Technische Dokumentation

Version: 3.0 vom 05.08.2026

Slideroom GmbH
Gewerbepark Cité 4
76532 Baden-Baden

Tel.: 07221 40 37 633
info@slideroom.de
www.slideroom.de

Grenke Bank AG
IBAN: DE21 2013 0400 0060 3067 27
BIC: GREB DEH1 XXX

USt-IdNr. DE367213821
Amtsgericht Mannheim
HRB 750 207

Geschäftsführer:
Christian Vollmer
Maximilian Supper

Management Summary für IT-Abteilungen und Sicherheitsverantwortliche

1. Hosting, Infrastruktur & Compliance

- **Anwendungsbetrieb und persistente Datenhaltung in Microsoft Azure / Germany West Central**; KI-Verarbeitung innerhalb der europäischen Microsoft Azure Data Zone.
- **Moderne Architektur:** Kotlin/Spring Boot Backend, Vue.js Frontend, PostgreSQL, Azure Blob Storage, Nginx Reverse Proxy.
- **Serverzugriffe** ausschließlich via SSH-Key-Authentifizierung (nur CTO + Vertretung).
- **Getrennte Entwicklungs-, Test- und Produktivumgebungen**

2. Datenschutz & Sicherheit (DSGVO Art. 25 & 32)

Vertraulichkeit

- **Mandantentrennung:** serverseitige Zuordnung aller Daten und Zugriffe zur jeweiligen Organisation; Dateien in organisationsbezogenen Blob-Containern.
- **Rollenbasierte Zugriffskontrolle** (Admin/User)
- **Keine unbefugten Entwicklerzugriffe** (Need-to-Know-Prinzip)
- **Physische Sicherheit** durch Microsoft Azure (Videoüberwachung, Zutrittskontrollen)

Pseudonymisierung

- **Alle Dateien werden pseudonymisiert gespeichert**
  → keine inhaltsbezogene Suche möglich
- **Analysedaten** nur anonymisiert/pseudonymisiert (Opt-in)

Integrität

- **Dateiintegrität** durch Checksummen
- **Kommunikation** ausschließlich über HTTPS/TLS

Verfügbarkeit

- **Vier automatisierte Backups täglich;** Wiederherstellung auf Basis der vorhandenen Sicherungen.
- **Soft-Delete & Versionierung** von Dateien
- **Disaster Recovery:** Plan bis Q4 2026 einschließlich vollständigem Wiederherstellungstest.

KI-Verarbeitung

- **Funktionen mit künstlicher Intelligenz basieren auf Microsoft Azure AI Foundry und Azure OpenAI** mit einer aufgaben- und komplexitätsabhängigen Modellauswahl.
- **Verarbeitung erfolgt über Data Zone Standard** (EU)
- **Prompts, Eingaben und Modellantworten** werden ausschließlich innerhalb der europäischen Microsoft Azure Data Zone verarbeitet.
- **Übermittelte Inhalte** werden nicht zum Training oder zur Verbesserung der zugrundeliegenden KI-Modelle verwendet.
- **Tool-Zugriffe:** requestbezogen, serverseitig validiert und ausschließlich unter den Berechtigungen des authentifizierten Benutzers ausgeführt.
- **Menschliche Kontrolle:** KI-Ausgaben und lokale Änderungen bleiben sichtbar und reversibel; die dauerhafte Speicherung erfolgt durch den Benutzer.

3. Authentifizierung & Autorisierung

- **Microsoft Single Sign-On (OAuth2)** vollständig integriert
- **Alternative Login-Option:** E-Mail + Passwort (mit starken Anforderungen)
- **Mehrfaktor-Authentifizierung (MFA)** optional
- **Tokens in Secure** / HttpOnly / SameSite = Strict Cookies
- **Tokens** sind gerätegebunden → Schutz gegen Token-Theft
- **Session-Timeouts** klar definiert (5 Minuten / 1 Tag / 1 Jahr)

4. Datenverarbeitung & Datenspeicherung

Welche Daten werden verarbeitet?

- **Benutzerdaten: Name und E-Mail; beim E-Mail-Login wird ausschließlich ein Passwort-Hash gespeichert**
- **Login- & Gerätedaten:** für Session-Handhabung
- **Dateien & Präsentationen:** mandantenisoliert, pseudonymisiert

Wo werden Daten gespeichert?

- **PostgreSQL (Azure):** strukturierte Daten; Schutz der Datenbank auf Speicherebene durch Encryption at Rest.
- **Azure Blob Storage:** Dateien, nur für autorisierte Benutzer zugänglich, versioniert + Checksummen

5. Betrieb, Monitoring & Updates

- **Testsystem:** automatisiertes Deployment; Produktivsystem: manuell/kontrolliert
- **Monitoring:** Azure Monitor, Sentry, Prometheus und Grafana mit Alarmierung bei kritischen Betriebszuständen.
- **Automatische Sicherheitsupdates** für OS und Komponenten
- **Vollständige Protokollierung** aller administrativen Zugriffe

6. Technische Sicherheitsmaßnahmen

- **HTTPS/TLS + HSTS und MFA-Unterstützung**
- **CSRF- & XSS-Schutz** durch Frameworks + Validierung
- **SQL-Injection-Schutz** (Hibernate/JPA)
- **Rate Limiting** gegen Missbrauch
- **Geplante Integration** von Azure DDoS Protection
- **Schwachstellenmanagement:** systemweite Prüfungen mit OWASP Dependency-Check, Trivy und Gitleaks sowie eine CycloneDX-SBOM je Produktversion.
- **Strikte Beschränkung** administrativer Schnittstellen

