RO Assistant Pilot

Security, Architecture, and Governance Review Packet

To: Dealer IT, Information Security, and Compliance
From: RO Assistant Project Team
Date: December 2025
Subject: Pilot Review Package – RO Assistant

Purpose of This Package

You are receiving this document set to support technical, security, and compliance review of the RO Assistant pilot.

The RO Assistant is a limited-scope system intended to make historical repair order information more accessible to dealership personnel in a controlled, read-only manner. The pilot has been deliberately designed to minimize operational risk, data exposure, and integration complexity while allowing meaningful evaluation of the underlying capability.

This packet is intended to clearly explain **how the system works, what data it touches, how access is controlled, and how data is retained or removed**, so that the pilot can be reviewed on its merits without ambiguity.

Pilot Intent and Design Philosophy

The pilot deployment is intentionally conservative.

- It operates in a **single-dealer context**
- It is **read-only** with respect to dealership systems
- It does **not** write back to the DMS or modify records
- It does **not** automate decisions or diagnostics

All data ingested into the system is explicitly provided by the dealer and remains logically isolated by tenant. Sensitive information is excluded by default. The platform architecture assumes a cautious posture: least privilege, clear trust boundaries, and auditable access.

While the system internally uses modern search and language-processing techniques to surface relevant historical information, outputs are informational only and are grounded in existing repair order records.

Handling of Personally Identifiable Information (PII)

The RO Assistant pilot is designed to operate **without reliance on personally identifiable information (PII)**. As deployed for the pilot, ingestion and processing focus on repair order content necessary for technical reference, and PII is intentionally excluded wherever practicable.

The platform architecture, however, is capable of securely handling PII **if and only if** such use is explicitly approved by the dealer group. This capability is not enabled by default.

If approved in a future phase, PII handling would adhere to the following principles:

- PII ingestion would be explicitly scoped, documented, and limited to approved data fields
- PII would be logically and physically separated from non-PII operational data
- Access to PII would be restricted by role, tenant, and purpose, and fully audited
- PII would be encrypted at rest and in transit using industry-standard mechanisms
- Retention and deletion of PII would follow dealer-defined policy and documented retention schedules

No PII is shared across tenants, used for model training outside the tenant boundary, or exposed to unauthorized users. Any expansion of PII usage would require updated documentation and formal review prior to enablement.

About AI Usage in This Pilot

The RO Assistant uses AI techniques internally to improve search relevance and summarization of historical repair information. These capabilities are deliberately constrained:

- AI components do not have direct access to dealership systems
- No autonomous actions are taken
- Outputs are always traceable to underlying source records
- The system does not learn from or modify dealer data outside the defined ingestion flow

What This Packet Contains

This packet includes the following documents, which are intended to be read together:

- **Security Summary** – High-level overview of the system's security posture and design controls
- **Data Flow Diagram** – Clear depiction of how data enters, moves through, and exits the system
- **Pilot Scope and Non-Goals** – Explicit definition of what the pilot does and does not include
- **Access Control Summary** – How users, roles, and tenant boundaries are enforced
- **Data Retention and Deletion Policy** – How pilot data is retained, reviewed, and removed

Scope Control and Future Expansion

This pilot does not imply approval for broader deployment, expanded integrations, or production-scale use. Any expansion—such as multi-store deployment, write-back functionality, or broader use of customer or vehicle PII—would require separate review and updated documentation.

The current documentation reflects the **pilot as deployed**, not a future or hypothetical system state.

RO Assistant Security Summary

Document Version: 1.0
Last Updated: December 2025
Pilot Scope: Single-Dealer, Read-Only Pilot

Applies To: RO Assistant Pilot Deployment

1. Purpose & Scope

This document provides a high-level security overview of the RO Assistant, a read-only system designed to make historical repair order (RO) knowledge searchable for dealership personnel.

This summary is intended to support:

- Dealer IT review
- Security and compliance evaluation
- Pilot approval decisions

The current deployment is a single-dealer, read-only pilot with intentionally constrained scope.

