vM VECTORMATRIX

SICHERHEITS- & PLATTFORM-WHITEPAPER

Selbst gehostete Enterprise-KI, gebaut für Vertrauen

Architektur, Datenresidenz-Kontrollen und Plattformfunktionen des VectorMATRIX Self-Hosted-KI-Stacks — für Sicherheits-, Beschaffungs- und Compliance-Prüfungen in Unternehmen.

DSGVO-konform | Keine Dokumentenspeicherung | Kundenkontrollierte Infrastruktur | Deutschland & weltweit

VectorMATRIX — Self-Hosted Enterprise AI
contact@vectormatrix.io · vectormatrix.io

Version 2.0
Juli 2026

Inhaltsverzeichnis

1. Zusammenfassung
2. Plattformfunktionen
    - 2.1 Automatisierte Management-Präsentationen
    - 2.2 Corporate-Template-Engine
    - 2.3 Report Mode — Periodische Finanzberichte
    - 2.4 Strategische Quadranten-Analyse
    - 2.5 Workflow-Automatisierung mit n8n
    - 2.6 Privater KI-Chat
3. Bereitstellungsmodell
4. Datenfluss und Datenresidenz
5. Zugriffskontrolle
6. Anwendungssicherheit
7. DSGVO-Konformität
8. Verfügbarkeit und Betrieb
9. Zusammenfassung für Sicherheitsprüfer

1. Zusammenfassung

Die VectorMATRIX-Plattform stellt KI-Funktionen — Präsentationsautomatisierung, periodische Berichterstellung, Workflow-Intelligenz und privaten KI-Chat — vollständig in kundenkontrollierter Infrastruktur bereit. Die Architektur folgt einem einzigen Grundprinzip: **Kundeninhalte verlassen niemals kundenkontrollierte Systeme.**

Die gesamte KI-Verarbeitung erfolgt entweder auf lokalen Modellen innerhalb des Kundenperimeters oder — sofern Cloud-Inferenz vom Administrator ausdrücklich aktiviert wird — über ein verpflichtendes Anonymisierungs-Gateway, das sensible Daten vor jedem externen API-Aufruf entfernt und jede ausgehende Anfrage zu Prüfzwecken protokolliert. Es gibt keine vielmandantenfähige Komponente beim Anbieter, keine Dokumentenspeicherung und kein Tracking.

> **IN EINEM SATZ**
>
> Ihre Daten bleiben auf Ihrem Server. Die KI kommt zu den Daten — nicht umgekehrt.

Dieses Dokument beschreibt die Funktionen der Plattform (Abschnitt 2) sowie die technischen Maßnahmen, die ihr Sicherheitsprinzip durchsetzen (Abschnitte 3–9).

2. Plattformfunktionen

Die Plattform ist ein modularer Stack containerisierter Dienste. Jede der folgenden Funktionen läuft auf derselben gehärteten Basisarchitektur, die in diesem Dokument beschrieben wird; Module können je Bereitstellung unabhängig voneinander aktiviert werden.

2.1 Automatisierte Management-Präsentationen LIVE

Erstellt PowerPoint-Decks auf C-Level aus einem Thema, eingefügten Daten oder hochgeladenen Excel-/CSV-Dateien. Neun Layout-Typen — Aufzählungen, KPI-Zeilen, zweispaltige Vergleiche, Prozess-Pfeile, Management-Zitate, native Kreis-/Ring-, Balken- und Liniendiagramme sowie ganzseitige Fotofolien — werden von einem Sprachmodell zu einer beratungsgerechten Dramaturgie zusammengesetzt (Ausgangslage → Optionen → Risiken → Empfehlung → nächste Schritte). Ein Editor im Browser ermöglicht das Regenerieren, Verschieben und Umgestalten einzelner Folien vor dem Export. Optionale Web-Recherche reichert Decks mit zitierten, aktuellen Marktdaten an; jedes Diagramm trägt eine Quellenangabe.

2.2 Corporate-Template-Engine LIVE

Laden Sie einmalig ein Corporate-Master-Deck (.pptx/.potx) hoch: Die Engine extrahiert Farbschema, Schriftarten und Logo direkt aus den Office-Theme-Dateien und erkennt automatisch helle oder dunkle Markenwelten. Jede erzeugte Folie — Hintergründe, Karten, Diagramme, Akzente — erscheint im Corporate Design des Kunden. Vorlagen werden ausschließlich im Arbeitsspeicher verarbeitet und können jederzeit ersetzt oder entfernt werden; keine Vorlageninhalte werden auf der Festplatte gespeichert.

