



SECURITY & PLATFORM WHITEPAPER

Self-Hosted Enterprise AI, Engineered for Trust

Architecture, data-residency controls, and platform capabilities of the VectorMATRIX self-hosted AI stack — for enterprise security, procurement, and compliance review.

GDPR / DSGVO aligned

Zero document persistence

Customer-controlled infrastructure

Germany & Worldwide

Contents

1. Executive Summary
2. Platform Capabilities
 - 2.1 Executive Presentation Automation
 - 2.2 Corporate Template Engine
 - 2.3 Report Mode — Periodic Financial Reporting
 - 2.4 Strategic Quadrant Analysis
 - 2.5 Workflow Automation with n8n
 - 2.6 Private AI Chat
3. Deployment Model
4. Data Flow and Residency
5. Access Control
6. Application Security
7. GDPR / DSGVO Alignment
8. Availability and Operations
9. Summary for Security Reviewers

1. Executive Summary

The VectorMATRIX platform delivers AI capabilities — presentation automation, periodic reporting, workflow intelligence, and private AI chat — entirely within customer-controlled infrastructure. The architecture follows a single governing principle: **customer content never leaves customer-controlled systems.**

All AI processing occurs either on local models inside the customer perimeter or, when cloud inference is explicitly enabled by the administrator, passes through a mandatory sanitization gateway that redacts sensitive data before any external API call and logs every outbound payload for audit. There is no vendor-operated multi-tenant component, no document persistence, and no telemetry.

THE ONE-SENTENCE VERSION

Your data stays on your server. The AI comes to the data — not the other way around.

This document describes the platform's capabilities (Section 2) and the technical controls that enforce its security principle (Sections 3–9).

2. Platform Capabilities

The platform is a modular stack of containerized services. Each capability below runs on the same hardened base architecture described in this document; modules can be enabled independently per deployment.

2.1 Executive Presentation Automation

LIVE

Generates C-level PowerPoint decks from a topic, pasted data, or uploaded Excel/CSV files. Nine layout types — bullet narratives, KPI stat rows, two-column comparisons, process chevrons, executive quotes, native pie/donut, bar, and line charts, and full-bleed photo slides — are assembled by a large language model into a consulting-grade storyline (situation → options → risks → recommendation → next steps). An in-browser editor allows per-slide regeneration, reordering, and layout changes before export. Optional web research enriches decks with cited, current market data; every chart carries a source footnote.

2.2 Corporate Template Engine

LIVE

Upload any corporate master deck (.pptx/.potx) once: the engine extracts the company's color scheme, fonts, and logo directly from the Office theme files and automatically detects light or dark branding. Every generated slide — backgrounds, cards, charts, accents — renders in the client's own corporate design. Templates are processed in memory only and can be replaced or removed at any time; no template content is stored on disk.

2.3 Report Mode — Periodic Financial Reporting

ROADMAP

A dedicated pipeline for quarterly and monthly business reports: P&L and KPI tables are ingested via the existing upload channel (Excel/CSV), then rendered as branded report decks with waterfall charts for result bridges, quarter-over-quarter delta bars, and formatted data tables. Designed for recurring board and investor reporting where figures are highly sensitive — the full pipeline runs inside the customer perimeter under the same data-residency controls as all other modules.

2.4 Strategic Quadrant Analysis

ROADMAP

A 2×2 matrix layout for portfolio and strategy work — positioning business units, competitors, or initiatives on configurable axes (e.g., growth vs. profitability, impact vs. effort). Quadrant charts render natively in PowerPoint with axis labels, quadrant titles, and positioned data points, matching the active corporate theme.

2.5 Workflow Automation with n8n

ROADMAP

n8n — a self-hosted, source-available workflow automation platform — connects the AI stack to the customer's existing systems: ERP exports, CRM events, email inboxes, file shares, and databases. Typical flows include scheduled report generation (data pulled, deck built, draft delivered for review), document intake and classification, and alerting. n8n runs as a container on the internal network with no external exposure; credentials for connected systems remain on the host.