2. System Overview

The RO Assistant:

- Ingests historical RO text provided via export
- Converts non-PII text into embeddings for semantic search
- Returns citation-backed, reference-only answers
- Does not write back to dealership systems
- Does not perform automated diagnosis or recommendations

The system is architected as a platform with explicit trust boundaries, not a monolithic AI application.

3. Core Security Principles

The system is designed around the following principles:

1. Least Privilege
2. Fail-Closed by Default
3. Strong Tenant Isolation
4. Separation of AI and Sensitive Data
5. Full Auditability
6. Policy-Driven Expansion

These principles apply consistently across authentication, data storage, AI usage, and access control.

4. Authentication & Access Control

Authentication

- Users authenticate via /auth/login
- Credentials are verified against an internal user table
- A signed JWT is issued upon successful authentication
- All subsequent requests derive identity only from the verified JWT

There is no trust in request headers or request body for identity or tenant context.

Authorization

- Role-Based Access Control (RBAC) is enforced at the API layer
- Roles determine access to:
  - Ingestion (Admin only)
  - Search and answer (Technician / Admin)
  - Audit logs (Admin only)

Tenant Isolation

- Each request is scoped to a single dealer tenant
- Tenant isolation is enforced:
  - In application middleware
  - In database session context
  - Via PostgreSQL Row-Level Security (RLS)
- Cross-tenant access is not possible by design

5. Data Handling & Storage

Data Sources

- Repair Order text is provided via explicit export (CSV)
- The RO Assistant does not connect directly to dealership systems
- The DMS remains the system of record at all times

Storage

- Data is stored in PostgreSQL with tenant-scoped tables
- Vector embeddings are stored using pgvector
- All data is encrypted at rest and in transit

No Write-Back

The system:

- Does not modify repair orders
- Does not update customer records
- Does not write back to the DMS or any dealership system

6. PII Handling (Current and Approved Modes)

Current Pilot Mode (Default)

In the pilot configuration:

- PII ingestion is disabled by policy
- If PII is detected during ingestion, the operation fails
- No PII is stored
- No PII is sent to embeddings or LLMs
- No PII appears in logs or audit records

This allows pilots to proceed with minimal compliance overhead.

Approved PII Mode (Optional)

If a dealer group explicitly approves PII usage, the system supports it without architectural changes:

- PII is extracted and stored only in a dedicated, encrypted pii_vault
- AI components (embeddings, LLMs) never receive PII
- PII access is:
  - Role-restricted
  - Explicit
  - Fully audited
- PII is tenant-isolated and encrypted at all times

PII support is a policy decision, not a technical limitation.

7. AI / LLM Usage Controls

The RO Assistant uses Azure OpenAI services for:

- Text embeddings
- Optional answer summarization

Strict controls apply:

- Only non-PII, redacted text is sent to AI services
- AI is used for reference-only assistance
- The system does not generate recommendations or diagnoses
- Every answer includes citations to source ROs
- If insufficient data exists, the system returns a deterministic fallback

The system does not:

- Train models on dealer data
- Share data across tenants
- Retain prompts or responses beyond request scope

8. Audit Logging & Monitoring

All security-relevant actions are audited, including:

- Authentication attempts
- Authorization denials
- Ingestion events (success/failure)
- Query and answer requests

- PII access (if enabled)

Audit logs:

- Are tenant-scoped
- Contain metadata only (no raw text, no PII)
- Are immutable and viewable by authorized admins

9. Failure Behavior & Resilience

The system is designed to fail predictably and safely:

- Missing configuration service fails fast
- Embeddings unavailable search/answer return controlled 503
- PII detected ingestion fails with no partial state
- Unauthorized access denied and audited

There are no silent failures.

10. Pilot Scope & Limitations

This pilot intentionally excludes:

- DMS write-back
- Cross-store data sharing
- Automated decision-making
- Production SSO integration
- Distributed rate limiting
- Antivirus scanning of uploads

These can be added later under formal change control if required.

11. Summary for Reviewers

In short:

- The system is read-only

RO Assistant — Security Summary
Pilot Deployment

- Data remains under dealer control
- AI is isolated from sensitive data
- Tenant boundaries are strictly enforced
- PII is excluded by default but supported securely if approved
- All access is auditable

This design allows dealer groups to evaluate value safely before expanding scope.

