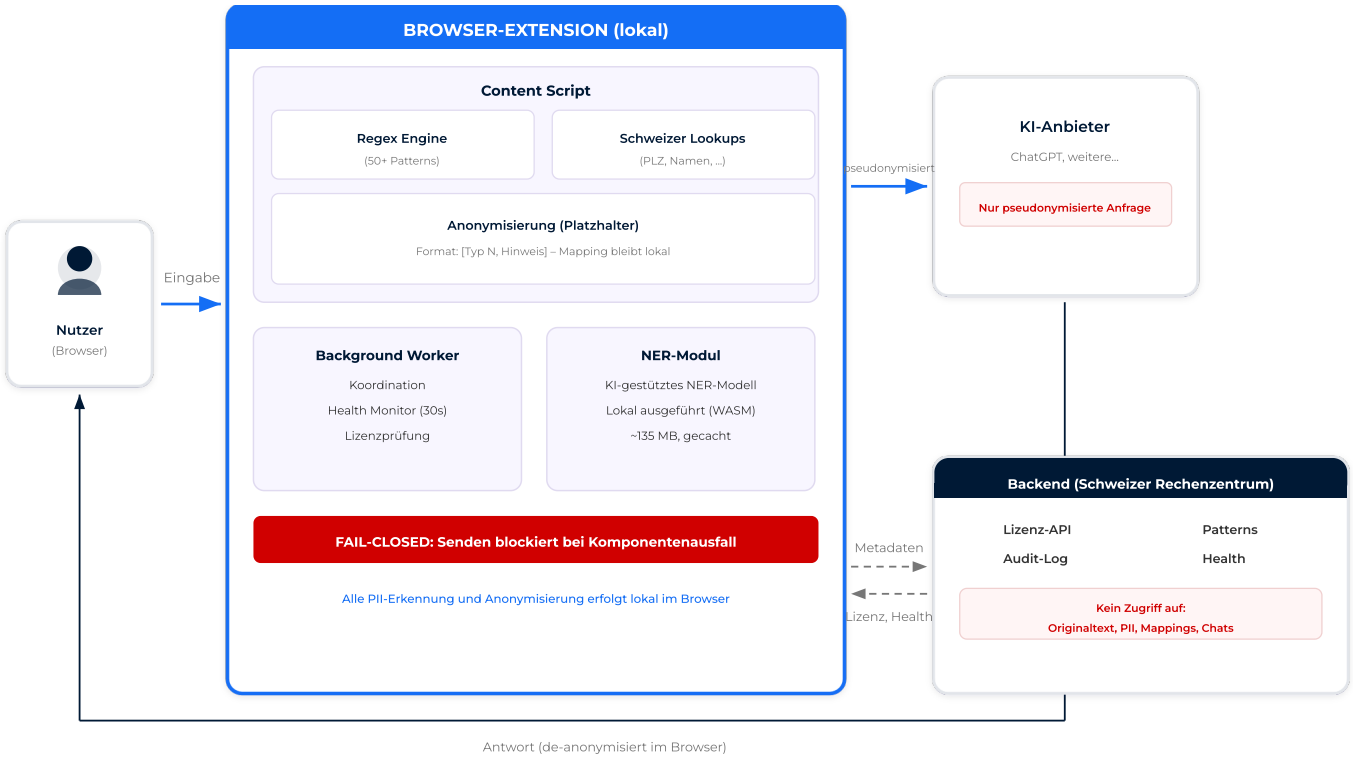
Promptspace Guard

Sicherer KI-Zugriff für Kanzleien, Treuhand und Beratungen

Technischer Überblick zu Architektur, Datenfluss und Compliance

Promptspace ermöglicht die kontrollierte Nutzung von KI-Tools, indem sensible Inhalte vor der Übermittlung erkannt, je nach Konfiguration pseudonymisiert und nachvollziehbar protokolliert werden.

