AI & ORCHESTRATION

ModelNative

Multi-phase model planning, sharding, memory fit, calibrated latency/bandwidth support, SLO and cost/energy policy.

IP-4888 · Commercial software · CLI / SDK / API

Function and architecture

- Model↔Hardware Planner
- Independent installation
- CLI, Python SDK and server surfaces
- Tenant, entitlement and evidence controls

Integration summary

ModelNative: Multi-phase model planning, sharding, memory fit, calibrated latency/bandwidth support, SLO and cost/energy policy. Delivered as CLI / SDK / API. Integrate the selected version at its documented input/output boundary; retain source and configuration identity with each result. Use the executable quickstart and bounded server adapter linked on this page.

Technical record

Canonical identity	IP-4888
Responsibility	Multi-phase model planning, sharding, memory fit, calibrated latency/bandwidth support, SLO and cost/energy policy.
Delivery boundary	CLI / SDK / API
Evidence scope	Commercial software
Engineering boundary	Multi-phase model planning, sharding, memory fit, calibrated latency/bandwidth support, SLO and cost/energy policy.
Deployment	CLI / SDK / API; 1.2.0
Operational scope	Customer-specific performance, economics and physical effects require customer data and validation. Optional PostgreSQL live integration was not executed in the recorded release.

Delivery and release binding

- Versioned package and dependency lock
- Deployment environment and access controls
- Input/output schemas and integration examples

• Evaluation scope, support and update policy

Confirm for the selected release: Environment and dependencies, Version / build hash, API compatibility, License and support scope.

Evidence and qualification

Our records show product-specific tests within the 280-test commercial suite, 20,000 randomized cases and HTTP/ASGI/import/install campaigns. Source-bound external exercises are documented.

Customer-specific performance, economics and physical effects require customer data and validation. Optional PostgreSQL live integration was not executed in the recorded release.

Canonical status: P2 EXECUTABLE COMMERCIAL SOFTWARE / TEST SUITE + ADVANCED CAMPAIGN + HTTP/ASGI/IMPORT/INSTALL EVIDENCE; Immutable public model shape + synthetic hardware inventory: legal multi-engine placement where single-engine baseline had none.

Evaluation and licensing

Evaluation: Evaluate the named software release. Production: Team or enterprise deployment. Custom work: Workflow and API integration. Rights, pricing, support and acceptance criteria are agreed for the exact scope.

Contact: patrick@garmolabs.com | Product path: /software/modelnative