Diagram Pilot 1. Executive Overview (12 minutes to read)

Document Version: 1.0
Last Updated: December 2025
Pilot Scope: Single-Dealer, Read-Only Pilot
Applies To: RO Assistant Pilot Deployment

The RO Assistant is a read-only, tenant-isolated system that ingests historical repair order (RO) text, converts it into searchable embeddings, and returns citation-backed answers to user queries.

Key properties:

- No write-back to dealership systems
- No cross-tenant data sharing
- No PII sent to LLMs
- All access audited
- All data scoped to a single dealer tenant
- PII ingestion is fail-closed

The system can operate with synthetic data only until IT approves access to real RO exports.

2. Formal Data Flow (System of Record)

Actors

- User: Technician, Service Writer, or Admin (authenticated)
- Dealer IT / DMS: System of record (external, read-only export)
- RO Assistant Platform: Application backend
- PostgreSQL (Azure): Primary data store
- Azure OpenAI: Embeddings + LLM (no PII)

Flow A — Authentication & Session Establishment

1. User authenticates via /auth/login
2. Credentials verified against app.users
3. System issues a signed JWT containing:
    a. user_id
    b. tenant_id

   c. role

4. JWT is attached to subsequent requests
5. All requests derive identity only from JWT, not headers or body

Security Controls

- JWT verification on every request
- Role-Based Access Control (RBAC)
- Tenant isolation enforced at DB and app layers
- Login lookup uses a locked-down SECURITY DEFINER function

Flow B — RO Ingestion (Admin Only)

Trigger: Admin uploads RO export (PDF / TXT)

1. File uploaded to /workloads/ro/ingest
   a. PII redacted before leaving dealer group IT, unless authorized.
2. File validation:
   a. Type allowlist
   b. Size limits
3. Text extraction occurs in-app
4. PII Scan
   a. If PII detected ingestion fails (no partial writes)
5. Text is chunked
6. Chunks are embedded via Azure OpenAI Embeddings
   a. Only non-PII text
7. Data stored in Postgres:
   a. repair_orders
   b. ro_chunks
   c. ro_embeddings
8. Audit events written for ingestion success/failure

Security Controls

- Admin-only route
- Fail-closed PII detection
- No raw files stored
- No PII in embeddings
- Transactional writes (no partial state)
- Full audit trail

Flow C — Query & Retrieval (Tech / Admin)

Trigger: User submits a query

1. User sends query to /workloads/ro/search
2. Query embedded (no PII allowed in query)
3. Vector search performed via pgvector
    a. Tenant-scoped
    b. RLS enforced
4. Matching chunks returned (redacted text only)
5. Citations constructed from source ROs
6. Audit event logged (query hash only)

Security Controls

- Tenant isolation via RLS
- No raw RO text returned
- No PII in logs
- No cross-tenant visibility

Flow D — Answer Generation

1. Retrieved chunks sent to LLM
2. LLM instructed:
    a. Reference-only
    b. No recommendations
    c. No hallucinations
3. Answer returned with citations
4. Audit event logged

Security Controls

- LLM sees only redacted chunks
- No system instructions allowing speculation
- Deterministic fallback if data insufficient

The system does NOT:

- Write back to DMS
- Modify repair orders
- Share data across dealers
- Store raw files long-term
- Send PII to external services
- Train models on dealer data
- Execute autonomous actions

Design Intent Regarding PII

The RO Assistant platform is architected to securely handle PII, but PII ingestion is currently disabled by policy for the pilot phase.

This decision is governance-driven, not technical.

The system already includes the controls required to ingest, store, and access PII if and only if a dealer group explicitly approves its use.

