

PII Firewall

Pseudonymisierung von Arbeitsdokumenten, bevor sie ein externes Sprachmodell erreichen.

Zeitraum: 08/2026 – heute

Rolle: Produktverantwortlicher für Architektur, Erkennung, Verschlüsselung, Oberfläche und Sicherheitsprüfungen in der CI
Quellcode nicht öffentlich

AUSGANGSLAGE

Verbote externer Sprachmodelle halten meist nur bis zur ersten dringenden Aufgabe. Dann landet ein Vertrag mit personenbezogenen Daten im Browser eines Drittanbieters. Fertige PII-Detektoren sind auf westliche Formate ausgelegt und unterscheiden eine russische Steuernummer nicht von einer gleich langen Vertragsnummer.

LÖSUNG

- Eingang sind PDF-, DOCX- und XLSX-Dateien. Ausgang sind pseudonymisierter Text und eine verschlüsselte Ersetzungstabelle, die Antwort des Modells kommt mit den Originalwerten zurück.
- Vier Erkennungsschichten: Strukturregeln mit Prüfsummen für Kennungen, ein GLiNER-NER-Modell auf der GPU, ein Wörterbuch mit morphologischem Schlüssel und eine Schicht für Zugangsdaten. Gefundene Zugangsdaten blockieren das ganze Dokument.
- Die Pseudonymisierungszone hat weder eine Route ins Internet noch zum Modell. Ein Fehler in ihrem Code kann deshalb keine ausgehende Anfrage erzeugen.
- Die Ersetzungstabelle wird nur im Perimeter und nur für die Person entschlüsselt, die das Dokument hochgeladen hat.
- Dokumente mit Krankenversicherungsnummer gelten als Gesundheitsdaten und gehen standardmäßig nur an das lokale Modell.

ERGEBNISSE

- Synthetischer Referenzsatz, 40 Dokumente mit 322 Vorkommen: Recall 100 %, Precision 100 %.
- Wiederherstellung in 20 von 20 Fällen bytewise identisch mit dem Original.
- Nächste Prüfphase: echte Dokumente und Korpora öffentlicher Gerichtsentscheidungen.

TECHNOLOGIEN

Python · GLiNER · OCR · Envelope Encryption · Keycloak SSO · Docker · GitLab CI