7. Risikoübersicht & Gegenmaßnahmen

Risiko	Schutzmaßnahmen
Unbefugter Zugriff	SSO, MFA, starke Passwörter, Token-Security
Datenmanipulation	Checksummen, HTTPS/TLS
Datenverlust	Vier automatisierte Backups täglich, Versionierung
DDoS	Rate Limiting + geplante Azure DDoS Protection
Web-Angriffe (XSS, SQLi)	Framework-basierte Security + Validierung
Fehlkonfiguration	Getrennte Umgebungen, Best Practices

Zusammenfassung

slideroom wurde speziell für den professionellen Unternehmenseinsatz entwickelt und erfüllt höchste Anforderungen an Datenschutz, Informationssicherheit, Compliance und die verantwortungsvolle Nutzung von Künstlicher Intelligenz. Die Plattform orientiert sich konsequent an den Vorgaben des EU AI Acts, der DSGVO sowie etablierten Best Practices für Enterprise-Software und Microsoft-basierte IT-Infrastrukturen. slideroom verfolgt dabei einen **„Security, Privacy & Compliance by Design"-Ansatz.** Datenschutz, Informationssicherheit und regulatorische Anforderungen werden bereits bei der Entwicklung neuer Funktionen berücksichtigt und kontinuierlich überprüft.

EU AI Act & verantwortungsvolle KI-Nutzung

Die in slideroom integrierten KI-Funktionen dienen ausschließlich der Unterstützung von Anwendern, beispielsweise bei Textoptimierungen, Übersetzungen, Formulierungsvorschlägen oder Designempfehlungen. Die KI trifft dabei keine automatisierten Entscheidungen mit rechtlicher oder vergleichbarer Wirkung auf Personen.

Alle KI-generierten Inhalte werden dem Nutzer transparent angezeigt und können jederzeit geprüft, angepasst oder verworfen werden. Die finale Entscheidung verbleibt stets beim Anwender. Dadurch werden zentrale Anforderungen des EU AI Acts hinsichtlich Transparenz, menschlicher Kontrolle und verantwortungsvoller KI-Nutzung erfüllt.

slideroom beobachtet die laufende Weiterentwicklung des EU AI Acts kontinuierlich und passt Prozesse sowie technische Maßnahmen bei Bedarf an neue regulatorische Anforderungen an.

Datenschutz & DSGVO

slideroom erfüllt höchste technische und organisatorische Datenschutzanforderungen für den Unternehmenseinsatz.

Zu den wesentlichen Datenschutzmaßnahmen gehören:

- DSGVO-konforme Datenverarbeitung
- Berücksichtigung von „Privacy by Design“ und „Privacy by Default“
- Pseudonymisierte Speicherung von Dateien und Inhaltsdaten, soweit technisch und fachlich sinnvoll
- Datensparsame Verarbeitung nach dem Need-to-Know-Prinzip
- Verschlüsselte Datenübertragung mittels aktueller TLS-Standards
- Verschlüsselte Datenspeicherung („Encryption at Rest“)
- Klare Mandantentrennung zwischen Kundenorganisationen
- Rollen- und Berechtigungskonzepte
- Unterstützung von Auftragsverarbeitungsverträgen (AVV)

Informationssicherheit

slideroom wurde für den sicheren Betrieb in Unternehmensumgebungen entwickelt und verfügt über eine mehrschichtige Sicherheitsarchitektur.

Dazu gehören unter anderem:

- Getrennte Entwicklungs-, Test- und Produktivumgebungen
- Mehrstufige Zugriffsschutzmechanismen
- Regelmäßige Sicherheitsupdates und Wartungsprozesse
- Backup- und Wiederherstellungsstrategien
- Protokollierung sicherheitsrelevanter Vorgänge
- Mandantenfähige Systemarchitektur
- Schutz vor unbefugten Zugriffen und Datenmanipulationen
- Kontinuierliche Überwachung kritischer Systemkomponenten

Microsoft-Integration & Enterprise Readiness

slideroom ist speziell für Microsoft-basierte Unternehmensumgebungen konzipiert und lässt sich nahtlos in bestehende IT-Landschaften integrieren.

Die Plattform bietet unter anderem:

- Single Sign-on (SSO) über Microsoft Entra ID (Azure AD)
- Automatische Benutzer- und Gruppenverwaltung über Microsoft Entra Connect Sync
- Direkte Integration in Microsoft PowerPoint
- Zentrale Benutzerverwaltung für IT-Administratoren
- Einfache Skalierung von Pilotprojekten bis zum unternehmensweiten Rollout
- Minimalen Betriebs- und Administrationsaufwand auf Kundenseite

Hosting & Datenstandort

Der Anwendungsbetrieb und die persistente Datenhaltung von slideroom erfolgen innerhalb der Microsoft Azure Cloud in Deutschland. Die KI-Inferenz wird innerhalb der europäischen Microsoft Azure Data Zone ausgeführt.

