

Technisch-organisatorische Maßnahmen (TOM)

gemäß Art. 32 DSGVO für den PII-Anonymisierungsdienst

Stand: 15.05.2026 · Version 1.0

Übersicht

Dieses Dokument beschreibt die technisch-organisatorischen Maßnahmen (TOM) gemäß Art. 32 DSGVO, die der PII-Anonymisierungsdienst von Michael SCHILLER - Organisation. Digital. implementiert hat, um ein dem Risiko angemessenes Schutzniveau zu gewährleisten.

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

1.1 Zutrittskontrolle

Schutz vor unbefugtem Zutritt zu Datenverarbeitungsanlagen

- › Server in einem professionellen Rechenzentrum in Deutschland (Hetzner, Nürnberg)
- › Physischer Zugangsschutz durch den Rechenzentrumsbetreiber (ISO 27001 zertifiziert)
- › Kein physischer Zugang durch Mitarbeiter des Auftragsverarbeiters erforderlich (vollständig remote verwaltet)

1.2 Zugangskontrolle

Schutz vor unbefugter Nutzung der Datenverarbeitungssysteme

- › SSH-Zugang ausschließlich über schlüsselbasierte Authentifizierung (ED25519)
- › Kein Root-Zugang, nur dedizierte Benutzerkonten
- › Anubis Bot-Protection vor allen öffentlichen Endpunkten
- › Rate Limiting: 60 Anfragen/Minute pro IP-Adresse
- › Fail2Ban für SSH-Brute-Force-Schutz

1.3 Zugriffskontrolle

Schutz vor unbefugtem Lesen, Kopieren, Verändern oder Entfernen

- › API-Endpunkte sind öffentlich zugänglich (keine Authentifizierung erforderlich) – dies ist beabsichtigt, da keine personenbezogenen Daten dauerhaft gespeichert werden
- › Zuordnungstabellen sind session-basiert isoliert (UUID) und nur mit der korrekten Session-ID abrufbar
- › Keine Administrations-Oberfläche für PII-Daten
- › Keine Möglichkeit, verarbeitete Texte nachträglich einzusehen

1.4 Trennungskontrolle

Getrennte Verarbeitung für unterschiedliche Zwecke

- › Jede Anonymisierungs-Session erhält eine eigene UUID
- › Session-Daten sind vollständig voneinander isoliert
- › Keine gemeinsame Nutzung von Zuordnungstabellen zwischen Sessions
- › Produktions- und Staging-Umgebung physisch getrennt (separate Container, separate Datenbanken)

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

2.1 Weitergabekontrolle

Schutz bei Datenübertragung

- › TLS 1.3 Verschlüsselung für alle API-Kommunikation (HTTPS)
- › Caddy Reverse Proxy mit automatischer Zertifikatserneuerung (Let's Encrypt)
- › Keine unverschlüsselten Übertragungswege
- › Keine Datenübermittlung an Drittländer
- › Keine Unterauftragsverarbeiter

2.2 Eingabekontrolle

Nachvollziehbarkeit der Datenverarbeitung

- › Keine Protokollierung von übermittelten Texten oder PII-Inhalten (Privacy by Design)
- › Nur technische Metadaten werden geloggt: Zeitstempel, IP-Adresse, Anzahl erkannter Entities, Verarbeitungsdauer
- › Zuordnungstabellen werden automatisch nach maximal 1 Stunde gelöscht

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

3.1 Verfügbarkeitskontrolle

Schutz gegen Zerstörung oder Verlust

- › Automatische Container-Neustarts bei Ausfällen (Kamal/Docker)
- › Monitoring mit Grafana-Dashboard
- › Tägliche Backups der Systemkonfiguration
- › Keine dauerhafte Speicherung personenbezogener Daten → kein Datenverlust möglich
- › Bei Systemausfall gehen maximal die temporären Zuordnungstabellen verloren (max. 1h Daten)

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

Da keine personenbezogenen Daten dauerhaft gespeichert werden, beschränkt sich die Wiederherstellung auf die Systemverfügbarkeit:

- › Container-basierte Deployment-Architektur (Kamal) ermöglicht schnelle Wiederherstellung
- › Automatische Health-Checks und Neustart-Mechanismen
- › Disaster Recovery durch Image-basierte Deployments: Vollständige Neubereitstellung in unter 10 Minuten

