KAMI WAZA

Kamiwaza: Bringing intelligence to your data — wherever it lives

kamiwaza.ai

2025

KAMI WAZA

Kamiwaza: Bringing intelligence to your data — wherever it lives

Today, businesses are under intense pressure to deliver results. (And fast.) Enter AI: your companion in efficiency.

But the race to adopt AI has created more noise than clarity. Leaders are stuck navigating a maze of tools — and the path forward is blocked by three significant challenges:

- **Complexity.** The AI landscape is fragmented across dozens of models, frameworks, and integration approaches. Teams spend months evaluating LLMs, vector databases, orchestration tools, and deployment stacks instead of solving business problems.
- **Data fragmentation.** Data lives across clouds and on-prem systems. Centralizing it for AI is costly, slow, and (for some businesses) impossible.
- **Security.** Compliance and data governance are non-negotiable. And high-profile AI missteps have put even well-intentioned initiatives under scrutiny.

Kamiwaza cuts through AI complexity by providing a complete, pre-integrated AI stack with proven technical choices already made. Our lightweight, Docker-based architecture runs securely within your existing infrastructure, bringing AI to your data. So you can avoid costly centralization and stay in control.

And because it's compatible with open-source models, you're free to choose the fastest, smartest, or most secure option for you. No vendor lock-in, no compromises.

Orchestrating AI across your enterprise.

Kamiwaza transforms enterprise AI adoption by orchestrating intelligence across your existing infrastructure, bringing AI to your data — wherever it lives. No forced, potentially impossible data, centralization.

Simplify & accelerate AI adoption.

Deploying Kamiwaza as your open-source AI orchestration solution dramatically reduces the complexity and risk typically associated with enterprise-scale AI initiatives. Our Docker-based deployment model allows you to implement AI with minimal disruption — and using your existing infrastructure.

A typical Kamiwaza implementation can be operational within days, not months, allowing you to realize ROI immediately, while maintaining complete control over your data and infrastructure.

KAMI WAZA

Break free from vendor lock-in.

Its vendor-neutral architecture prevents lock-in, allowing you to leverage AI without dependency on a single provider. This gives you the freedom to choose the best tools for each specific use case without architectural constraints.

Future-proof your AI investment.

As new models and capabilities emerge, Kamiwaza's modular design lets you incorporate these advancements without rewriting applications or replacing infrastructure. This future-proofing lets your AI investments remain relevant even as tech changes, giving you both immediate results and long-term strategic flexibility.

Features & benefits.

Feature	Description	Outcomes
Hardware-agnostic	Integrates with NVIDIA, Intel, AMD, Ampere, and other accelerators across cloud, on-premises, and edge environments	Use existing hardware investments and avoid vendor lock-in
Distributed data engine	Maintains a global catalog and metadata layer that identifies data locality and routes inference requests to nodes positioned alongside your data	Ensures complete data isolation between security domains while supporting cross-repository insights
Model context protocol	Normalized across major foundation models	Prevents dependency on any single AI vendor, allowing model switching without rewriting applications
Docker-based deployment	Packaged as lightweight containers that run on existing infrastructure	Rapid implementation with minimal disruption — and operational within days, not months

Enterprise use cases.

- **Agentic AI.** Deploy AI agents that navigate systems and complete multi-step workflows in minutes.
- **Advanced analytics.** Analyze data across secure boundaries in real time, so you get fast, accurate insights without risking compliance.
- **Workflow automation.** Replace clunky manual tasks with automation that adapts to your processes and frees up your team's time.

- **Knowledge discovery.** Build a connected view of your data so you can uncover insights and answer complex questions across departments.
- **Document processing**. Automatically extract important info from reports, emails, and contracts to speed up responses and reduce manual work.

Organizations getting real results with Kamiwaza.

Organization	Challenge	Solution	Result
U.S. Department of Homeland Security	Predict weather impacts on critical infrastructure by analyzing 97 years of data (33,000 files) across security boundaries	Processed data where it resided using distributed intelligence engine	• Completed in one week, saving two to three years of manual effort • Maintained all security classifications • Enabled predictive emergency response planning
Campbell's	Manual processing of broker rebate invoices requiring staff to transcribe PDFs into spreadsheets	Automated document extraction, validation, and processing workflow	• 87% reduction in processing time • Elimination manual errors • Staff redeployed to higher-value activities

Kamiwaza pricing.

