OFFICE TIMESHEETS

AI Telemetry Dashboard

Security & Compliance Overview

Lookout Software, LLC.

Contents

1. Purpose and Scope

This document describes the security architecture, data-handling practices, and compliance posture of the AI-powered features in the Office Timesheets (OTS) AI Telemetry Dashboard. It is intended for customer IT, security, and compliance teams evaluating how artificial intelligence is applied to their organization's timesheet and labor cost data.

The AI Telemetry Dashboard provides two AI-assisted capabilities:

- **Narrative generation —** a short, natural-language commentary summarizing the data already displayed on a user's dashboard.
- **"Ask Your Data" —** a conversational interface that lets users pose questions about their timesheet data in plain English and receive calculated, verified answers.

A central design principle underpins both features: the language model never touches your database, never performs calculations, and never returns unverified numbers. The model's role is deliberately narrow — interpreting questions and describing data — while all data access, filtering, and computation remain inside the Office Timesheets platform, governed by a constrained query language (DSL) that acts as a hard guardrail.

2. Executive Summary

- **Minimal data exposure.** Only the data already visible on the user's dashboard (employees, elements, and entries information, serialized as JSON) plus the user's question is sent to the model. The model receives no database credentials, no connection strings, and no direct data access of any kind.
- **The LLM never writes SQL and never calculates metrics.** The model emits only a structured query intent in a restricted DSL. Office Timesheets validates that intent, generates parameterized read-only SQL itself, and performs every calculation server-side.
- **Constrained by design.** The DSL supports exactly five categories of question. Anything outside that registry — destructive operations, schema changes, arbitrary queries — is structurally impossible to express and is rejected.
- **Zero-retention inference.** Inference runs on the Ollama harness. Prompt and response data is never logged, never retained, and never used to train models.
- **Deployment choice.** Customers may use the OTS-managed inference layer (Google's Gemma 4 – 31B model on Ollama's US-based cloud), or connect their own Ollama harness running a cloud model of their choosing — keeping all AI traffic entirely within their own boundary.
- **Tenant isolation.** The per-tenant schema catalog and all query execution remain on the OTS side, scoped to the authenticated tenant. One customer's configuration and data are never visible to another's AI interactions.

3. Architecture Overview

The AI Telemetry Dashboard architecture is built on four pillars:

- **Schema catalog (per tenant).** Each customer's semantic data model — the levels, elements, and structures administrators configure — is maintained on the Office Timesheets side. The catalog gives the AI layer a vocabulary for the tenant's data without ever exposing the underlying database.
- **Constrained SQL generation via DSL.** For “Ask Your Data,” the LLM receives the user's question, dashboard-scoped data (employees, elements, entries information), and a strict system prompt. It emits a structured query DSL — not SQL. OTS translates validated intents into parameterized queries.
- **Pre-aggregated metrics and server-side calculation.** The LLM is never asked to do arithmetic. For “Ask Your Data,” OTS retrieves the metric intent from the LLM and calculates the result on the OTS side. Trend metrics are pre-aggregated by the platform.
- **Dashboard narrative generation.** For narratives, OTS provides dashboard data to the LLM for analysis and receives back a short descriptive phrase, which is displayed as natural-language commentary.

