

PII Firewall

De-identification of work documents before they reach an external language model.

Period: 08/2026 – present

Role: product owner for architecture, detection, encryption, interface and security checks in CI

Source code not public

BACKGROUND

Bans on external language models usually last until the first urgent task. Then a contract with personal data ends up with a third-party provider. Off-the-shelf PII detectors are built for Western formats and cannot tell a Russian tax number from a contract number of the same length.

SOLUTION

- Input is PDF, DOCX and XLSX files. Output is de-identified text and an encrypted substitution map, and the answer of the model comes back with the original values.
- Four detection layers: structural rules with checksums for identifiers, a GLiNER NER model on the GPU, a dictionary with a morphological key and a layer for credentials. Detected credentials block the entire document.
- The de-identification zone has no route to the internet or to the model. A bug in its code therefore cannot produce an outgoing request.
- The substitution map is decrypted only inside the perimeter and only for the person who uploaded the document.
- Documents with a health insurance number count as health data and go only to the local model by default.

RESULTS

- Synthetic gold set, 40 documents with 322 mentions: recall 100%, precision 100%.
- Restoration byte-identical to the original in 20 of 20 cases.
- Next test phase: real documents and corpora of published court decisions.

TECHNOLOGIES

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