OntoGuard — Decision Authorization Infrastructure for Autonomous AI

What it is (in one sentence)

OntoGuard is a **decision authorization layer** that sits **above your AI model and data** and determines whether an AI output is allowed to become a real-world action—or must be **blocked** or **escalated to a human**—with proof attached.

The problem it solves (why executives care)

Companies want AI to drive high-stakes workflows—approvals, filings, healthcare operations, security actions, customer exceptions—but raw model output isn't "business-admissible" because it can be wrong, unauditable, and risky.

Once AI output can trigger action, you need an **authorization + proof layer** so someone can say: **"This was allowed, here's why, and here's the evidence."** OntoGuard is built to be that layer.

What OntoGuard does:

Think of the model as a brilliant fast talker.

OntoGuard is the **adult in the room**:

- Treats the model's output as a **claim, not the truth**
- Grounds it to structured enterprise reality (**objects + relationships**) and builds a **provenance-rich symbolic trace (L1)**
- Cross-checks with **semantic consensus (L2)** and then applies **alignment feedback + arbitration (L3)** to resolve conflicts and drive a single decision
- When confidence is low, it **auto-abstains / routes to review**; when confidence is high, it produces **governance-ready artifacts** that satisfy legal, risk, audit, and ops

The 3 deliverables executives actually get (every step)

1. **Decision API (the product): ALLOW / BLOCK / ESCALATE**, plus confidence/uncertainty gating and safe abstention (no silent failures).
2. **Evidence Pack (the moat):** sources + hashes + policy snapshot + replayable audit context ("receipts" for the decision).
3. **Governance Report (the enterprise artifact): schema-stable JSON + PDF** with trust/risk/uncertainty, clause-level citations, fallback badges, and telemetry.

How it plugs in (enterprise reality)

OntoGuard is designed to integrate into existing stacks and run at the **decision boundary (output → action)**:

- **Inline gateway (synchronous checkpoint):** best for high-stakes flows where you must block/escalate before action
- **Async auditor (post-run governance):** best for fast pilots; still generates evidence packs + reports

And it's **LLM-agnostic and data-agnostic**: it works with your chosen model and sources—**you can swap models without rebuilding governance** (not "multiple LLMs at once" as the core promise).

It also makes AI systems improve over time (training signals)

OntoGuard doesn't just police outputs. It can export **training signals** from governed decisions:

- What the AI/agent tried to do (intent/tool call)
- What evidence was missing/conflicting
- What policy/rule triggered the outcome
- What the safe/correct action should have been

This creates "clean examples" to improve RAG/agents faster—without relying on noisy internet data.

The “Peak Data” angle (why it can be massively profitable)

Models go stale because new facts, policies, and edge cases (“Peak Data”) keep arriving—traditional fixes require expensive retraining and re-validation cycles.

OntoGuard is positioned as a **runtime governed mitigation layer**: update knowledge + constraints without touching base weights, enforce policy at runtime, and ship proof artifacts for audit/compliance—turning drift into a manageable (and monetizable) maintenance surface.

What you license (clean expansion path)

The deck’s licensing ladder:

1. **Integration Packs**
2. **Agent Control Plane**
3. **Proof Packs**
4. **Training Signal Export**
5. **Platform Expansion**

The simplest way to say it

LLMs generate intelligence. OntoGuard turns that intelligence into authorized, auditable actions—with receipts.