2.6 Private AI Chat

LIVE

A self-hosted chat interface (Open WebUI) over local language models gives employees a ChatGPT-like experience with zero data leakage: prompts and documents never leave the perimeter, there is no per-seat licensing, and usage is unrestricted. Models are unloaded from memory when idle, keeping the infrastructure footprint minimal.

3. Deployment Model

The platform deploys as a set of containerized services on customer-owned infrastructure — on-premise server, datacenter, or customer-controlled VPS. There is no vendor-operated multi-tenant component.

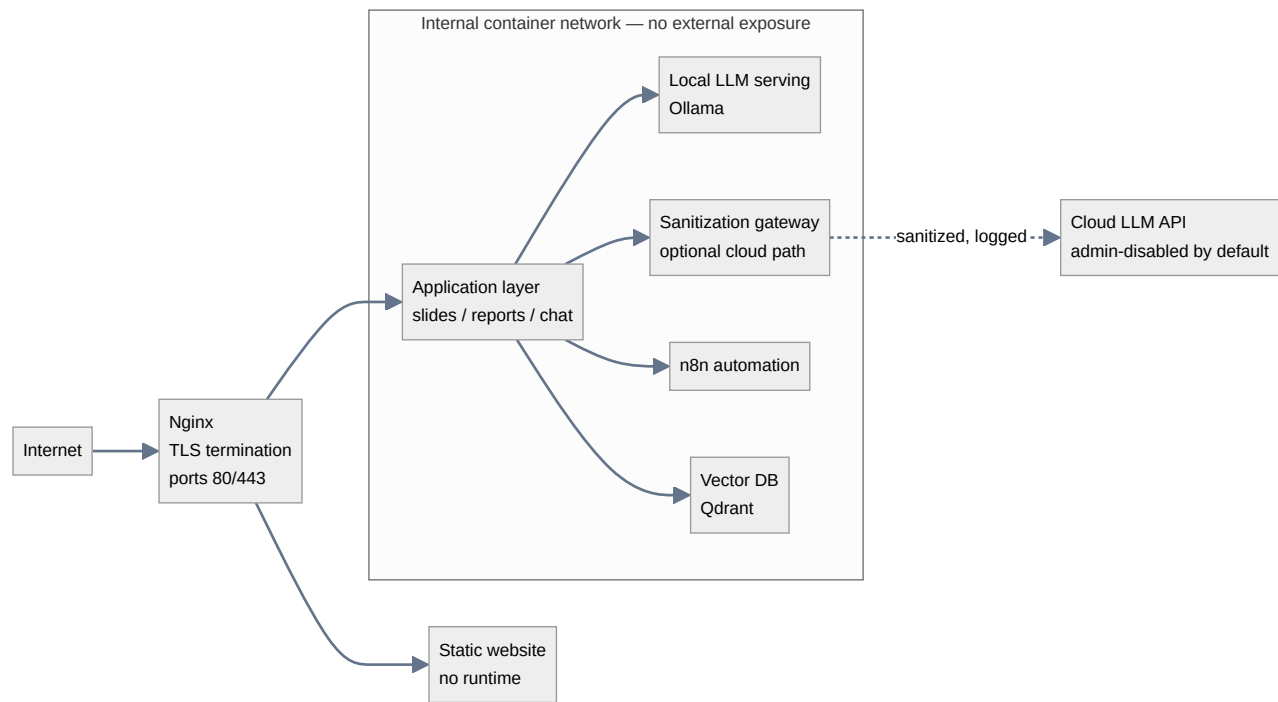


Figure 1 — Reference deployment architecture

Isolation properties:

- All application services bind to 127.0.0.1 or the internal container network. No AI service port is reachable from the public internet.
- Inter-service communication (application ↔ LLM ↔ database ↔ automation) never traverses public networks.
- Host firewall (UFW) admits only ports 22 (SSH), 80, and 443. Default policy: deny inbound.

4. Data Flow and Residency

4.1 Local processing (default)

User content is processed by AI models running inside the deployment. No content, prompts, or generated artifacts are transmitted externally. Generated files are produced in memory and streamed to the user's browser; the platform retains no document storage.