5. Verfahren zur regelmäßigen Überprüfung (Art. 32 Abs. 1 lit. d DSGVO)

5.1 Datenschutz-Management

- › Regelmäßige Überprüfung der technischen Maßnahmen
- › Automatisierte Sicherheitstests (Brakeman, Bundler-Audit) in der CI/CD-Pipeline
- › Dependency-Scanning via Dependabot
- › Code-Reviews vor jedem Deployment

5.2 Privacy by Design & Default

- › Originaltexte werden ausschließlich im RAM verarbeitet – keine Speicherung auf Festplatte
- › Zuordnungstabellen mit automatischem TTL (max. 1 Stunde)
- › Keine Inhalts-Logs – es werden keine übermittelten Texte protokolliert
- › Session-Isolation durch UUIDs – kein Session-übergreifender Zugriff möglich
- › 5 parallele Erkennungsmethoden für maximale Abdeckung bei minimaler Datenhaltung

6. Erkennungssysteme

Der PII-Anonymisierungsdienst setzt 5 parallele Erkennungsmethoden ein:

Pattern Detection – Regex-basierte Erkennung strukturierter Daten (E-Mail, Telefon, IBAN, Kreditkarten, EU-VAT)

SpaCy NER – KI-basierte Named Entity Recognition (lokal, kein Cloud-Service)

Address Detection – Spezialisierte Adresserkennung für 30+ Länder

Crypto-Wallet-Erkennung – 12 Kryptowährungen (Bitcoin, Ethereum, Litecoin, Ripple u.a.) per Pattern-Matching

Abbreviation Detection - Erkennung von Unternehmens-/Bereichsabkürzungen (standardmäßig aktiv, ~665 Whitelist-Einträge)

Alle Erkennungssysteme laufen lokal auf eigenen Servern in Deutschland. Es werden keine Daten an externe Dienste übermittelt.

7. Pseudonymisierungs-Modi

Nach der Erkennung werden personenbezogene Werte je nach Zielsystem in einem von drei Modi pseudonymisiert:

Tagged-Modus - HMAC-basierte Marker im Format [TYPE:8hex], deterministisch pro Session. Beispiel: [PERSON:a1b2c3d4]. Eingesetzt an externen Schnittstellen (MCP, öffentliche API, Logs). Das empfangende System erkennt klar, dass das Datum anonymisiert ist.

Faker-Modus - Auswahl aus deutschen Faker-Pools (~30.000 Einträge, Namen, Adressen, Firmen). HMAC-indiziert mit deterministischer linearer Probing, sodass derselbe Eingabewert in einer Session immer dasselbe Pseudonym erhält. Eingesetzt an internen KI-Pipelines (AI-Chat, Voice, Mailer). Sprachmodelle verarbeiten plausible Pseudonyme besser als Bracket-Marker.

Sentinel-Modus - Fester Sentinel-String <unterdrückt> für hochsensible Werte wie IBAN und Telefonnummer. Plausibel aussehende Fake-Werte würden das Sprachmodell verleiten, die Daten als echt zu behandeln; deshalb wird hier ein eindeutig erkennbarer Platzhalter eingesetzt.

In allen drei Modi verlassen die Original-Werte unsere Verarbeitungsgrenze nie als Klartext, weder zu Cloud-KI-Diensten noch zu externen Konsumenten. Vor der Auslieferung an Endnutzer werden die Pseudonyme serverseitig wieder durch die echten Werte ersetzt - Endnutzer sehen also immer ihre Original-Daten.

8. Infrastruktur-Übersicht

Server

- Hetzner Cloud, Nürnberg, Deutschland
- Ubuntu Linux (gehärtet)
- Kamal (Docker-basiert)

Dienste

- Rails 8 (Ruby 3.4) - API und Web-Oberfläche
- SpaCy NER Service - Lokaler KI-Container
- Crypto & EU-ID Pattern-Matching - Integriert in Pattern-Detector

- Caddy - Reverse Proxy mit TLS 1.3
- Redis - Session-Cache mit TTL

9. Ansprechpartner

Bei Fragen zu den technisch-organisatorischen Maßnahmen:

Michael Schiller (Geschäftsführer)

Michael SCHILLER - Organisation. Digital.

info@schiller-partners.de