Wesentliche Merkmale:

- Hosting in deutschen Microsoft Azure-Rechenzentren
- Persistente Datenhaltung innerhalb Deutschlands; KI-Inferenz innerhalb der europäischen Microsoft Azure Data Zone
- Moderne Cloud-Infrastruktur auf Enterprise-Niveau
- Hohe Verfügbarkeit, Skalierbarkeit und Ausfallsicherheit
- Regelmäßige Sicherungen und Disaster-Recovery-Konzepte

Fazit

slideroom bietet Unternehmen eine sichere, datenschutzkonforme und zukunftssichere Plattform für die Nutzung von KI direkt in Microsoft PowerPoint.

Durch die Kombination aus DSGVO-konformer Datenverarbeitung, Hosting in deutschen Azure-Rechenzentren, moderner Sicherheitsarchitektur, mandantenfähigem Betrieb, Microsoft-SSO-Integration sowie der konsequenten Berücksichtigung der Anforderungen des EU AI Acts erfüllt slideroom höchste Anforderungen an Datenschutz, Informationssicherheit und Compliance im Unternehmensumfeld.

Einleitung

Diese Dokumentation beschreibt die Architektur, Sicherheit und den Betrieb des PowerPoint Add-Ins slideroom.

Hinweis: Diese Dokumentation wird regelmäßig aktualisiert, um Änderungen an der Architektur, Sicherheitsmaßnahmen und anderen relevanten Informationen zu berücksichtigen. Eine Überprüfung und Aktualisierung erfolgt mindestens jährlich oder nach bedeutenden Änderungen an der Applikation.

Übersicht der Applikation

- **Name der Applikation:** slideroom
- **Technologie-Stack:** Vue.js mit TypeScript im Frontend, Kotlin mit Spring Boot im Backend, Deployment via Gradle. Nginx als Reverse Proxy.

Technische Dokumentation

1. Architektur

Präsentationsverarbeitung

- **Präsentationsdateien** werden zur Analyse und Verarbeitung an einen dedizierten Worker-Service übergeben
- Der Worker nutzt **Aspose.Slides for Java** zur Verarbeitung von Microsoft PowerPoint-Dateien (.pptx)
- Der Dienst wird innerhalb der von slideroom betriebenen Microsoft-Azure-Infrastruktur als Azure App Job in der Region **Germany West Central** ausgeführt
- Die **technische Verarbeitung der PowerPoint-Datei durch den Worker** erfolgt innerhalb der von slideroom betriebenen Azure-Infrastruktur. Für ausdrücklich aufgerufene KI-Funktionen können relevante Inhalte gemäß Abschnitt 3.1 an den Azure OpenAI Service innerhalb der europäischen Microsoft Azure Data Zone übermittelt werden. Eine Übermittlung an weitere KI-Dienste oder Consumer-Angebote findet nicht statt.

- Die **Verarbeitung** umfasst unter anderem die Analyse von Präsentationen, die Extraktion relevanter Metadaten sowie weitere für die Funktionalität der Anwendung erforderliche Verarbeitungsschritte
- **Datenbank:** PostgreSQL für strukturierte Daten und Microsoft Blob Storage für Dateispeicherung.
- **Protokolle:**
  - **HTTPS** zur sicheren Kommunikation
  - **REST API** zwischen Frontend und Backend sowie Nutzung der externen REST APIs von Pexels und Pixabay (optional wählbar für Benutzer).

2. Sicherheitsmaßnahmen (gem. Art. 32 DSGVO)

2.1 Vertraulichkeit (Art. 32 Abs. 1 lit. b DSGVO)

Physische Sicherheit / Zutrittskontrolle

- **Hosting** in Azure-Rechenzentren in Deutschland. Höchster Sicherheitsstandard durch Azure.
- **Keine unbefugten Personen vor Ort:** Zutritt nur für autorisierte Personen.
- **Alarm- und Videoüberwachungssysteme** durch den Cloud-Anbieter.

Zugangskontrolle

- **Keine unbefugte Systembenutzung:** sichere Passwörter, automatische Sperrmechanismen, Zwei-Faktor-Authentifizierung.
- **Verschlüsselung von Datenträgern** auf den Entwicklungs- und Administrationsgeräten.
- **Für unsere Infrastruktur:** SSH nur mit Schlüssel-Authentifizierung (nur CTO und Vertretung).

Zugriffskontrolle

- **Rollenbasierte Zugriffskonzepte** innerhalb der Anwendung (aktuell User/Admin) pro Organisation.
- **Entwickler-/Mitarbeiterzugriff** auf Infrastruktur nur für notwendige Wartung (Need-to-know-Prinzip).
- **Protokollierung** sämtlicher Zugriffe auf administrative Systeme.

Schlüssel- und Secret-Management

- **Administrative Zugangsdaten:** zentrale Verwaltung in 1Password; Zugriff ausschließlich für berechtigte Personen nach dem Need-to-know-Prinzip.
- **Technische Secrets:** Speicherung von API-Schlüsseln, Datenbankzugängen und Client-Secrets im Azure Key Vault.
- **Dienstauthentifizierung:** wo möglich über Managed Identities anstelle statischer Zugangsdaten; Schlüssel und Zugangsdaten werden standardmäßig nach 90 Tagen erneuert.

Trennungskontrolle

- **Mandantentrennung:** Jeder Zugriff wird im Backend anhand der authentifizierten Identität und Organisation autorisiert. Der Mandantenkontext kann nicht über frei wählbare Client- oder Modellparameter verändert werden.
- **Separate Entwicklungs-, Test- und Produktivumgebungen** (Sandboxing).

