Deterministic behavioral control, inside the model.

PRIOR is a self-hosted control plane for the open-weight LLMs you run yourself. It reads the intent behind each request, decides whether it's within your policy, and when it isn't, **holds the model to your rules from inside**, steering it back on course or refusing outright, before a wrong token reaches a user. **No second LLM, nothing added to your prompts, no telemetry.**

WHAT IT DOES

- **Enforces policy from the inside.** A tri-state gate routes every request green (answer), yellow (confirm), or red (refuse) against packs you author from plain-language docs.
- **Reads intent, not keywords.** Survives jailbreaks, personas, encodings, and injection hidden in retrieved content.
- **Grounds facts the model can't ignore**: in latent state, zero prompt tokens, overriding stale or tampered sources.

HOW IT DEPLOYS

- Headless engine container, **OpenAI-compatible** `/v1` API + a separate optional **console** container (web dashboard).
- Compiled **Rust** over native **llama.cpp**; no Python in the serving path. Runs GGUF.
- Runs in your **VPC or on-prem**, node-locked and **self-hosted**; your data never leaves it. Air-gapped tier available.
- **Calibrated builds across every major open-weight family** (Llama, Qwen, Gemma, Phi, Mistral, DeepSeek), up to 31B; steering runs on dense, **MoE** and hybrid state-space architectures.

MEASURED RESULTS

RESULT	NUMBER
Flagged harmful requests refused	**100% · 79/79**
StrongREJECT harm, Mistral-Nemo-12B	**0.60 → 0.084 (-86%)**
Disguised jailbreaks recovered	**96%**
Generalizes to unseen attacks	**AUC ~0.93**
Capability tax (math/knowledge)	**Δ 0.0**
Asserted a planted false source (worst model, unprotected)	**45%**
Regression to a false source	**eliminated → 0%**
Added latency (idle / steering)	**None / negligible**

1B to 14B fleet pass. Independent cais/HarmBench judge, 11 models · StrongREJECT scored by its own classifier on Mistral-Nemo-12B · GSM8K/MMLU byte-identical, 12 models · knowledge-conflict on a 100-item web-verified corpus, n=15 to 51 per model, worst was Mistral-Nemo-12B at 45% and the axis is model deference, not size; a pinned value is a compliance guarantee, not an accuracy gain · cross-check 93.6% (n=800) as the conservative floor · across-turn composition above 8B in progress.

INJECTION SURVIVAL (NEW IN 1.1.0)

DEFENSE, UNDER A ONE-LINE INJECTION	REFUSAL RATE
Safety system prompt, unattacked	**64%**
Safety system prompt, injected	**8%**
PRIOR, injected	**48%**
Benign output left byte-identical	**236/236**

Separate 4B-class pass, not combined with the fleet numbers above. StrongREJECT, n=50, judged by StrongREJECT's own fine-tuned classifier, on Qwen3-4B · on abliterated Qwen3-4B-heretic the system prompt scores 0% unattacked while PRIOR holds 44% · benign retention on XSTest-safe, 236 of 250 routed GREEN, byte-identical on Qwen3-4B-heretic, Qwen3-4B and Llama-3.2-3B · the 48% shortfall is detector recall, not steering.

PRICING

Licensed per node, annually. **Single Node $995** · **Team (5) $2,450** · **Platform (12) $4,950**. 30-day free trial, full capabilities, no card. Enterprise & air-gapped on request.