SECURITY-FIRST

DEFENSE ARCHITECTURE

LAYER 01
Pre-LLM
Write-intent and prompt-injection detection before the question reaches the model.

LAYER 02
Post-LLM
AST schema validation, regex against DDL/DML, LIMIT cap, column auto-correction.

LAYER 03
Execution
Per-dialect statement timeout, work_mem limit, Cartesian detection, EXPLAIN gate.

LAYER 04
Connection
Read-only pool per tenant, driver-level enforcement, AES-256 credentials, 5-min idle TTL.

GUARDRAILS IMPLEMENTED

IDENTIFIER	NAME	WHAT IT DOES	STATUS
G-PRE-1	detectWriteIntent	Blocks questions with write or DDL intent before calling the LLM	OK
G-PRE-2	detectPromptInjection	Detects prompt injection attempts in natural language	OK
G-POST-1	checkSqlSecurity	Regex against INSERT, UPDATE, DELETE, DROP, TRUNCATE, ALTER, CREATE, MERGE, GRANT, REVOKE, EXEC	OK
G-POST-2	PreExecutionValidator	AST validation of SQL against the real schema; auto-correction via Levenshtein distance	OK
G-POST-3	LIMIT cap	Overrides any LIMIT greater than 500 and injects one when missing	OK
G-EXEC-1	Statement timeout	Dedicated timeout per dialect on PG, MySQL, SQL Server, and Oracle	OK
G-EXEC-2	work_mem limit	Per-session working memory limit on PostgreSQL	OK
G-EXEC-3	Cartesian guard	Detects Cartesian products before execution and blocks the plan	OK
G-EXEC-4	EXPLAIN cost gate	Rejects plans with cost above the configured threshold (feature-flag)	OK
G-CONN-1	Read-only enforcement	SqlExecutionGuard.ensureReadOnly guarantees a read-only connection at the driver level	OK
G-CONN-2	Pool size limit	Maximum of 5 simultaneous connections per workspace to prevent saturation	OK
G-CONN-3	Connect timeout	Connection-open timeout implemented on every dialect	OK
G-CONN-3b	Shared pool singleton	Single pool per connectionId with 5-minute idle TTL	OK
G-RESULT-1	Row limit safety net	LIMIT 500 / FETCH FIRST / TOP applied even when the SQL omits it	OK
G-RESULT-2	Truncation notice	Visible banner to the user when the result was truncated by the LIMIT	OK

DATA GOVERNANCE · GDPR / LGPD NATIVE

Governance, audit, and isolation *by construction.*

Isolated multi-tenant, encryption at rest and in transit, granular RBAC, PII masking, and append-only audit log. GDPR and LGPD compliance mapped at the architecture level.

Data classification

Four levels with distinct access and retention rules.

- **Restricted** — Credentials, tokens, passwords
- **Confidential** — Customer PII, tax data
- **Internal** — Internal metrics, logs
- **Public** — Marketing content, docs

Three-role RBAC

Permissions enforced at API, UI, and query layers.

CAPABILITY	ADMIN	USER	VIEWER
DB connections	CRUD	—	—
Dashboards	All	Own	Read
NL queries	Yes	Yes	Yes
Audit log	Read	—	—
Users	CRUD	—	—

Encryption and isolation

Data never crosses tenant boundaries.

- **AES-256-GCM** for database credentials and tokens at rest
- **bcrypt (12 rounds)** for user passwords
- **TLS 1.2+** mandatory on every connection
- **Mandatory tenant filter** on every metastore query
- **Workspace isolation**: pool, credentials, and audit

PII masking and LGPD

Automatic masking in logs and responses.

- **11 patterns detected**: email, tax ID, phone, card, IP, etc.
- **Email** masked in logs (u***@domain.com)
- **User IP** stored as salted SHA-256 hash
- **User-Agent** truncated to prevent fingerprinting
- **Right to be forgotten**: soft delete + scheduled purge

Append-only audit

Immutable log of every sensitive action.

- Required schema: **tenantId, userId, action, resource, IP hash, outcome**
- No **update/delete** methods on the audit repository
- Covers login, queries, connection changes, RBAC, and export
- Queryable by the tenant admin through a dedicated UI

Retention and disposal

Policy aligned with LGPD and GDPR.

- **Audit log**: 5 years active + 10 years archive
- **Query history**: 1 year (configurable per tenant)
- **Sessions**: 30 days, revocable by admin
- **Backups**: encrypted with customer-managed key (BYOK optional)

NL2SQL maturity *in the high-end enterprise tier.*

Internal assessment based on the HKUST NL2SQL Handbook (5-level taxonomy). Entendo implements 35+ techniques from the top 10% of the academic literature and features 7

8.2
/ 10
GLOBAL SCORE

SCORECARD BY PIPELINE STAGE

Pre-Processing	8.0	Generation (SQL)	7.5
Post-Processing	9.0	Security	9.5
Multi-turn	8.0	Cross-Dialect	8.5
Visualization	9.0	Analytics / Prediction	8.5

SIX DIFFERENTIATORS WITH NO ACADEMIC EQUIVALENT

DIFF 01
4-Layer Defense-in-Depth
Unique security architecture: pre-LLM, post-LLM, execution, and connection, all mandatory.

DIFF 02
Business semantic layer
Term dictionary mapped to the real schema, with configurable synonyms and hierarchies.

DIFF 03
Cross-Dialect Error Classification
13 error classes x 5 dialects x 100+ SQLSTATE codes translated into business-friendly messages.

DIFF 04
GROUP BY Auto-Fix
Recursive auto-correction via CTE when the generated aggregation violates dialect rules.

DIFF 05
Schema Evolution Detection
Proactive detection of schema changes and re-indexing of the semantic layer.

DIFF 06
PII Detection & Masking
11 patterns automatically detected in logs and responses, with IP and email hashing.

COMPARISON WITH MARKET REFERENCES

CRITERION	THOUGHTSPOT SAGE	TABLEAU ASK DATA	METABASE AI	ENTENDO
Defense-in-depth (4 layers)	Partial	Partial	Partial	**Complete**
SQL dialects supported	3	Via driver	4+	**5 (with BigQuery)**
Read-only by design	Optional	Optional	Optional	**Mandatory**
On-prem deploy	No	Yes	Yes	**Yes**
Business semantic layer	Yes	Partial	No	**Yes**
PII masking native	No	No	No	**Yes (11 patterns)**
LGPD / GDPR native	Generic	Generic	Generic	**Mapped**