4.2 Sanitized cloud inference (optional, off by default)

When higher-quality output is required, an optional cloud provider may be enabled by the administrator. Before any request leaves the perimeter, a **sanitization gateway** enforces redaction:

Table 1 — Sanitization rules applied to all outbound payloads

CATEGORY	EXAMPLES	ACTION
Personal data	Email addresses, phone numbers	Redacted → [EMAIL] , [PHONE]
Financial identifiers	IBAN, card numbers	Redacted → [IBAN] , [CARD]
Credentials	API keys, tokens, secrets	Redacted → [API_KEY]
Infrastructure	IP addresses	Redacted → [IP]
Customer-defined terms	Client names, project codenames	Redacted → [REDACTED]

- The denylist of customer-specific terms is administrator-configurable per deployment.
- **Every outbound payload is written to an audit log** (content and timestamp), available for compliance review at any time.
- The cloud path can be disabled with a single configuration change, reverting the platform to fully local operation without data loss.

5. Access Control

- **Authentication:** password-based login with signed, expiring session cookies (12-hour lifetime, HTTPS-only, HttpOnly, SameSite).
- **Session integrity:** cookies are HMAC-signed with a server-side secret; tampering invalidates the session.
- **Administrative access (host):** SSH restricted to public-key authentication; password login disabled; brute-force protection via fail2ban with automatic IP banning.
- **Network access:** TLS 1.2+ for all client traffic (Let's Encrypt, automated renewal); HTTP permanently redirected to HTTPS.

Roadmap: OIDC/SAML single sign-on for enterprise directory integration.

6. Application Security

- **No persistent content storage:** presentations, reports, and uploaded files are processed in memory and delivered directly to the browser. Restarting the service removes all session data. Corporate templates are processed in memory only.
- **Security headers:** X-Frame-Options, X-Content-Type-Options, Referrer-Policy on all responses.
- **Upload validation:** file type allowlists, size limits, in-memory parsing only.
- **Dependency management:** containerized builds from pinned dependency versions; the host receives automatic security patches (unattended-upgrades).

- **TLS certificates:** automated issuance and renewal; no long-lived certificate risk.

7. GDPR / DSGVO Alignment

Table 2 — GDPR requirements mapped to platform implementation

REQUIREMENT	IMPLEMENTATION
Data minimization	No document persistence; in-memory processing only
Storage limitation	Session data expires with service restart
Data residency	All processing within customer-chosen jurisdiction; no third-country transfer in default configuration
Processor transparency	Optional cloud path documented, sanitized, logged, and admin-controlled
Auditability	Complete outbound payload log; access logs via web server
Right to erasure	Trivially satisfied: no user content is retained

A Data Processing Agreement (DPA / Auftragsverarbeitungsvertrag) is available on request.

8. Availability and Operations

- **Resource isolation:** container memory/CPU limits prevent any single service from exhausting host resources.
- **Automatic recovery:** all services restart automatically on failure or host reboot.
- **Patch management:** unattended security updates for the host OS; container updates via versioned rebuilds with rollback capability.
- **Backup posture:** the platform is stateless by design; only configuration files require backup (< 1 MB).
- **Capacity:** the reference stack runs comfortably on 2 vCPU / 8 GB RAM with cloud inference; local-model chat at scale is the only workload that benefits from an upgrade (4 vCPU / 16 GB). No GPU is required.

9. Summary for Security Reviewers

1. No customer content leaves customer-controlled infrastructure in the default configuration.
2. The optional cloud inference path is sanitized, logged, and administrator-controlled.
3. No document persistence exists anywhere in the platform.
4. All network exposure is limited to TLS-terminated web traffic; AI services are network-internal.

5. Administrative access is key-based, rate-limited, and monitored.
6. Automation credentials (n8n) remain on the host; workflows run on the internal network only.

VectorMATRIX — Self-Hosted Enterprise AI

contact@vectormatrix.io · vectormatrix.io · Germany & Worldwide

This document describes the reference architecture as deployed. Customer-specific deployments may vary; a deployment-specific security review is available as part of the Assessment engagement.