2.3 Report Mode — Periodische Finanzberichte ROADMAP

Eine eigene Pipeline für Quartals- und Monatsberichte: GuV- und KPI-Tabellen werden über den bestehenden Upload-Kanal (Excel/CSV) eingelesen und als Berichts-Decks im Corporate Design gerendert — mit Wasserfall-Diagrammen für Ergebnisbrücken, Quartalsvergleichs-Balken und formatierten Datentabellen. Konzipiert für wiederkehrende Vorstands- und Investorenberichte mit hochsensiblen Zahlen — die gesamte Pipeline läuft innerhalb des Kundenperimeters unter denselben Datenresidenz-Kontrollen wie alle anderen Module.

2.4 Strategische Quadranten-Analyse ROADMAP

Ein 2×2-Matrix-Layout für Portfolio- und Strategiearbeit — zur Positionierung von Geschäftseinheiten, Wettbewerbern oder Initiativen auf frei konfigurierbaren Achsen (z. B. Wachstum vs. Profitabilität, Wirkung vs. Aufwand). Quadranten-Diagramme werden nativ in PowerPoint gerendert — mit Achsenbeschriftungen, Quadranten-Titeln und positionierten Datenpunkten, passend zum aktiven Corporate Theme.

2.5 Workflow-Automatisierung mit n8n ROADMAP

n8n — eine selbst gehostete, quelloffene Workflow-Automatisierungsplattform — verbindet den KI-Stack mit den bestehenden Systemen des Kunden: ERP-Exporte, CRM-Ereignisse, E-Mail-Postfächer, Dateifreigaben und Datenbanken. Typische Abläufe sind die terminierte Berichtserstellung (Daten abrufen, Deck erstellen, Entwurf zur Prüfung bereitstellen), Dokumenteneingang und -klassifizierung sowie Benachrichtigungen. n8n läuft als Container im internen Netzwerk ohne externe Erreichbarkeit; Zugangsdaten für angebundene Systeme verbleiben auf dem Host.

2.6 Privater KI-Chat LIVE

Eine selbst gehostete Chat-Oberfläche (Open WebUI) über lokalen Sprachmodellen bietet Mitarbeitenden ein ChatGPT-ähnliches Erlebnis ohne Datenabfluss: Eingaben und Dokumente verlassen den Perimeter niemals, es fallen keine Lizenzkosten pro Arbeitsplatz an, und die Nutzung ist unbegrenzt. Modelle werden bei Inaktivität aus dem Arbeitsspeicher entladen — der Infrastruktur-Fußabdruck bleibt minimal.

3. Bereitstellungsmodell

Die Plattform wird als Satz containerisierter Dienste auf kundeneigener Infrastruktur bereitgestellt — On-Premise-Server, Rechenzentrum oder kundenkontrollierter VPS. Es gibt keine vielmandantenfähige Komponente beim Anbieter.

