

Krista LLM Access



Enterprise-governed access to every LLM, with control over models, prompts, cost, and data in one place.

The Problem

Employees did not wait for permission. By the time a company has 500 or 1,000 people, its workforce is already prompting ChatGPT, Claude, and Gemini every day. Some use personal accounts. Some expense a subscription and label it as something else. Many paste company information into models no one in IT has reviewed. The capability arrived faster than the controls.

Two things happen in nearly every enterprise at once. Cost explodes with no clear return to justify it. And the security team has no way to see what is happening, so the risk keeps growing. A corporate license to one public model does not fix this. It standardizes a single tool. It does not route across models, govern prompts, manage budgets, or produce an audit trail, and it does not stop employees from using the other tools they already prefer.

Without a governed front door, leadership cannot answer three basic questions:

1

Which models are employees using, and for what?

2

What is each model costing, by person and by department?

3

Has sensitive data left the building, and into whose model?


User A
Strategy

Compare our top 3 targets, recommend which to pursue, with the key risks.


User B
AP clerk

I need budget-friendly gift ideas for our upcoming Holiday Party.


User C
Sales Rep

Our widget costs \$3.00 to make. How much should we charge?



Frontier · gpt-5.4

Multi-step reasoning. For the ~20% that truly needs it.

Krista LLM · Private

Extraction & formatting. The ~80% routine work. Contained & private.

BLOCKED

"No confidential information prompts."

Krista monitors every prompt and routes to optimize speed, quality, and cost — so routine work never pays frontier prices. And keeps your data privacy policies enforced.

 Reasoning → Frontier  Routine → Krista LLM  Blocked by Rule

The Solution

Krista LLM Access gives an enterprise one governed place for employees to use LLMs. It blocks direct access to public models and gives employees a governed prompt surface inside the firewall instead. Behind that surface, Krista reads each prompt, applies the company's policies, routes the work to the right model, and records everything.

Krista LLM Access connects to a wide and growing range of models across families: open-source models, Krista's own model, models the customer hosts on its own equipment, and public or frontier models from OpenAI, Anthropic, and Google. The employee still gets their answer. The company gets cost control, data control, and a complete record of activity.

Krista reads every prompt and routes it to the model that best fits the work on speed, accuracy, cost, and safety. The employee never picks the model. Routine work runs on Krista's own model. Only the work that truly needs a frontier model gets one.



Why It Matters

No executive is going to pull LLM access off everyone's desktop. The pattern repeats the early fight over personal phones, which IT swore would never touch the network and then had to secure instead of block. The job is not to block access. The job is to govern it.

The cost case is now urgent. Through 2026 the industry named the behavior driving the bills: defaulting to the most expensive model for every task with no routing, budget, or visibility. On routine work, Krista's own model delivers roughly 94 to 98% of premium quality at a fraction of the price, up to 4 to 6 times faster. The per-token difference is stark: Krista's model runs at \$0.30 per million tokens against a frontier model's \$15.00, roughly 50 times cheaper on the work that makes up most of the load.



Note: Routing is automatic, and the safety rule is also a control. It limits data leakage and keeps off-topic personal prompts, such as an employee checking FIFA scores, off the company's expensive models.

Key Capabilities



Multi-Source Model Access

Connects to a wide and growing range of models across families: open-source, Krista's own model, customer-hosted, and public or frontier models. Not a fixed list. New models are added as they prove out, and administrators decide which roles can reach which models.



Intelligent Routing

Krista reads each prompt and routes it to the best-fit model on speed, accuracy, cost, and safety. The employee never picks the model. Routine, high-volume work runs on Krista's own model. The roughly 20% of work that needs absolute fidelity or specialized reasoning is reserved for premium models.



Cost Controls and Budgets

Budgets are attached to roles and departments. As a budget fills, Krista LLM Access warns and applies graceful model tiering, moving work to less expensive models rather than enforcing a hard cutoff.



Prompt Guardrails

Policies act on every prompt in flight, scoped by role. Block a forbidden topic so it never leaves the building. Warn on a sensitive area before it sends. Redact a sensitive value so the model processes the request but never sees the real data.



Data Protection and DLP

Sensitive values are redacted before a prompt leaves, and controls are destination-aware. Public models can be locked down hard while private and self-hosted models stay open, because data sent to a contained model never leaves the instance.



Monitoring, Audit, and Reporting

Records every request, the policy decisions applied, the model used, tokens consumed, and cost. Reports break spend down by user, by role, and by department. The audit log is immutable.



Role-Based Access and Model Portfolios

Administrators build model portfolios and assign them to roles, then attach guardrails and budgets. A research role can get looser guardrails, a broader portfolio, and a higher budget. An entry-level role gets a portfolio of inexpensive models and a cap. Access maps to job and clearance, and routing, guardrails, budgets, and audit all key off role.

Who This Is For

Security and Compliance Leaders

- Need every LLM interaction governed, recorded, and auditable
- Want sensitive data redacted before it reaches a public model
- Require contained, private inference for regulated and confidential work

Finance Leaders

- Can see the AI bill but cannot tie it to value
- Want spend tracked and enforced by user, role, and department
- Need routine work on a \$0.30 model, not a \$15 one

IT and Platform Leaders

- Own the platform but have no visibility into LLM use
- Want one governed place to reach every model instead of scattered accounts
- Need role-based policy, model portfolios, and a complete record of activity

Enterprises Managing LLM Adoption

- Already living with shadow AI on personal accounts
- Want to govern access, not ban it, the way they secured personal phones
- Need cost control and data control before scaling AI across the business



NEXT STEP

Talk to Krisha about deploying Krisha LLM Access in your environment.

krisha.ai