Cloudflare Open-Sources Internal AI Workspace Used by Entire Workforce Since May—Agents Start with Zero Access, Gain Only Task-Specific Permissions

Cloudflare just released the AI workspace its 4,000+ employees have used for three months. The architecture inverts everything enterprises assume about agent security.

The News: Cloudflare Ships Its Own AI Infrastructure to the World

On August 4, 2026, Cloudflare announced Cloudflare OS—an open-source AI agent workspace released under Apache 2.0 license. The platform went live on GitHub immediately, with full source code available for any organization to deploy.

This isn't vaporware or a limited preview. Cloudflare's entire global workforce has been running this system internally since May 2026. The public release represents version 2.0, a complete rewrite built on six core Cloudflare primitives: Workers, Durable Objects, Dynamic Workers, Facets, AI Gateway, and Cloudflare Access.

The architecture rests on three pillars: an agent workspace where employees interact with AI, a security and governance framework called "Gatekeepers," and a platform for building persistent internal applications. Unlike typical AI assistants that produce one-off chat responses, agents in Cloudflare OS can generate apps and workflows that become permanent internal tools.

According to the technical blog post, the platform routes requests through Cloudflare AI Gateway, allowing organizations to use multiple model providers while maintaining cost visibility broken down by person, team, or application. Data stays within the organization's own Cloudflare environment rather than flowing through vendor-managed SaaS infrastructure.

A managed deployment option will arrive later through the Cloudflare dashboard. Partner implementations are available immediately through firms like Presidio and Happy Cog. No pricing has been announced for the managed version.

Why It Matters: The Procurement Nightmare Gets a Solution

Enterprise AI procurement in 2026 has crashed into a single, unavoidable question: what happens to company data when an agent touches it?

Microsoft Copilot embeds into the Microsoft 365 ecosystem. Glean indexes enterprise knowledge. Google's Duet AI lives in the Google Cloud. Each approach assumes the AI needs broad access to be useful. Security teams have spent the past 18 months trying to retrofit controls onto systems designed for maximum data availability.

Cloudflare flipped the model entirely. Agents in Cloudflare OS start with zero permissions—no credentials, no access to systems, no ability to read data. They gain only scoped, revocable permissions specific to each task, granted through the Gatekeepers

security layer.

This capability-based security model addresses the fundamental trust problem differently than any major enterprise AI platform. Traditional approaches grant access and then try to audit usage. Cloudflare OS grants nothing by default and requires explicit permission grants for every capability an agent needs.

The winners here are organizations with complex compliance requirements—financial services, healthcare, government contractors—that have watched AI adoption stall because security teams couldn't sign off on data exposure. They now have an open-source reference implementation they can audit line by line.

The losers are incumbent enterprise AI vendors who built their moats on managed data access. When a CTO can deploy an equivalent workspace on infrastructure they already control, with security guarantees they can verify themselves, the value proposition of black-box SaaS AI shifts dramatically.

Technical Deep Dive: How Capability-Based Security Actually Works

The Gatekeepers system implements capability-based security, a model that dates back to the 1970s but rarely appears in modern enterprise software. Understanding why Cloudflare chose this architecture requires examining what makes agent security fundamentally different from human user security.

The Problem with Traditional Agent Permissions

Human users authenticate once and receive role-based permissions. An analyst might have read access to financial reports, write access to their department's folders, and no access to HR systems. The assumption: the human makes reasonable decisions about how to use their access.

AI agents break this assumption completely. An agent doesn't "decide" to stay within appropriate boundaries. It executes whatever instructions it receives, potentially chaining together capabilities in ways no human anticipated. Grant an agent access to email and

calendar "to schedule meetings," and you've also granted it access to read every email thread and every calendar event.

The capability-based approach inverts this entirely. Instead of granting agents permissions they might need, Cloudflare OS requires explicit capability tokens for every action.

Architecture of the Gatekeepers Layer

When an agent in Cloudflare OS needs to perform an action—say, reading a specific Slack channel to summarize recent discussions—it doesn't authenticate as a service account with Slack access. Instead, the workflow looks like this:

- **Request phase:** The agent declares what capability it needs: "Read messages from #engineering-standup from the past 7 days."
- **Gatekeeper evaluation:** The Gatekeepers layer checks whether the requesting user has permission to grant this capability, whether the scope is acceptable under organizational policy, and whether any audit requirements apply.
- **Capability token issuance:** If approved, the agent receives a scoped, time-limited token that grants exactly the requested access—nothing more.
- **Execution:** The agent uses the token to perform the specific action.
- **Revocation:** The token expires after use or after a short time window, whichever comes first.

The critical difference: the agent never "has access" to Slack. It receives permission to perform one specific read operation, and that permission evaporates immediately after use.

Building on Cloudflare Primitives

The six core primitives—Workers, Durable Objects, Dynamic Workers, Facets, AI Gateway, and Cloudflare Access—each serve specific architectural purposes:

Workers handle stateless compute at the edge. Every agent interaction runs as a Worker, meaning compute happens close to users with millisecond cold-start times.