The value you get from an AI-powered data orchestration tool depends on how you put it to work. And pricing should match that flexibility. That's why our plans are built to support everything from solo builders to businesses at scale.

Edition	Price	Nodes	Outcomes
Community	Free	1 node	Self-service
Flex	$25,000/year	3 nodes (1 CPU socket or 1 GPU < 128 GB VRAM)* *Add more nodes for $3,500/year per node	1/year
Starter	$75,000/year	Up to 3 mixed CPU/GPU nodes (500 GB VRAM per node)	4/year (quarterly)
Enterprise	$125,000/year	3 mixed nodes (no VRAM cap)* *Add more nodes for $25,000/year per node	12/year (monthly)

KAMI WAZA

We deliver outcomes.

We focus on guaranteed results — not just traditional software licensing. Each paid subscription includes a specific number of outcome-based projects where we work directly with you to set, and meet, outcomes. Each outcome includes solution design, guided implementation, and optimization until measurable business value is achieved.

Technical specs.

Category	Outcomes
Technical requirements	• **Minimum hardware:** Community Edition runs on a single GPU/node, while enterprise deployments support diverse hardware (Intel, AMD, Ampere, Nvidia, Qualcomm) • **Supported OS:** Windows and Linux • **Environment:** Enterprise editions work across cloud, on-premises, and edge • **Network requirements:** Hybrid deployments require connectivity for distributed inference mesh and locality-aware data processing
Security & compliance	• **Encryption:** Data protected via stringent security protocols • **Compliance:** Supports GDPR and sector-specific frameworks • **Architecture:** Identity management (SAML 2.0, Active Directory), data lineage tracking, and enterprise-grade stability
Integration capabilities	• **APIs/connectors:** REST API, SDK, OpenAI-compatible endpoints, and integrations with Milvus, DataHub, Hugging Face • **Enterprise systems:** SAP, ServiceNow, Workday, CRMs, ERPs • **Authentication:** SAML 2.0, Active Directory, API keys
Performance benchmarks	• **Throughout:** 8,000 tokens/sec for a 70B-parameter model on an 8-way server
Implementation timeline	• **Setup:** Community edition installs locally in minutes, Enterprise (paid) deployments involve cluster configuration • **Integration:** Data pipeline setup (1–2 weeks) using prebuilt connectors • **Deployment:** Model tuning and workflow automation (2–4 weeks) • **Optimization:** Continuous performance tracking (2-4 weeks) • **Training:** Uses Jupyter notebooks, SDK examples, and Discord community support

KAMI WAZA

Kamiwaza AI model support.

Capability	Details
Model Backend	• llama.cpp • vLLM • Hugging Face • Private/enterprise models via Model Context Protocol (MCP)
Model source	• Hugging Face • Private/enterprise models via Model Context Protocol (MCP)
Example supported models	• Llama family models • Qwen models (including Qwen2.5-7B-Instruct) • Other open-source LLMs compatible with llama.cpp/vLLM
API compatibility	• OpenAI-compatible endpoints • Native SDK and CLI tools
Private/custom model support	• Yes, via MCP • Tools for model search, download, deployment, and evaluation • Support for both general-purpose and customized AI workflows

Next steps.

We know you need AI results fast. That's why our Kamiwaza Quick Start delivers measurable impact within 30 days using your existing setup. Our team will partner with yours to find a high-value use case and deploy Kamiwaza securely — no data migration needed.

Let's schedule a 45-minute workshop next week to review your priorities and see how Kamiwaza's distributed intelligence can help. Once you experience Kamiwaza, you'll see why top organizations rely on us for AI without limits.

Looking to see Kamiwaza in action first-hand?

- Book a live demo featuring our unique MCP-enabled browser automation, or explore a proof-of-concept with guaranteed outcomes. Reach out: hello@kamiwaza.ai
- Try our Community edition for developers