2.2 Pseudonymisierung (Art. 32 Abs. 1 lit. a; Art. 25 Abs. 1 DSGVO)

- **Hochgeladene Dateien** werden pseudonymisiert gespeichert, sodass keine gezielte Suche nach Inhalten möglich ist.
- **Personenbezogene Daten** werden nur mit separaten Referenzen (IDs) verknüpft.
- Wenn möglich und sinnvoll, werden **Daten in der Datenbank pseudonymisiert**.
- **Analytics Tools** bekommen (sofern zugestimmt) nur pseudonymisierte Issue Reports.

2.3 Integrität (Art. 32 Abs. 1 lit. b DSGVO)

Weitergabekontrolle

- **Sichere Übertragung über HTTPS/TLS**
- **Für Dateien werden Checksummen erstellt** die dann beim Abruf überprüft werden um eine Manipulation zu verhindern.

2.4 Verfügbarkeit und Belastbarkeit (Art. 32 Abs. 1 lit. b DSGVO)

Verfügbarkeitskontrolle

- **Backups:** viermal täglich automatisierte Sicherung der virtuellen Maschine einschließlich Datenträgern über Azure Backup; regelmäßige Sicherung der Daten im Azure Blob Storage.
- **Firewall, Virenschutz, USV** in Azure-Infrastruktur.

Rasche Wiederherstellbarkeit (Art. 32 Abs. 1 lit. c DSGVO)

- **Disaster-Recovery-Plan:** Fertigstellung bis Q4 2026 und Validierung durch einen vollständigen Wiederherstellungstest. Berücksichtigt werden insbesondere Ransomware, versehentliches Löschen, Fehlkonfigurationen und kompromittierte Zugangsdaten.
- Rollback-Mechanismen durch regelmäßige Backups.
- **Versionierung** und Soft-Delete auf Dateiebene.
- **Komponentenweise Wiederherstellung:** Die Datenbank wird auf einem separaten Datenträger und in einem eigenen Container betrieben.

2.5 Verfahren zur regelmäßigen Überprüfung, Bewertung und Evaluierung (Art. 32 Abs. 1 lit. d; Art. 25 Abs. 1 DSGVO)

Datenschutz-Management

- **Regelmäßige Kontrollen** der Datenverarbeitungsprozesse, Dokumentation und Anpassung an aktuelle DSGVO-Anforderungen.

Incident-Response-Management

- **Incident Response:** Sicherheitsrelevante Ereignisse einschließlich KI-bezogener Tool-Fehler und technischer Ausnahmen werden im bestehenden Incident-Management analysiert, bewertet und bearbeitet.

2.6 Datenschutzfreundliche Voreinstellungen (Art. 25 Abs. 2 DSGVO)

- **Standardmäßig minimale Datenerhebung**, Opt-in für erweiterte Analysen.

Auftragskontrolle

- **Keine Weitergabe oder Verarbeitung** ohne Weisung des Auftraggebers (Auftraggeber = Benutzer innerhalb der Organisation).
- Streng geregelte Auswahl etwaiger Subdienstleister (z. B. Cloud-Provider).

3. Datenfluss und Verarbeitung

Erhobene Daten

- **Personenbezogene Daten:** Name und E-Mail werden im Klartext in der Datenbank gespeichert. Vom Passwort wird ausschließlich ein Hashwert gespeichert; das Passwort selbst wird nicht gespeichert. Durch die notwendige Zugehörigkeit zu einer Organisation ist indirekt auch die Firmenzugehörigkeit bekannt.

- **Nutzungsdaten:** Login-Zeiten und Geräteinformationen über die Laufzeit der Session. Optional Opt-in für Seitenaufrufe und Klickverhalten (pseudonymisiert).
- **Dateien und Inhalte:** Hochgeladene Präsentationen und Medien und Metadaten.

Datenspeicherung

- **PostgreSQL:** Strukturierte Benutzerdaten und Metadaten. Name und E-Mail werden in der Datenbank im Klartext gespeichert. Die Datenbank ist auf Speicherebene durch „Encryption at Rest“ geschützt. Vom Passwort wird ausschließlich der Hashwert gespeichert.
- **Microsoft Blob Storage:** Medieninhalte und Präsentationen, zugänglich nur für autorisierte Benutzer. Zusätzlich werden die Dateien vor der Ablage pseudonymisiert sodass nicht gezielt nach Inhalten gesucht werden kann. Gegen Manipulation der Dateien werden Checksummen berechnet und gespeichert.

Datenverarbeitung

- **Benutzerregistrierung und Authentifizierung:** Per E-Mail/Passwort oder Microsoft SSO. Bei der Anmeldung per E-Mail/Passwort verarbeitet Spring Security das Passwort. In der Datenbank wird ausschließlich der Passwort-Hash gespeichert.
- **Dateiverwaltung:** Dateien können innerhalb der Applikation hochgeladen, gespeichert und abgerufen werden.
- **Analyse und Monitoring:** Seitenaufrufe und Klickverhalten (Opt-in), anonymisierte oder pseudonymisierte (wenn zugestimmt) Fehlerberichte und Logs zur Verbesserung der Anwendung.