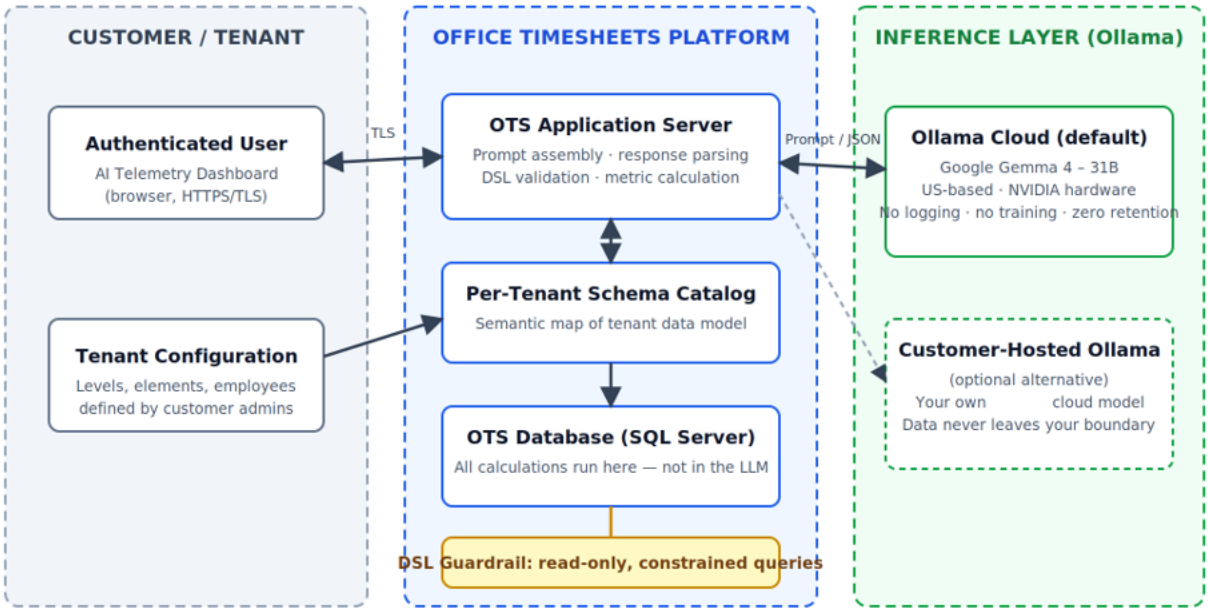

Figure 1 — System architecture and trust boundaries. The LLM sits behind two boundaries and never has a path to the database.

Three trust boundaries separate the system: the customer boundary (users and tenant configuration), the Office Timesheets platform boundary (application logic, schema catalog, database, and DSL enforcement), and the inference boundary (the Ollama harness, whether OTS-managed or customer-hosted). The only data that crosses into the inference boundary is a prompt containing dashboard-scoped JSON; the only thing that comes back is text or a structured JSON intent.

4. How the AI Features Work

4.1 Narrative Generation

When a dashboard narrative is generated, Office Timesheets serializes the data currently shown on the dashboard into JSON and sends it to the LLM together with a system prompt asking for a couple of sentences describing that data. The model's response — plain natural-language commentary — is displayed to the user. No follow-on actions, queries, or data access result from a narrative request; it is a strictly read-and-describe operation on data the user can already see.

4.2 “Ask Your Data”

“Ask Your Data” follows a more structured lifecycle, illustrated below.