Durable Objects provide strongly consistent state. When an agent builds a persistent

application—say, a dashboard that tracks project status—the application state lives in Durable Objects with guaranteed consistency across regions.

Dynamic Workers enable runtime code generation. Agents can create new Workers on the fly, allowing them to build and deploy tools without human intervention for each deployment.

Facets (a newer Cloudflare primitive) provide identity and capability management. The Gatekeepers layer builds directly on Facets to manage capability tokens.

AI Gateway routes requests to model providers. Organizations can use Claude, GPT-4, Gemini, or any other provider through a unified interface, with full logging and cost tracking.

Cloudflare Access handles human authentication. Users authenticate through Access, which then determines what capabilities they can grant to agents.

The Persistence Layer: Apps, Not Just Answers

Most AI assistants produce ephemeral outputs. Ask a question, get an answer, the interaction ends. Cloudflare OS treats agent outputs as potentially persistent artifacts.

When an agent creates a workflow—say, a process for onboarding new vendors—that workflow becomes a deployable application within the platform. Other employees can invoke it without understanding the underlying implementation. The agent has effectively programmed a new capability into the organization's toolkit.

This creates a flywheel effect. As agents build tools, those tools become available for other agents to use. A vendor onboarding workflow might call a contract review tool that another agent built last month. Organizations accumulate capability over time rather than repeating similar agent interactions.

The Contrarian Take: What Most Coverage Gets Wrong

Early coverage of Cloudflare OS has focused on the open-source angle—Cloudflare giving away valuable IP to compete with Microsoft and Google. This framing misses the strategic reality.

This Isn't Altruism, It's Infrastructure Lock-In

Cloudflare OS runs on Cloudflare infrastructure. The Workers runtime, Durable Objects, AI Gateway, and Access are all Cloudflare services. Open-sourcing the application layer makes perfect sense when the infrastructure layer generates revenue.

Every organization that deploys Cloudflare OS becomes a Cloudflare infrastructure customer. The managed version, when it arrives, will presumably carry a premium. But even organizations running the open-source version directly will pay Cloudflare for Workers invocations, Durable Objects storage, and AI Gateway requests.

This is the same playbook MongoDB and Elastic ran for years: open-source the database, monetize the managed service. Cloudflare just applied it to AI agent infrastructure.

The Security Model Has Real Limitations

Capability-based security is genuinely more secure than role-based access control for AI agents. That doesn't mean it solves every problem.

Capability accumulation: If an agent requests 50 capabilities to complete a complex task, the user must approve 50 requests—or the system needs a way to batch approvals, which reintroduces some of the risks the model was designed to avoid.

Indirect information flows: An agent with capability A and capability B might be able to infer information that neither capability alone would reveal. The Gatekeepers layer doesn't (and can't) analyze these compositional risks.

Prompt injection remains unsolved: If an agent reads content that contains adversarial instructions, the capability-based model doesn't prevent the agent from being manipulated. It limits what the manipulated agent can do, but that's a containment strategy, not a solution.

Implementation complexity: Organizations with hundreds of systems need to build capability interfaces for every system an agent might access. The open-source release provides the framework but not the connectors. Integration work will be substantial.

What's Actually Underhyped

The most significant aspect of this release isn't the security model—it's the persistence layer.

Enterprise AI has been stuck in "chat interface" mode. Users ask questions and receive answers. The next interaction starts fresh. This creates two problems: useful outputs disappear, and organizations don't accumulate intelligence over time.

Cloudflare OS treats agent outputs as deployable artifacts. A well-crafted workflow becomes institutional knowledge that other employees can invoke without understanding the underlying AI interaction that created it. Over time, an organization builds a library of agent-created tools.

This is closer to traditional software development than to chat-based AI assistance. The agent acts as a developer who can ship code, not just a consultant who provides advice.

If this model gains adoption, it changes how organizations think about AI investment. The question shifts from "how many AI seats do we need?" to "what capabilities have we built?" The value accumulates in the capability library, not in individual user licenses.

Practical Implications: What Technical Leaders Should Do Now

For Organizations Evaluating Enterprise AI

Cloudflare OS creates a new reference point for enterprise AI procurement. Before signing an agreement with any vendor, technical teams should now ask:

- **What's the default permission state?** Does the agent start with access, or does it request capabilities per task?
- **Can we audit capability grants?** Does the platform log what permissions were granted, when, by whom, and for what purpose?
- **Where does data live?** Is data processed in vendor infrastructure, or can it stay

within environments we control?
- **What persists after interactions?** Are useful outputs ephemeral, or can they become organizational tools?

These questions had no good answers six months ago. Now they do. Vendors that can't match Cloudflare's security model will need to explain why their approach is acceptable.

For Teams Considering Deployment

The open-source release is available immediately, but production deployment requires significant integration work. Before committing engineering resources, evaluate:

Cloudflare infrastructure dependency: If your organization isn't already on Cloudflare, you're adopting a new infrastructure vendor along with the AI platform. Calculate total cost of ownership including Workers, Durable Objects, and AI Gateway charges.