Current Pilot Mode (Default)

PII Handling Policy:
PII ingestion is disabled (fail-closed)

In the current pilot configuration:

- Any detected PII during ingestion causes the operation to fail
- No PII is stored
- No PII is sent to embeddings or LLMs
- No PII appears in logs, prompts, or vector stores

This mode is intended to:

- Minimize compliance friction
- Enable evaluation without data sensitivity concerns
- Reduce approval timelines for initial pilots

Approved PII Mode (Optional, Not Enabled by Default)

If a dealer group approves the use of PII, the system supports a separate, hardened PII data flow with the following properties:

PII Data Flow (When Enabled)

1. RO text is ingested
2. PII is detected and extracted
3. PII is stored only in a dedicated PII store (pii_vault)
4. Non-PII text proceeds through the normal embedding/search pipeline
5. PII is never embedded
6. PII is never sent to LLMs
7. PII access is:
    a. Role-restricted
    b. Explicitly audited
    c. Tenant-isolated
    d. Encrypted at rest and in transit

Access Control for PII (If Enabled)

If PII storage is enabled:

- Only explicitly authorized roles may access PII
- PII access requires:
    - Elevated role
    - Explicit endpoint
    - Full audit logging
- PII access is never implicit or automatic

Example:

- A Return service history for this customer request resolves customer identity outside the LLM pipeline and joins data after retrieval.

RO Assistant Access Control Summary

Document Version: 1.0
Last Updated: December 2025
Pilot Scope: Single-Dealer, Read-Only Pilot
Applies To: RO Assistant Pilot Deployment

1. Purpose

This document summarizes how authentication, authorization, tenant isolation, and auditing are enforced within the RO Assistant during the pilot phase.

Its goals are to:

Clearly define who can access the system
Specify what actions each role is permitted to perform
Explain how access is technically enforced
Demonstrate how access is monitored and audited

This document should be read alongside the Security Summary and Pilot Scope & Non-Goals.

2. Identity & Authentication

Authentication Method

Users authenticate via /auth/login using email and password
Credentials are verified against an internal user table
Passwords are stored as secure cryptographic hashes (e.g., bcrypt)
Upon successful authentication, the system issues a signed JWT

JWT Properties

The JWT includes:

user_id
tenant_id
role
Expiration (exp)

JWTs are:

- Signed with a server-side secret
- Verified on every request
- Never trusted from request headers beyond the Authorization bearer token

There is no implicit trust in:

- Client-provided tenant identifiers
- Request body or query parameters

3. Roles & Responsibilities

The pilot defines a minimal role model to reduce complexity and risk.

Admin Role

Admins may:

- Ingest repair order data
- Search and retrieve RO references
- View audit logs
- Manage pilot users (if enabled)

Admins may not:

- Write back to the DMS
- Modify repair orders
- Access other tenants data

Technician Role

Technicians may:

- Search historical ROs
- View citation-backed answers

Technicians may not:

- Ingest data
- View audit logs
- Access PII
- Modify system configuration

4. Authorization Enforcement

API-Level Enforcement

All routes are protected by authentication middleware
Role-Based Access Control (RBAC) is enforced before request handling
Unauthorized access attempts are denied and audited

Database-Level Enforcement

PostgreSQL Row-Level Security (RLS) enforces tenant isolation
Each request sets session-scoped variables:
app.tenant_id
app.user_id
app.role
Queries without a valid tenant context fail closed

This ensures that even application bugs cannot bypass tenant isolation.

5. Tenant Isolation

Tenant isolation is enforced at multiple layers:

Layer	Control
Application	Tenant context derived from JWT only
Middleware	Tenant guard validates tenant active status
Database	RLS policies restrict row access
Storage	Tenant-scoped primary keys and indexes

Cross-tenant access is not possible by design.

6. PII Access Controls (If Enabled)

Default Pilot Mode

PII ingestion is disabled
No users can access PII
Any detected PII causes ingestion to fail

Approved PII Mode (Optional)

If explicitly approved by the dealer:

- PII is stored only in a dedicated encrypted pii_vault
- PII access:
- Requires elevated role
- Occurs via explicit endpoints
- Is never implicit
- All PII access events are fully audited

AI components:

- Never receive PII
- Never access the PII vault

7. Audit Logging & Monitoring

All access control decisions are logged, including:

- Authentication successes and failures
- Authorization denials (RBAC, tenant inactive)
- Data ingestion events
- Search and answer requests
- PII access (if enabled)

Audit logs:

- Contain metadata only (no RO text, no PII)
- Are tenant-scoped
- Are immutable
- Are accessible only to Admin users

8. Failure Handling

Access control failures are handled predictably:

Scenario	**Behavior**
Invalid or expired token	Request denied (401) + audit
Insufficient role	Request denied (403) + audit
Inactive tenant	Request denied (403) + audit
Missing tenant context	Fail closed + audit

There are no silent failures.