3.1 KI-gestützte Funktionen und Datenverarbeitung

Eingesetzte KI-Dienste und Modelle

- **Für KI-gestützte Funktionen** nutzt slideroom Microsoft Azure AI Foundry sowie den Azure OpenAI Service.
- **GPT-5-nano:** Klassifizierungsaufgaben, Prompt-Routing und Einschätzung der Aufgabenkomplexität zur automatischen Auswahl eines geeigneten Modells.
- **GPT-5-mini:** Spezialisierte KI-Funktionen wie Übersetzungen, Textaufgaben und Bildersuche.
- **GPT-5.6-luna:** Hauptmodell für die zentrale Agentensteuerung sowie für komplexere spezialisierte KI-Funktionen.
- **GPT-5.6-terra:** Eskalationsstufe für besonders komplexe Aufgaben. Die Übergabe kann automatisiert durch die Routing-Komponente oder das Hauptmodell erfolgen.

Hosting und Datenverarbeitung

- **Die KI-Dienste** werden innerhalb der Microsoft-Azure-Infrastruktur betrieben.
- **Die Verarbeitung** erfolgt über die Bereitstellungsoption Data Zone Standard (EU).
- **Prompts, Eingaben und Modellantworten** werden ausschließlich innerhalb der europäischen Microsoft Azure Data Zone verarbeitet.
- **Eine Verarbeitung im Rahmen der KI-Inferenz außerhalb der EU Data Boundary findet nicht statt.**

Datenschutz und Verwendung der Eingaben

- **Inhalte, die im Rahmen der KI-Funktionen verarbeitet werden** (z. B. Texte, Präsentationsinhalte oder Benutzereingaben), werden ausschließlich zur Bearbeitung der jeweiligen Anfrage verwendet.
- **Übermittelte Inhalte** werden nicht zum Training, zur Verbesserung oder zur Weiterentwicklung der zugrunde liegenden KI-Modelle verwendet.

- **Es erfolgt keine Verwendung von Kundendaten** für Lern- oder Trainingszwecke durch slideroom.
- **Keine eigenen Modelle:** slideroom erstellt, trainiert oder fine-tuned keine Foundation Models. Die Modelle werden als verwaltete Dienste über Microsoft Azure AI Foundry bereitgestellt.
- **Verantwortungsabgrenzung:** Modellgewichte und ursprüngliche Trainingsdaten liegen beim Modellanbieter und sind für slideroom und Endnutzer weder zugänglich noch exportierbar.
- **KI-Funktionen werden ausschließlich über Dienste von Microsoft Azure OpenAI umgesetzt.** Es erfolgt keine direkte Nutzung öffentlich zugänglicher KI-Dienste oder Consumer-Angebote.

Technische und organisatorische Maßnahmen

- **Zugriff auf KI-Funktionen** erfolgt ausschließlich für authentifizierte Benutzer innerhalb der jeweiligen Organisation.
- **Berechtigungsprüfung** erfolgt entsprechend der rollenbasierten Zugriffskontrolle der Anwendung.
- **Übertragung aller Daten** erfolgt verschlüsselt über HTTPS/TLS.
- **Verarbeitung von KI-Anfragen** erfolgt innerhalb der Sicherheits- und Compliance-Rahmenbedingungen des Microsoft Azure OpenAI Service.
- **Modellendpunkte und Secrets:** Anfragen werden ausschließlich über das Backend ausgeführt; API-Schlüssel und direkte Endpunktinformationen werden nicht an Clients oder Endnutzer übertragen.
- **Requestbezogene Tools:** Das Modell besitzt keinen dauerhaften System- oder MCP-Zugriff. Nur für den aktuellen Request freigegebene Tools werden bereitgestellt; Tool-Parameter werden vor der serverseitigen Ausführung validiert.
- **Berechtigungen und Mandantenkontext:** Jeder Tool-Aufruf wird unter den serverseitig ermittelten Rechten des aktuellen Benutzers ausgeführt. Das Modell kennt keine frei wählbaren Organisations- oder Mandanten-IDs und kann Rechte oder Mandantenzuordnung nicht verändern.

3.2 Embeddings und Vektorsuchindizes

Verwendung und Lebenszyklus

- **Embedding-Vektoren:** numerische Repräsentationen von Dokumentinhalten, die für eine semantische Suche verwendet werden. Aus den jeweils zu indexierenden Dokumentinhalten werden Embeddings erzeugt und im zugehörigen Suchindex gespeichert.
- **Medienpool:** Hochgeladene Dokumente werden eingebettet und in einem Suchindex erfasst, damit ihre Inhalte durchsuchbar sind. Der zugehörige Suchindexeintrag besteht so lange wie das Dokument im Medienpool und wird zusammen mit dem Dokument entfernt.
- **KI-Chat:** Dokumente, die innerhalb eines KI-Chats hochgeladen werden, werden ebenfalls eingebettet und für die Suche in einem sitzungsbezogenen Index erfasst. Dieser Suchindex besteht ausschließlich während der jeweiligen Sitzung und wird anschließend gelöscht.
- **Trennung der Suchindizes:** Medienpool- und Chat-Indizes werden ausschließlich im berechtigten Benutzer- und Mandantenkontext verwendet und nicht mandantenübergreifend zusammengeführt. Vektordaten werden nicht an Endgeräte ausgegeben.