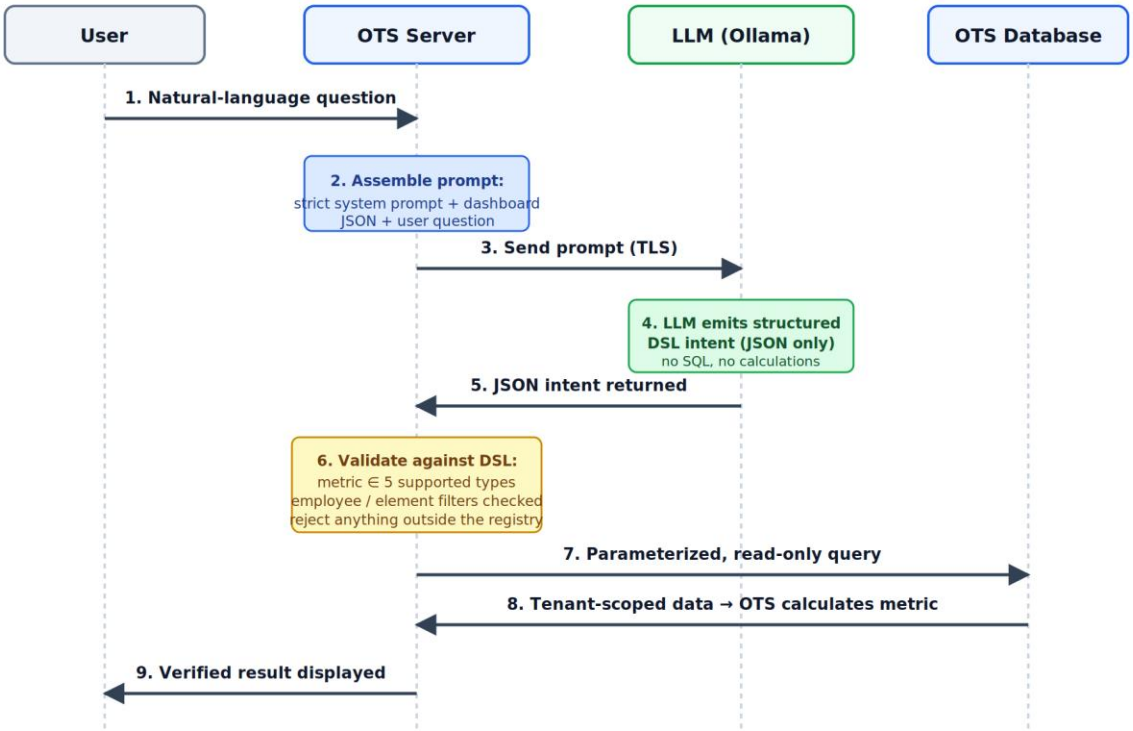

Figure 2 — “Ask Your Data” request lifecycle. The model produces an intent; Office Timesheets produces the answer.

The steps are:

- The user submits a natural-language question from the dashboard.
- OTS assembles a prompt from a strict system prompt, the dashboard data (employees, elements, entries information in JSON), and the user's question.
- The prompt is sent to the LLM over an encrypted connection.
- The LLM responds with a JSON structure identifying what the user wants calculated — the metric type and any employee or element filters. Five categories of questions are currently supported.
- OTS validates the returned intent against the DSL registry. Intents referencing unsupported metrics or invalid filters are rejected.

- OTS pulls the required data from its own database using parameterized, read-only queries scoped to the tenant, and calculates the metric itself.
- The calculated, verified result is displayed as the answer.

The consequence of this design is that **every number a user sees is computed by Office Timesheets from its own database** — the model's output determines only which supported calculation to run, never the result of that calculation. This eliminates the two most common risks of AI analytics tools: hallucinated figures and model-generated SQL.

5. Data Handling

5.1 What Is and Is Not Sent to the Model

Data category	Sent to the LLM?	Notes
Dashboard-scoped data (employees, elements, entries info)	Yes (JSON)	Limited to what the authenticated user's dashboard already displays
User's natural-language question	Yes	“Ask Your Data” only
Database credentials / connection strings	Never	The model has no data-access capability
Raw database tables or full tenant datasets	Never	Queries are executed by OTS, after DSL validation
Authentication tokens, passwords, payment data	Never	Outside the scope of prompt assembly

5.2 Data in Transit and at Rest

All traffic between the user's browser, the Office Timesheets platform, and the inference layer is encrypted in transit. Timesheet data at rest remains in the customer's Office Timesheets database exactly as it does without the AI features — the AI layer introduces no new persistent data store. On the inference side, prompt and response data is never logged and never retained (see Section 7), so no copy of your data accumulates at the model provider.

5.3 Data Retention at the Inference Layer

Inference requests are processed transiently: the prompt is used to generate a response and is then discarded. Prompt and response data is never logged and is never used for model training. Where Ollama partners with NVIDIA Cloud Providers to host models, no-logging, no-training, and zero-data-retention policies are contractually required.

6. AI Guardrails: The DSL as a Security Boundary

When AI agents perform actions — such as interacting with databases — they need boundaries. Office Timesheets implements those boundaries as a domain-specific language (DSL): a restricted, safe interface that gives the AI just enough expressive power to do its job while making destructive or out-of-scope operations impossible to express.

Key properties of the DSL guardrail:

- **Closed vocabulary.** The model can only select from the registry of supported query intents — currently five categories of question — with employee/element filter parameters. There is no free-form escape hatch.
- **No SQL from the model.** SQL is generated exclusively by Office Timesheets from validated intents, using parameterized queries. Model output is treated as data, not as code, which forecloses SQL injection via the AI pathway.
- **Read-only semantics.** The DSL expresses questions about data, not commands against it. Inserts, updates, deletions, schema changes, and administrative operations are not representable.
- **Server-side validation.** Every model response is parsed and checked before anything touches the database. Malformed, unsupported, or out-of-scope intents are rejected rather than "best-effort" executed.