9. Separation of Duties

The system enforces separation between:

- Data ingestion vs. data consumption
- Administrative actions vs. technician access
- AI processing vs. sensitive data storage

This limits blast radius and reduces insider risk.

RO Assistant Data Retention & Deletion Policy

Document Version: 1.0
Last Updated: December 2025
Pilot Scope: Single-Dealer, Read-Only Pilot

Applies To: RO Assistant Pilot Deployment

1. Purpose

This document defines how data is retained, protected, and deleted within the RO Assistant during the pilot phase.

The goals of this policy are to:

- Minimize data retention risk
- Ensure data is not held longer than necessary
- Enable clean, auditable deletion
- Maintain dealer control over all data

This policy applies to all data stored or processed by the RO Assistant.

2. Data Ownership

- All repair order (RO) data remains the property of the dealer group
- The RO Assistant does not claim ownership or reuse rights
- Dealer data is never used for model training
- Data is not shared across tenants or customers

The RO Assistant acts solely as a data processor during the pilot.

3. Data Categories & Retention Rules

3.1 Repair Order Text (Non-PII)

Examples

- Technician notes

- Complaint descriptions
- Repair descriptions (redacted)

Storage

- Stored in PostgreSQL
- Tenant-isolated via Row-Level Security
- Encrypted at rest and in transit

Retention

- Retained for the duration of the pilot
- Deleted upon pilot termination or upon dealer request

3.2 Vector Embeddings

Description

- Numeric vector representations of non-PII RO text
- Used solely for semantic search

Storage

- Stored in PostgreSQL using pgvector
- Tenant-scoped
- Cannot be reverse-engineered to reconstruct original text

Retention

- Retained only while corresponding RO text exists
- Automatically deleted when source RO records are deleted

3.3 Audit Logs

Description

- Security and access metadata
- Examples: login attempts, ingestion events, access denials

Storage

- Stored in PostgreSQL
- Metadata only (no RO text, no PII)

Retention

- Retained for a minimum of 90 days during the pilot
- Configurable based on dealer preference
- Deleted upon pilot termination, unless otherwise requested

3.4 Personally Identifiable Information (PII)

Pilot Default

- PII ingestion is disabled
- Any detected PII causes ingestion to fail
- No PII is stored during the default pilot

If PII Is Explicitly Approved

- Stored only in a dedicated, encrypted pii_vault
- Never embedded
- Never sent to LLMs
- Access is explicitly audited

Retention (If Enabled)

- Retained only for approved use cases
- Subject to dealer-defined retention limits
- Deleted upon request or pilot termination

4. Data Deletion Procedures

4.1 Dealer-Initiated Deletion

Upon dealer request, the following can be deleted:

- All RO data for a tenant
- All embeddings derived from that data
- All associated audit logs (unless legally required to retain)

Deletion is:

- Explicit
- Logged
- Verifiable

4.2 Pilot Termination

At the conclusion of the pilot:

- All dealer-provided data is deleted within 30 days
- This includes:
  - RO text
  - Embeddings
  - PII (if enabled)
- Audit logs may be retained temporarily for closure documentation, then deleted

A deletion confirmation can be provided upon request.

4.3 Partial Deletion (Optional)

If requested, the system supports:

- Deletion by date range
- Deletion by RO batch
- Deletion by data category

Partial deletion does not impact other tenants.

5. Backups & Replication

Pilot Mode

- Backups are limited and time-bound
- Backup retention mirrors primary retention where feasible
- Deleted data is not restored from backup after retention windows expire

If longer backup retention is required, this will be disclosed and approved in advance.