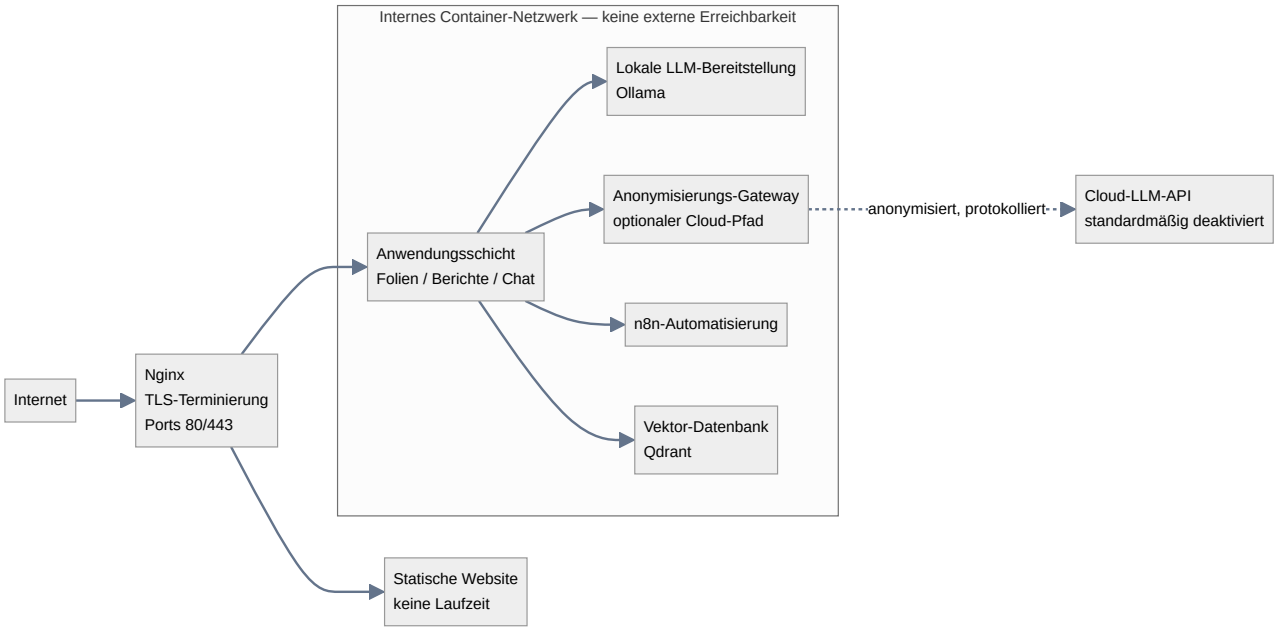

Abbildung 1 — Referenzarchitektur der Bereitstellung

Isolationseigenschaften:

- Alle Anwendungsdienste binden an `127.0.0.1` oder das interne Container-Netzwerk. Kein KI-Dienst-Port ist aus dem öffentlichen Internet erreichbar.
- Die Kommunikation zwischen den Diensten (Anwendung ↔ LLM ↔ Datenbank ↔ Automatisierung) durchquert niemals öffentliche Netzwerke.
- Die Host-Firewall (UFW) lässt nur die Ports 22 (SSH), 80 und 443 zu. Standardrichtlinie: eingehend verweigern.

4. Datenfluss und Datenresidenz

4.1 Lokale Verarbeitung (Standard)

Nutzerinhalte werden von KI-Modellen innerhalb der Bereitstellung verarbeitet. Keine Inhalte, Eingaben oder erzeugten Artefakte werden extern übertragen. Erzeugte Dateien entstehen im Arbeitsspeicher und werden direkt an den Browser des Nutzers übertragen; die Plattform speichert keine Dokumente.

4.2 Anonymisierte Cloud-Inferenz (optional, standardmäßig deaktiviert)

Wenn höhere Ausgabequalität erforderlich ist, kann der Administrator optional einen Cloud-Anbieter aktivieren. Bevor eine Anfrage den Perimeter verlässt, erzwingt ein **Anonymisierungs-Gateway** folgende Entfernungen:

Tabelle 1 — Anonymisierungsregeln für alle ausgehenden Anfragen

KATEGORIE	BEISPIELE	MASSNAHME
Personenbezogene Daten	E-Mail-Adressen, Telefonnummern	Entfernt → `[EMAIL]`, `[PHONE]`
Finanzkennungen	IBAN, Kartennummern	Entfernt → `[IBAN]`, `[CARD]`
Zugangsdaten	API-Schlüssel, Tokens, Secrets	Entfernt → `[API_KEY]`
Infrastruktur	IP-Adressen	Entfernt → `[IP]`
Kundendefinierte Begriffe	Kundennamen, Projekt-Codenamen	Entfernt → `[REDACTED]`

- Die Sperrliste kundenspezifischer Begriffe ist je Bereitstellung durch den Administrator konfigurierbar.
- **Jede ausgehende Anfrage wird in ein Audit-Log geschrieben** (Inhalt und Zeitstempel) und steht jederzeit für Compliance-Prüfungen zur Verfügung.
- Der Cloud-Pfad lässt sich mit einer einzigen Konfigurationsänderung deaktivieren — die Plattform kehrt ohne Datenverlust zum vollständig lokalen Betrieb zurück.

5. Zugriffskontrolle

- **Authentifizierung:** passwortbasierte Anmeldung mit signierten, ablaufenden Sitzungs-Cookies (12 Stunden Laufzeit, nur HTTPS, HttpOnly, SameSite).
- **Sitzungsintegrität:** Cookies sind mit einem serverseitigen Geheimnis HMAC-signiert; Manipulation macht die Sitzung ungültig.
- **Administrativer Zugang (Host):** SSH ausschließlich mit Public-Key-Authentifizierung; Passwort-Anmeldung deaktiviert; Brute-Force-Schutz über fail2ban mit automatischer IP-Sperrung.
- **Netzwerkzugang:** TLS 1.2+ für den gesamten Client-Verkehr (Let's Encrypt, automatische Verlängerung); HTTP wird dauerhaft auf HTTPS umgeleitet.