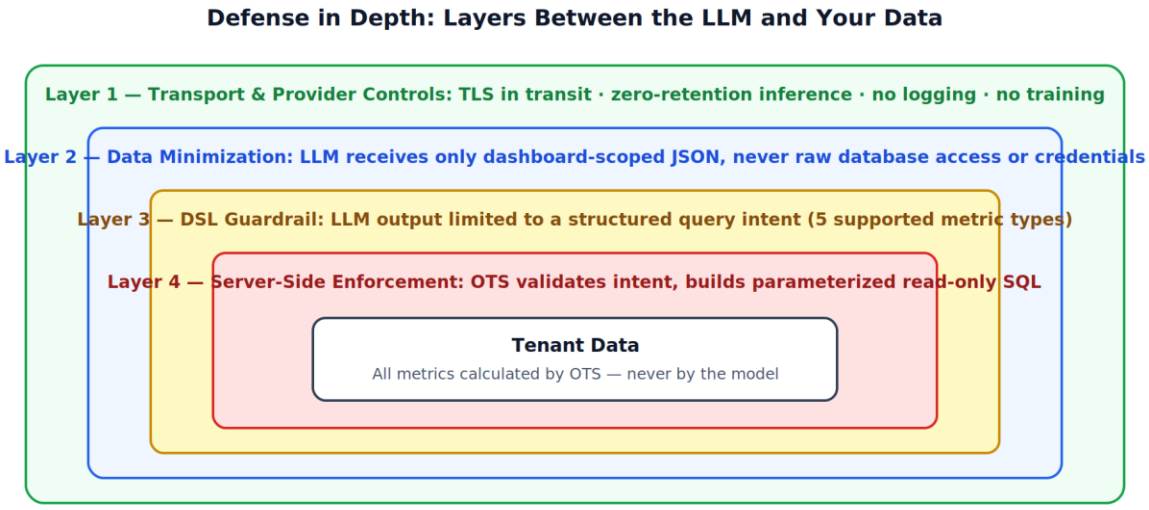

Figure 3 — Defense in depth. Four independent layers stand between the language model and tenant data.

7. Inference Layer: Ollama Harness and Model Options

7.1 The Ollama Harness

Office Timesheets uses the Ollama harness as its provider-agnostic inference layer:

- Trusted, US-based harness.

- Does not train on user data; operates servers only in the United States, Europe, and Singapore; and runs on modern NVIDIA hardware.
- Prompt and response data is never logged or trained on.
- Collaborates with NVIDIA Cloud Providers (NCPs) to host open models, and requires no-logging, no-training, and zero-data-retention policies from those partners.

7.2 Default Model: Google Gemma 4 – 31B

When customers use the OTS-managed inference option, inference is performed by Google's Gemma 4 – 31B cloud model, selected for being:

- Fast, accurate, and comprehensive for the structured tasks the dashboard requires.
- US-based in its hosting.
- Backed by a trusted company: Google.

7.3 Deployment Options

Option	How it works	Best for
OTS-managed inference (default)	OTS routes prompts to Gemma 4 – 31B on Ollama's US-based cloud under zero-retention terms.	Customers who want AI features with no infrastructure to run
Customer-hosted Ollama	You connect Office Timesheets to your own Ollama harness, running a cloud-based model you control.	Organizations with stricter data control policies

Both options use the identical application-side architecture — the schema catalog, DSL validation, and server-side calculation apply regardless of where inference runs.

8. Tenant Isolation and Access Control

- **Per-tenant schema catalog.** Each tenant's semantic catalog is configured and stored on the OTS side by that customer's administrators. AI interactions are grounded exclusively in the requesting tenant's catalog.
- **Authenticated, session-scoped requests.** AI features are available only to authenticated Office Timesheets users, and prompts are assembled from the data that user's dashboard exposes.
- **Tenant-scoped execution.** When OTS executes a validated query intent, the query runs against the requesting tenant's data only. There is no cross-tenant query path in the DSL.
- **Customer-controlled configuration.** Because customers configure their own levels and elements, administrators directly control the vocabulary the AI can reason over for their organization.

