Technical and Organizational Measures (TOM)

pursuant to Art. 32 GDPR for the PII Anonymization Service

Last updated: 25.05.2026 · Version 1.0

Overview

This document describes the technical and organizational measures (TOM) pursuant to Art. 32 GDPR that the PII anonymization service by Michael SCHILLER - Organisation. Digital. has implemented to ensure a level of protection appropriate to the risk.

1. Confidentiality (Art. 32(1)(b) GDPR)

1.1 Physical Access Control

Protection against unauthorized access to data processing facilities

- Servers in a professional data center in Germany (dataforest GmbH, Frankfurt am Main)
- Physical access protection by the data center operator (ISO 27001 certified)
- No physical access by Processor employees required (fully remotely managed)

1.2 System Access Control

Protection against unauthorized use of data processing systems

- SSH access exclusively via key-based authentication (ED25519)
- No root access, only dedicated user accounts
- Anubis bot protection for all public endpoints
- Rate limiting: 60 requests/minute per IP address
- Fail2Ban for SSH brute force protection

1.3 Data Access Control

Protection against unauthorized reading, copying, modification, or removal

- API endpoints are publicly accessible (no authentication required) – this is intentional as no personal data is permanently stored
- Mapping tables are session-isolated (UUID) and only accessible with the correct session ID
- No administration interface for PII data
- No ability to retrospectively view processed texts

1.4 Separation Control

Separate processing for different purposes

- Each anonymization session receives its own UUID
- Session data is completely isolated from each other
- No shared use of mapping tables between sessions
- Production and staging environments physically separated (separate containers, separate databases)

2. Integrity (Art. 32(1)(b) GDPR)

2.1 Transfer Control

Protection during data transmission

- TLS 1.3 encryption for all API communication (HTTPS)
- Caddy reverse proxy with automatic certificate renewal (Let's Encrypt)
- No unencrypted transmission channels
- No data transfer to third countries
- No sub-processors

2.2 Input Control

Traceability of data processing

- No logging of transmitted texts or PII content (Privacy by Design)
- Only technical metadata is logged: timestamp, IP address, number of detected entities, processing duration
- Mapping tables are automatically deleted after maximum 1 hour

3. Availability and Resilience (Art. 32(1)(b) GDPR)

3.1 Availability Control

Protection against destruction or loss

- Automatic container restarts on failures (Kamal/Docker)
- Monitoring with Grafana dashboard
- Daily backups of system configuration
- No permanent storage of personal data → no data loss possible
- In case of system failure, only temporary mapping tables are lost (max. 1h of data)

4. Recoverability (Art. 32(1)(c) GDPR)

Since no personal data is permanently stored, recovery is limited to system availability:

- Container-based deployment architecture (Kamal) enables rapid recovery
- Automatic health checks and restart mechanisms
- Disaster recovery through image-based deployments: complete re-provisioning in under 10 minutes

5. Regular Review Procedures (Art. 32(1)(d) GDPR)

5.1 Data Protection Management

- Regular review of technical measures
- Automated security tests (Brakeman, Bundler-Audit) in the CI/CD pipeline
- Dependency scanning via Dependabot
- Code reviews before every deployment

5.2 Privacy by Design & Default

- Original texts are processed exclusively in RAM – no storage on disk
- Mapping tables with automatic TTL (max. 1 hour)
- No content logs – no transmitted texts are logged
- Session isolation via UUIDs – no cross-session access possible
- 5 parallel detection methods for maximum coverage with minimal data retention

6. Detection Systems

The PII anonymization service employs 5 parallel detection methods:

Pattern Detection – Regex-based detection of structured data (email, phone, IBAN, credit cards, EU-VAT)

GLiNER NER – PII-specialized Named Entity Recognition (local, no cloud service, ONNX-optimized)

Address Detection – Specialized address detection for 30+ countries

Crypto Wallet Detection – 12 cryptocurrencies (Bitcoin, Ethereum, Litecoin, Ripple etc.) via pattern matching

Abbreviation Detection – Detection of company/department abbreviations (enabled by default, ~665 whitelist entries)

All detection systems run locally on own servers in Germany. No data is transmitted to external services.

7. Pseudonymization Modes

After detection, personal values are pseudonymized in one of four modes depending on the data type and target system:

Tagged mode – HMAC-based markers in the format [TYPE:8hex], deterministic per session. Example: [PERSON:a1b2c3d4]. Used on external surfaces (MCP, public API, logs). The receiving system clearly recognizes that the value is anonymized.

Faker mode – Selection from German faker pools (~30,000 entries; names, addresses, companies). HMAC-indexed with deterministic linear probing, so the same input value within a session always maps to the same pseudonym. Used on internal AI pipelines (AI chat, voice, mailer). Language models handle plausible pseudonyms better than bracket markers.

Strict-Pool mode – For highly sensitive financial and identifier values (IBAN, credit-card numbers, BIC, German tax/social/health-insurance IDs, Swiss AHV, Dutch BSN, French INSEE, Belgian Rijksregister, Spanish DNI/NIE, German ID card, Austrian social-security number) we generate format-preserving pseudonyms from reserved test ranges (ISO-3166-non-allocated IBAN prefixes, Stripe-test PAN 4242, leading-zero IDs, T-prefix markers). Each pseudonym passes the format check but deliberately FAILS its own algorithm validator (re-detection guard) — structurally impossible to hit a real identifier. The language model sees a plausible shape but cannot trigger a real bank or government action.

In all four modes, original values never leave our processing boundary as plaintext, neither to cloud AI services nor to external consumers. Before delivery to end users, pseudonyms are server-side replaced with the real values – end users always see their original data.

8. Infrastructure Overview

Server

- dataforest GmbH, Frankfurt am Main, Germany
- Ubuntu Linux (hardened)
- Kamal (Docker-based)

Services

- Rails 8 (Ruby 4.0) – API and web interface
- GLiNER NER Service – Local AI container (PII-specialized)
- Crypto & EU-ID Pattern Matching – Integrated in Pattern Detector
- Caddy – Reverse proxy with TLS 1.3
- Redis – Session cache with TTL

9. Contact

For questions about the technical and organizational measures:

Michael Schiller (Managing Director)

Michael SCHILLER - Organisation. Digital.

info@schiller-partners.de