6. Data Portability

Upon request:

- Dealer data can be exported in a standard format
- Export does not include system-derived embeddings unless requested
- Export operations are audited

7. Data Residency

- pilot deployments may run locally or in-region (e.g., Azure)
- Data does not leave the approved region
- No cross-region replication without approval

8. Verification & Auditability

The RO Assistant provides:

- Auditable deletion actions
- Confirmation of deletion events
- Evidence of retention enforcement upon request

9. Policy Changes

Any changes to this policy require:

- Dealer notification
- Updated documentation
- Explicit approval if retention is expanded

RO Assistant Pilot Scope & Non-Goals

Document Version: 1.0
Last Updated: December 2025
Pilot Scope: Single-Dealer, Read-Only Pilot
Applies To: RO Assistant Pilot Deployment

1. Purpose of This Document

This document defines the explicit scope and boundaries of the RO Assistant pilot.

Its purpose is to:

- Prevent scope creep
- Set clear expectations for dealer IT and operations
- Reduce risk during evaluation
- Ensure the pilot remains read-only, low-impact, and reversible

Anything not explicitly listed as In Scope is considered out of scope for the pilot.

2. Pilot Objectives (What This Pilot Is Meant to Prove)

The pilot is designed to evaluate:

- Whether historical repair order (RO) text contains reusable institutional knowledge
- Whether that knowledge can be surfaced quickly and safely
- Whether technicians find citation-backed references useful in daily work
- Whether this can be done without changing existing workflows or systems

The pilot is not intended to automate decisions or replace existing systems.

3. In-Scope Capabilities (Explicitly Included)

3.1 Read-Only RO Ingestion

- Historical RO text is ingested via explicit export

- Supported formats: CSV
- No direct connection to the DMS
- No scheduled or automated sync

3.2 Search & Reference Retrieval

- Users may search historical ROs using natural language
- Results are:
    - Citation-backed
    - Read-only
    - Tenant-scoped
- Answers reference prior work but do not instruct actions

3.3 AI Summarization

- AI is used to summarize retrieved RO excerpts
- AI outputs are:
    - Reference-only
    - Non-diagnostic
    - Non-prescriptive
- All summaries include source citations

3.4 Authentication & Role-Based Access

- Authenticated users only
- Roles limited to:
    - Admin
    - Technician
- All access is logged

3.5 Single-Tenant Operation

- Pilot operates on one dealer tenant
- No cross-store or cross-group visibility
- Tenant isolation enforced at application and database levels

4. Out-of-Scope (Explicit Non-Goals)

The following are explicitly excluded from the pilot.

4.1 No DMS Write-Back

The system does not:

- Update repair orders
- Modify customer records
- Create, close, or edit ROs
- Push recommendations back into the DMS

The DMS remains the system of record at all times.

4.2 No Automated Recommendations or Diagnosis

The system does not:

- Diagnose vehicle issues
- Recommend repairs
- Suggest parts
- Determine labor operations
- Provide procedural instructions

All outputs are informational references only.

4.3 No Cross-Dealer or Cross-Store Sharing

- Data from one dealer is never visible to another
- There is no aggregation across stores or groups
- Each tenant is fully isolated

4.4 No Training or Performance Evaluation

The pilot does not:

- Evaluate technician performance
- Score repair quality
- Track productivity
- Replace training programs

The tool is assistive, not evaluative.

4.5 No Workflow Enforcement

The system does not:

- Require usage during repairs
- Enforce steps or processes
- Integrate into existing workflows

Use is optional and advisory.

4.6 No Production Identity or SSO Integration

For the pilot:

- Enterprise SSO (e.g., Entra ID) is not required
- Authentication is local and scoped
- Production IAM integration is out of scope

4.7 No PII Usage

- PII is excluded by default
- PII ingestion requires explicit dealer approval
- PII handling is not enabled implicitly

5. Pilot Constraints (Deliberate Limitations)

The pilot is intentionally constrained to:

- One dealer tenant
- Read-only access
- Time-bounded evaluation period
- Explicit opt-in for any scope expansion

These constraints are designed to:

- Minimize operational risk
- Simplify approval
- Allow clean rollback

6. Expansion Policy (Post-Pilot)

Any of the following require:

- Explicit dealer approval
- Updated security review
- Revised documentation

Examples:

- Multi-store visibility
- PII-enabled workflows
- DMS write-back
- Production SSO
- Advanced analytics

Expansion is policy-driven.