9. Threat Model and Mitigations

Threat	Risk if unmitigated	Mitigation in OTS
Prompt injection via user question	Malicious question tries to make the model exceed its role	Strict system prompt; model output constrained to DSL JSON; OTS validates every intent against the registry before any execution; unsupported intents rejected
SQL injection via model output	Model-generated SQL executes arbitrary commands	Model never emits SQL; OTS builds parameterized queries from validated intents; model output treated as data, not code
Destructive operations	AI modifies or deletes records	DSL is read-only by construction; write operations are not expressible
Hallucinated figures	Users act on invented numbers	LLM performs no calculations; all metrics computed by OTS from the database; trends use pre-aggregated metrics
Data leakage to model provider	Sensitive labor data retained or trained on externally	Zero-retention, no-logging, no-training inference; only dashboard-scoped JSON sent; customer-hosted option keeps prompts fully in-boundary
Cross-tenant exposure	One tenant's data appears in another's answers	Per-tenant schema catalog; tenant-scoped prompt assembly and query execution
Excessive data sent to inference	Whole datasets exposed to the model	Data minimization: only the data already on the user's dashboard is serialized into the prompt

10. Compliance Considerations

The AI Telemetry Dashboard is designed to slot into customer compliance programs without changing the data-protection posture of Office Timesheets itself:

- **Data minimization and purpose limitation.** Only dashboard-scoped data needed to interpret the request is sent for inference, aligning with data-minimization principles found in frameworks such as GDPR and CCPA.
- **No new data processor accumulation.** Because inference is zero-retention with no logging and no training, use of the AI features does not create a growing copy of personal data at a third party.
- **Data residency options.** The Ollama harness operates servers only in the United States, Europe, and Singapore, and the default model is US-based Gemma 4 (Alphabet/Google). Customers with stricter residency requirements can host their own Ollama harness (requires cloud model).

- **Human-verifiable outputs.** Because all displayed metrics are calculated by OTS from the database, results are auditable and reproducible — supporting internal controls that require traceable figures.
- **Customer-controlled scope.** Administrators define the tenant's levels and elements, so organizations decide what vocabulary and structures the AI features operate over.

Customers subject to specific regulatory regimes should evaluate the AI features under their own compliance frameworks; Lookout Software can support security reviews with architecture details from this document. This document describes technical design and provider policies and is not legal advice.

11. Shared Responsibility

Party	Responsibilities
Office Timesheets (Lookout Software)	Maintaining the schema catalog infrastructure, DSL registry and validation, parameterized query generation, server-side metric calculation, and the integration with the inference layer under zero-retention terms.
Inference provider (Ollama and hosting partners)	Operating inference on modern NVIDIA hardware in the US, Europe, and Singapore with no logging, no training on user data, and zero data retention.
Customer	Configuring tenant levels and elements; managing user accounts and access.

12. Frequently Asked Questions

Does the AI have access to our database? No. The model receives only dashboard-scoped JSON and a question. All database access is performed by Office Timesheets using parameterized, read-only queries after validating the model's structured intent.

Is our data used to train AI models? No. Prompt and response data is never logged or trained on, and Ollama's hosting partners are required to operate under no-logging, no-training, zero-retention policies.

Can the AI change or delete our timesheet data? No. The DSL cannot express write operations, and the model never produces SQL.

Can we keep AI processing entirely inside our own environment? Yes. Connect Office Timesheets to your own Ollama harness running cloud-based model you control.

Are the numbers the AI shows us reliable? The numbers are not produced by the AI. The model only identifies which supported metric to compute and which filters to apply; Office Timesheets calculates the result from your data.

Where does inference run? By default, on Ollama's US-based cloud using Alphabet's/Google's Gemma 4 – 31B model, on modern NVIDIA hardware. Ollama operates servers only in the US, Europe, and Singapore.