Roadmap: OIDC/SAML Single Sign-On für die Anbindung an Unternehmensverzeichnisse.

6. Anwendungssicherheit

- **Keine dauerhafte Inhaltsspeicherung:** Präsentationen, Berichte und hochgeladene Dateien werden im Arbeitsspeicher verarbeitet und direkt an den Browser ausgeliefert. Ein Neustart des Dienstes entfernt alle Sitzungsdaten. Corporate-Vorlagen werden nur im Arbeitsspeicher verarbeitet.
- **Sicherheits-Header:** X-Frame-Options, X-Content-Type-Options und Referrer-Policy auf allen Antworten.
- **Upload-Validierung:** Positivlisten für Dateitypen, Größenlimits, Verarbeitung nur im Arbeitsspeicher.
- **Abhängigkeitsmanagement:** containerisierte Builds mit festgelegten Versionsständen; der Host erhält automatische Sicherheitsupdates (unattended-upgrades).
- **TLS-Zertifikate:** automatisierte Ausstellung und Verlängerung; kein Risiko langlebiger Zertifikate.

7. DSGVO-Konformität

Tabelle 2 — DSGVO-Anforderungen im Verhältnis zur Plattform-Umsetzung

ANFORDERUNG	UMSETZUNG
Datenminimierung	Keine Dokumentenspeicherung; Verarbeitung nur im Arbeitsspeicher
Speicherbegrenzung	Sitzungsdaten verfallen mit dem Neustart des Dienstes
Datenresidenz	Verarbeitung ausschließlich in der vom Kunden gewählten Jurisdiktion; kein Drittlandtransfer in der Standardkonfiguration
Transparenz der Auftragsverarbeitung	Optionaler Cloud-Pfad dokumentiert, anonymisiert, protokolliert und administrativ kontrolliert
Prüfbarkeit	Vollständiges Protokoll ausgehender Anfragen; Zugriffsprotokolle über den Webserver
Recht auf Löschung	Trivial erfüllt: Es werden keine Nutzerinhalte vorgehalten

Ein Auftragsverarbeitungsvertrag (AVV) ist auf Anfrage erhältlich.

8. Verfügbarkeit und Betrieb

- **Ressourcentrennung:** Speicher- und CPU-Limits der Container verhindern, dass ein einzelner Dienst die Host-Ressourcen erschöpft.
- **Automatische Wiederherstellung:** Alle Dienste starten nach einem Fehler oder Host-Neustart automatisch neu.

- **Patch-Management:** unbeaufsichtigte Sicherheitsupdates für das Host-Betriebssystem; Container-Updates über versionierte Builds mit Rollback-Möglichkeit.
- **Backup-Konzept:** Die Plattform ist bewusst zustandslos; nur Konfigurationsdateien erfordern ein Backup (< 1 MB).
- **Kapazität:** Der Referenz-Stack läuft mit Cloud-Inferenz problemlos auf 2 vCPU / 8 GB RAM; lediglich lokaler Modell-Chat im größeren Maßstab profitiert von einer Erweiterung (4 vCPU / 16 GB). Eine GPU ist nicht erforderlich.

9. Zusammenfassung für Sicherheitsprüfer

1. In der Standardkonfiguration verlassen keine Kundeninhalte die kundenkontrollierte Infrastruktur.
2. Der optionale Cloud-Inferenz-Pfad ist anonymisiert, protokolliert und administrativ kontrolliert.
3. Es existiert keinerlei Dokumentenspeicherung in der gesamten Plattform.
4. Die gesamte Netzwerk-Exposition beschränkt sich auf TLS-terminierten Web-Verkehr; KI-Dienste sind netzwerkintern.
5. Der administrative Zugang ist schlüsselbasiert, ratenbegrenzt und überwacht.
6. Zugangsdaten der Automatisierung (n8n) verbleiben auf dem Host; Workflows laufen ausschließlich im internen Netzwerk.

VectorMATRIX — Self-Hosted Enterprise AI

contact@vectormatrix.io · vectormatrix.io · Deutschland & weltweit

Dieses Dokument beschreibt die Referenzarchitektur im bereitgestellten Zustand. Kundenspezifische Bereitstellungen können abweichen; eine bereitstellungsspezifische Sicherheitsprüfung ist Teil des Assessment-Angebots.