Schutzbedarf und Restrisiko

- **Keine Anonymisierung:** Embeddings sind weder eine Einwegverschlüsselung noch anonymisierte Daten. Unter bestimmten Voraussetzungen können Inversions- oder Vergleichsangriffe semantische Inhalte, Eigennamen oder Teile des Ausgangstextes rekonstruieren beziehungsweise vermutete Inhalte bestätigen.
- **Risikoeinordnung:** Bei kurzen Texten, einzelnen Folien oder kleinen indexierten Abschnitten ist das Risiko höher. Bei umfangreichen Dokumenten ist eine vollständige oder wortgetreue Rekonstruktion typischerweise erschwert, jedoch nicht grundsätzlich ausgeschlossen.
- **Schutzmaßnahmen und Aufbewahrung:** Vektordaten unterliegen denselben Anforderungen an Zugriffsschutz, Mandantentrennung und Verschlüsselung wie die zugrunde liegenden Dokumente. Die Aufbewahrung folgt dem jeweiligen Zweck: Medienpool-Embeddings bestehen dokumentgebunden, Chat-Embeddings ausschließlich sitzungsbezogen.

4. Authentifizierung und Autorisierung

Authentifizierung

- **E-Mail und Passwort:** Passwortanforderungen (mind. 8 Zeichen, mit Sonderzeichen, Zahl und Großbuchstaben).
- **Microsoft SSO:** Über OAuth2. Nach Anmeldung erfolgt Token- und Sessionverarbeitung wie bei E-Mail-Authentifizierung. Die Sessioninfos von Microsoft werden nur zum Abruf von Benutzerdaten verwendet (bei der Registrierung) und zur Authentifizierung des Users und dann verworfen.
- **Mehrfaktor-Authentifizierung (MFA):** Kann optional aktiviert werden: entweder per E-Mail-Verifizierung oder TOTP (Authenticator-App). Organisationsrichtlinien, damit Administratoren MFA innerhalb einer Organisation erzwingen können, sind geplant.

Sicherheitsmechanismen

- **Token-Sicherheit:** Tokens zur Sessionzuordnung werden in HTTP-Only Cookies gespeichert (Secure, HttpOnly und SameSite=Strict).
- **Gerätebindung der Tokens:** Interne Mechanismen binden Tokens an das jeweilige Gerät.
- **Session-Management:** Session-Timeouts für „Angemeldet bleiben“ (1 Jahr) oder ohne (1 Tag). Auth-Cookies 5 Minuten, für den Refresh-Token („Angemeldet bleiben“) 1 Jahr, sonst "Session".

Autorisierung

- **Rollenbasierte Zugriffskontrolle:** Admin und User-Rollen; zukünftige Erweiterung geplant. Rollen und Berechtigungen sollen vom Administrator verwaltet werden können.
- **Berechtigungsprüfung:** Zugriffskontrolle bei jeder Anfrage; Prinzip des geringsten Privilegs.

5. Infrastruktur

Hosting-Umgebung

- **Cloud-Plattform:** Microsoft Azure in der Region Germany West Central für deutsche Datenschutzbestimmungen.

Netzwerkarchitektur

- **Virtuelles Netzwerk (VNet):** Eine VM, keine Segmentierung, NSG für Traffic-Kontrolle.
- **Load Balancer:** Aktuell nicht vorhanden, geplant für zukünftige Skalierung.

Sicherheitsmaßnahmen für VMs

- **Betriebssystem:** Ubuntu.
- **Zugriffskontrolle:** SSH-Zugriff mit Schlüssel-Authentifizierung, nur für Administratoren (CTO und Vertretung).

Verfügbarkeit und Skalierbarkeit

- **Skalierung:** Aktuell vertikal, horizontale Skalierung und Load Balancer bei wachsender Last geplant.

CI/CD-Prozesse

- **Build- und Deployment-Prozesse:** Azure DevOps als cloudbasierter Microsoft-Dienst; administrative Zugriffe ausschließlich über berechtigte Benutzerkonten mit rollenbasierter Zugriffskontrolle und MFA.
- **Deployment:** automatisiert in der Testumgebung und manuell sowie kontrolliert in der Produktivumgebung.
- **Administrative Serverzugriffe:** über das integrierte macOS-Terminal mit SSH-Schlüsselauthentifizierung. PuTTY und Jenkins kommen nicht zum Einsatz.

Backup und Disaster Recovery

- **Backups:** viermal täglich über Azure Backup für die VM und ihre Datenträger; regelmäßige Sicherungen des Azure Blob Storage.
- **Wiederherstellung:** Rollback und komponentenweise Wiederherstellung auf Basis der Sicherungen; Datenbank auf separatem Datenträger und in eigenem Container.
- **Disaster-Recovery-Plan: Fertigstellung und vollständiger Wiederherstellungstest bis Q4 2026.**

Wiederherstellung der KI-Funktionen

- **Reproduzierbare Deployments:** Modelle, Versionen, Regionen, Deploymentkonfigurationen, Systemanweisungen, Tooldefinitionen und Integrationslogik sind dokumentiert beziehungsweise versioniert. Die Neuerstellung eines Deployments wurde praktisch getestet.
- **Alternative Deployments:** Bei Ausfall oder Abkündigung können getestete alternative Modelle oder Deployments eingesetzt werden, ohne die serverseitigen Sicherheitskontrollen zu verändern.
- **Vorübergehende Störungen:** Automatische Wiederholungsversuche; bei anhaltendem Fehler kontrollierter Abbruch mit Benutzerhinweis. Die übrigen Anwendungsfunktionen bleiben grundsätzlich verfügbar.
- **Temporäre Sitzungsdaten:** werden nicht gesichert oder wiederhergestellt. Im KI-Chat erzeugte Vektordaten können aus dem Originaldokument neu erzeugt und die Verarbeitung in einer neuen Sitzung wiederholt werden.
- **RTO und RPO:** Für die KI-Funktion bestehen derzeit keine separaten formellen Zielwerte, da Model Deployments keine dauerhaften Kundendaten oder trainierten Zustände enthalten. Die Wiederherstellung erfolgt nach betrieblicher Priorität und verfügbarer Azure-Kapazität.