Connector availability: The base platform doesn't include connectors for enterprise systems. Check whether your critical integrations (Salesforce, ServiceNow, SAP, etc.) have community-built connectors or require custom development.

Team capability: Managing a capability-based security system requires different skills than managing role-based access control. Security teams need training on how to define appropriate capability scopes and review capability policies.

Model selection strategy: The AI Gateway supports multiple providers, but that flexibility requires decisions. Which models for which use cases? What's the fallback if a provider has an outage? Who manages cost allocation?

Start with a limited pilot in a single department before organization-wide deployment. The permission model requires careful policy definition; getting it wrong creates either security gaps or usability friction.

Code and Architecture to Explore

The GitHub repository contains the full platform implementation. Priority areas for technical evaluation:

- **/gatekeepers:** Core capability-based security implementation. Study how capability tokens are issued, scoped, and revoked.
- **/workflows:** The persistence layer that turns agent outputs into deployable applications. This is the most novel architectural element.
- **/gateway:** AI Gateway integration showing how multi-model routing works. Look at the cost tracking implementation for budget management patterns.

Teams building similar systems on other infrastructure can treat these implementations as reference architectures even if they don't adopt the platform directly.

Forward Look: Where This Leads in 12 Months

Enterprise AI Splits into Two Markets

By mid-2027, enterprise AI will have clearly bifurcated.

The first market: organizations that accept vendor-managed AI with traditional permission models. They'll use Copilot, Glean, and similar products because the deployment simplicity outweighs security concerns. These tend to be smaller organizations, those in less regulated industries, and those with limited security engineering capacity.

The second market: organizations that require capability-based security and infrastructure control. Financial services, healthcare, defense contractors, and any company with meaningful IP protection needs will gravitate toward self-managed or partially-managed solutions. Cloudflare OS positions itself as the default choice for this market.

The Capability Library Becomes an Asset

Organizations running Cloudflare OS for 12+ months will have accumulated substantial libraries of agent-created tools. These libraries represent genuine competitive advantage—institutional knowledge encoded as deployable capabilities.

Expect to see early movers talk about their capability libraries the way companies currently talk about data assets. "We have 400 operational workflows our agents built" becomes a metric in investor presentations and acquisition discussions.

Some organizations will explore licensing capability libraries to others, creating a new category of intellectual property. The legal frameworks for this don't exist yet.

The Connector Ecosystem Matters More Than the Platform

Cloudflare shipped the platform; the community will determine its success by shipping connectors. A capability-based security model is only valuable if it can govern access to systems that matter.

The first enterprises to build production-quality connectors for SAP, Salesforce, Workday, and ServiceNow will shape the platform's trajectory. If these connectors don't materialize within 12 months, Cloudflare OS remains an interesting architectural experiment rather than a practical deployment option for most enterprises.

Watch for announcements from the integration partners (Presidio, Happy Cog) about which connectors they're prioritizing. That signaling will indicate where early enterprise demand is strongest.

Managed Version Pricing Sets the Market

Cloudflare hasn't announced pricing for the managed version. When they do, it establishes a reference point for the entire enterprise AI market.

If the managed version prices below current Copilot and Glean contracts while offering stronger security guarantees, incumbent vendors face immediate pricing pressure. If it prices higher, positioning as a premium security offering, it creates space for a third tier of competitors between the "security-first" and "convenience-first" segments.

The pricing decision will reveal whether Cloudflare sees this as an infrastructure play (price low to drive adoption and infrastructure revenue) or a standalone product play (price for margin on the managed service itself). The former seems more likely given the open-source strategy.

The Real Stakes: Who Controls Enterprise AI Infrastructure

This release is the most significant enterprise AI infrastructure announcement since Microsoft integrated Copilot into Office 365. Not because the technology is more advanced—it isn't—but because it changes the competitive structure of the market.

Microsoft and Google have built enterprise AI assuming centralized data access and vendor-managed infrastructure. That model works when organizations trust the vendor or lack alternatives. Cloudflare OS provides an alternative.

The capability-based security model addresses real concerns that have stalled enterprise AI adoption. Many CISOs and CIOs have refused to approve broad AI deployments because they couldn't explain to regulators or boards exactly what data AI agents could access. "Zero default permissions with scoped capability grants" is an answer that security teams can defend.

More importantly, the open-source release means this architecture is now available for anyone to examine, modify, and improve. If capability-based security becomes the expected standard for enterprise AI agents, every major vendor will need to implement something similar—and Cloudflare will have set the benchmark.

The three-month internal deployment provides Cloudflare with credibility that other open-source AI infrastructure projects lack. This isn't a proof of concept or a minimum viable product. It's a system that a global technology company bet its own operations on.

For technical leaders making enterprise AI decisions in the next 12 months, Cloudflare OS isn't necessarily the answer—but it's now the question that every other vendor must answer.

WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

Book a meeting □