang" transformation with a steady, measurable progression toward enterprise-wide automation.

The Practical Path Forward

Conclusion: A Smarter Way to Bring AI to Work

For IT leaders, the goal has never been simply to deploy new technology. The goal is to make work better, faster, and more secure. Artificial intelligence offers powerful ways to do that, but only when it can operate safely inside the systems the enterprise already depends on.

Vision Agents provide a practical way to bridge that gap. They work where the data lives, interacting with applications just as a person would. They follow the same security policies, run inside a controlled workspace, and produce the same visibility and audit data that leaders require.

This is not about replacing people or rewriting systems. It is about extending what already works. It is about giving IT teams a path to deliver measurable results from AI without adding risk or complexity.

By starting small, maintaining clear governance, and scaling with intention, technology leaders can move from proof-of-concept to production-level automation with confidence.

The future of enterprise AI will not be built on isolated experiments or ungoverned tools. It will be built by leaders who understand that innovation and control can coexist and who have the vision to bring AI to work, securely and responsibly.

sonet.io

Secure Workspaces for Humans and AI

sales@sonet.io

Find us online at **sonet.io**

3031 Tisch Way, 110 Plaza West
San Jose, CA 95128