6. Betrieb und Wartung

Updates und Deployment

- **Update-Prozesse:** Nach Bedarf; ausführliche Tests auf einem Testsystem vor Deployment. Unit- und UI-Tests werden beim Build ausgeführt.

Lebenszyklus und Freigabe der KI-Integration

- **Verantwortung:** Training und Pflege der Foundation Models liegen beim Modellanbieter. slideroom verantwortet Auswahl, Konfiguration, Systemanweisungen, Tools, Integration, Tests, Freigabe, Überwachung, Versionswechsel und Außerbetriebnahme der eigenen KI-Integration.
- **Versionierung und Review:** Systemanweisungen, Tooldefinitionen, Modellkonfigurationen und Integrationslogik werden im Quellcode versioniert. Änderungen erfolgen über Pull Requests und werden vor dem Merge von mindestens einer weiteren berechtigten Person geprüft.
- **Modellupdates:** keine automatische Übernahme. Neue Modelle oder Versionen werden als separates Deployment in der Testumgebung auf technische Erreichbarkeit, Prompt- und Toolkompatibilität, Antwortqualität sowie Fehler- und Grenzfälle geprüft und anschließend gezielt freigegeben.
- **Rollback und Außerbetriebnahme:** Das bisherige Deployment kann während einer Übergangsphase als Rückfalloption erhalten bleiben, sofern es unterstützt wird. Nicht mehr benötigte Deployments, Berechtigungen und Ressourcen werden kontrolliert deaktiviert beziehungsweise entfernt.

Überwachung und Monitoring

- **Monitoring-Tools:** Azure Monitor, Sentry, Prometheus und Grafana.
- **Überwachung und Alarmierung:** VM- und Container-Verfügbarkeit, CPU, RAM, Festplattenbelegung, Datenbank- und Thread-Pool-Auslastung, Heap, Garbage Collection und API-Antwortzeiten. Alarme bei VM-Ausfall, CPU-Auslastung über 80 Prozent und Festplattenbelegung über 90 Prozent.

- **Aufbewahrung:** Monitoring- und Protokolldaten werden gemäß den konfigurierten Standardfristen der jeweiligen Systeme aufbewahrt und anschließend automatisch gelöscht.

KI-bezogene Sicherheitsereignisse

- **Erfasste Ereignisse:** insbesondere Fehler bei Tool-Aufrufen und technische Ausnahmen während der KI-Verarbeitung.
- **Bearbeitung:** Auffälligkeiten werden im bestehenden Incident-Management analysiert und bewertet; erforderliche Fehlerbehebungen oder Konfigurationsanpassungen werden zeitnah umgesetzt. Ein separater KI-spezifischer Incident-Response-Prozess besteht derzeit nicht.

Zugriffskontrolle und Berechtigungsmanagement

- **Administrative Zugriffe:** Streng kontrolliert, nur CTO und Vertretung.
- **Protokollierung:** Alle Zugriffe werden zur Nachvollziehbarkeit protokolliert.

Wartung und Sicherheitsupdates

- **Patch-Management:** Automatische Updates für Betriebssystem und Anwendungsspezifisches.
- **Sicherheitsmeldungen:** Administrator verfolgt Sicherheitsupdates regelmäßig (mindestens monatlich). Zusätzlich wird auf Sicherheitshinweise seitens Azure umgehend reagiert.

Schwachstellenmanagement und Software Bill of Materials

- **Systemweiter Umfang:** Betriebssystem, Container, Anwendungsabhängigkeiten und Cloud-Infrastruktur werden in das Schwachstellenmanagement einbezogen.
- **Automatisierte Prüfungen:** OWASP Dependency-Check für Anwendungsabhängigkeiten, Trivy für Container-Images und Gitleaks zur Erkennung versehentlich eingecheckter Secrets in der CI/CD-Pipeline.

- **Behandlung:** Erkannte Schwachstellen werden nach Kritikalität bewertet, priorisiert und durch Sicherheitsupdates, Konfigurationsänderungen oder den Austausch betroffener Komponenten behoben.
- **SBOM:** Für jede Produktversion wird eine Software Bill of Materials im CycloneDX-Format erstellt und gepflegt; sie wird auf Anfrage bereitgestellt.

7. Risiken und Maßnahmen

- **Unautorisierter Zugriff auf Benutzerdaten:** MFA, HTTPS-Verschlüsselung, Passwort-Hashes und Begrenzung von Login-Versuchen.
- **Cross-Site Scripting (XSS):** XSS-Schutz durch serverseitige Validierung der Eingaben. Browserseitig werden Escapemechanismen von vue.js verwendet.
- **Schwache Passwortsicherheit:** Passwortanforderungen und -feedback, optional MFA.
- **Man-in-the-Middle (MitM) Angriffe:** HTTPS mit TLS, HSTS für HTTPS-Erzwingung, regelmäßige Erneuerung der Zertifikate.
- **SQL-Injection:** Verwendung von JPA/Hibernate mit parametrisierten Queries, Serverseitige Validierung.
- **Distributed Denial of Service (DDoS):** Rate Limiting, geplante Nutzung von Azure DDoS Protection Services (Firewall).
- **Sicherheitslücken in Drittanbieter-Bibliotheken:** Regelmäßige Aktualisierung und Sicherheitsaudits für externe Abhängigkeiten.
- **Datenverlust:** Regelmäßige Backups und geplante Implementierung eines Disaster Recovery Plans.
- **Fehlkonfiguration der Server:** Anwendung von Best Practices, geplante Verwendung von IaC-Tools.
- **Unautorisierter Zugriff auf Administrationsschnittstellen:** Beschränkter Zugriff auf Admin-Funktionen und Implementierung von MFA für Administratoren.

7.1 Direkte und indirekte Prompt Injection

- **Bewertung:** Generative Modelle sind grundsätzlich für direkte Manipulationsversuche in Benutzereingaben und indirekte Anweisungen in Dokumenten anfällig. Sämtliche Eingaben und Dokumentinhalte werden daher als nicht vertrauenswürdig behandelt.
- **Technische Begrenzung:** Tools werden nur für den aktuellen Request bereitgestellt. Parameter werden validiert; Ausführung, Berechtigungsprüfung und Mandantentrennung erfolgen ausschließlich im Backend. Secrets und frei wählbare Mandantenparameter werden dem Modell nicht bereitgestellt.
- **Mögliche Auswirkungen:** unerwünschte oder sachlich falsche Ausgaben, Fehlinterpretationen, reversible lokale Änderungen an der geöffneten Präsentation oder der Versuch eines Tool-Aufrufs mit unerwünschten Parametern.
- **Sicherheitsgrenzen:** Eine Prompt Injection kann keine nicht bereitgestellten Tools aktivieren, Benutzerrechte erweitern, den serverseitigen Mandantenkontext verändern oder Dateien eigenständig dauerhaft speichern. Lokale Änderungen bleiben sichtbar und können rückgängig gemacht werden.
- **Restrisiko:** Prompt Injection kann nicht vollständig verhindert werden. Das verbleibende Risiko betrifft unerwünschte Modellausgaben und reversible Aktionen innerhalb des legitimen Benutzer- und Mandantenkontexts.

7.2 Halluzinationen und fachliche Kontrolle

- **Bewertung:** KI-Ausgaben können plausibel formuliert, aber sachlich falsch, unvollständig oder nicht durch die Eingabedaten belegt sein. Dies gilt insbesondere bei unklaren, widersprüchlichen oder unvollständigen Informationen.
- **Technische Begrenzung:** Tool-Parameter, Objektkennungen und Berechtigungen werden im Backend geprüft. Halluzinierte IDs, unzulässige Parameter oder behauptete Berechtigungen führen nicht automatisch zu einem gültigen Zugriff oder einer Aktion.
- **Kontrolle durch den Benutzer:** KI-Ausgaben gelten als Vorschläge. Generierte oder geänderte Inhalte sind sichtbar, können geprüft, bearbeitet, verworfen oder rückgängig gemacht werden; die dauerhafte Speicherung verbleibt beim Benutzer.
- **Restrisiko:** Inhaltlich falsche oder irreführende Ausgaben können nicht vollständig ausgeschlossen werden. Bei fachlich, rechtlich, finanziell oder geschäftlich relevanten Inhalten bleibt eine menschliche Prüfung erforderlich.

7.3 Data Poisoning und nicht vertrauenswürdige Eingabedaten

- **Abgrenzung:** slideroom trainiert oder fine-tuned keine Modelle und verwaltet keine Trainingsdaten oder Modellgewichte. Kundendokumente und ihre Embeddings werden nicht für Modelltraining verwendet und verändern keine Modellgewichte.
- **Index- und Sitzungsisolierung:** Medienpool-Embeddings bleiben an das jeweilige Dokument und den berechtigten Mandantenkontext gebunden. Chat-Embeddings dienen ausschließlich der aktuellen Sitzung und werden anschließend gelöscht. Manipulierte Inhalte können Suchergebnisse oder das Ergebnis der jeweiligen Verarbeitung beeinflussen, verändern jedoch keine Modellgewichte.
- **Risikobehandlung:** Das verbleibende Szenario wird als Risiko durch nicht vertrauenswürdige Eingaben beziehungsweise mögliche indirekte Prompt Injection und nicht als klassisches Data Poisoning behandelt.

7.4 Model Stealing und missbräuchliche Nutzung

- **Modellgewichte:** werden vollständig durch Microsoft Azure AI Foundry betrieben und sind aus der slideroom-Infrastruktur weder direkt zugänglich noch exportierbar. Endnutzer kommunizieren nicht unmittelbar mit den Modellendpunkten.
- **Schutz der Integration:** API-Schlüssel, Systemanweisungen, Tooldefinitionen, Validierungsregeln und Orchestrierungslogik werden serverseitig geschützt und nicht gezielt als Modellausgabe offengelegt.
- **Begrenzung von Massenabfragen:** Modellzugriffe sind an authentifizierte Benutzer und Unternehmen gebunden. Unternehmensbezogene AI Credits begrenzen automatisierte Abfragen, Ressourcenmissbrauch und unkontrollierten Kostenverbrauch.
- **Restrisiko:** Berechtigte Benutzer können durch wiederholte Nutzung Teile des sichtbaren Systemverhaltens, von Prompts oder Toolstrukturen ableiten. Dadurch entstehen jedoch keine Zugriffsrechte auf Modellgewichte, Azure-Secrets oder Daten anderer Mandanten.