1 ARCHITEKTURÜBERSICHT



2 DATENFLUSS IM DETAIL

1. Mitarbeitende geben Inhalte in unterstützte KI-Oberflächen ein (aktuell: ChatGPT). Weitere KI-Tools werden auf Kundennachfrage kontinuierlich berücksichtigt.
2. Die Extension prüft Eingaben lokal mittels 50+ Regex-Patterns und KI-basierter Namenserkennung (KI-gestütztes NER-Modell, bereitgestellt über Promptspace-Infrastruktur, lokal im Browser ausgeführt via WASM).
3. Erkannte sensible Daten werden als Platzhalter pseudonymisiert (Format: [Typ N, Hinweis], z.B. [Person 1, männlich]).
4. Nur die pseudonymisierte Anfrage wird an den KI-Anbieter übermittelt. Das Mapping (Platzhalter → Originaldaten) bleibt lokal im Browser.
5. Die KI-Antwort wird im Browser automatisch de-anonymisiert (Platzhalter → Originaldaten) und dem Nutzer angezeigt.
6. Audit-Metadaten (Aktion, Risiko-Level, PII-Anzahl) werden protokolliert. Keine Originalinhalte werden übermittelt.
7. Erkennungsregeln und Policies werden zentral über das Backend verteilt und lokal gecacht.

3 DATENKATEGORIEN UND SPEICHERORTE

DATENKATEGORIE	SPEICHERORT	DAUER	ZUGRIFF PROMPTSPACE
Eingabetext (Original)	Browser-RAM	Sitzung	NEIN
Erkannte PII-Werte	Browser-RAM	Sitzung	NEIN
Anonymisierungs-Mapping	sessionStorage	Bis Tab geschlossen	NEIN
Pseudonymisierter Text	KI-Anbieter	Gemäss KI-Anbieter	NEIN
Audit-Metadaten	Backend (CH)	Gemäss Policy	ADMIN
Lizenzdaten	Backend (CH)	Vertragsdauer	ADMIN
Health Reports	Backend (CH)	Anonymisiert	SUPPORT

4 ZUGRIFFSKONTROLLE – PROMPTSPACE

Zugriff

- ✓ Technische Betriebsdaten
- ✓ Anonymisierte Audit-Metadaten
- ✓ Lizenzkonfiguration
- ✓ Pattern-Definitionen

Kein Zugriff

- ✗ Nutzereingaben (Originaltext)
- ✗ Erkannte PII-Werte
- ✗ Anonymisierungs-Zuordnungen (Mappings)
- ✗ KI-Antworten und Chat-Verläufe
- ✗ Browserdaten und Cookies

5 SICHERHEITSARCHITEKTUR

EIGENSCHAFT	DETAILS
Datenübertragung	TLS 1.2+ (Backend), HTTPS (KI-Anbieter)
Erkennung	50+ Regex-Patterns + KI-gestütztes NER-Modell (lokal, WASM)
Fail-Closed	Senden blockiert bei Komponentenausfall
Health Monitoring	30-Sekunden-Intervall, automatische Recovery
Audit-Logging	Nur Metadaten – keine PII-Werte, kein Originaltext
Session-Isolation	Mapping pro Browser-Tab, Löschung bei Tab-Schliessung

6 HOSTING UND UNTERAUFTRAGSBEARBEITER

KOMPONENTE	ANBIETER	STANDORT	ZWECK
Extension	Lokal (Browser)	–	PII-Erkennung, Anonymisierung
Backend	Schweizer Infrastruktur	Schweiz	Lizenz, Audit, Patterns
KI-Zugriff	Direkt vom Browser	Gemäss Anbieter	KI-Verarbeitung
NER-Modell	Promptspace (Download)	Lokal gecacht	Namenserkennung

> Promptspace betreibt keine eigenen KI-Modelle. Der Datenverkehr zwischen Nutzer und KI-Anbieter verläuft direkt – Promptspace fungiert als lokale Schutzschicht.

7 COMPLIANCE-MATRIX

REGULIERUNG	RELEVANZ	UMSETZUNG
nDSG	Datenminimierung	Lokale Verarbeitung, keine PII-Übermittlung an Dritte
DSGVO / GDPR	Privacy by Design	Keine personenbezogenen Daten in Audit-Logs
EU AI Act	Transparenz	Audit-Trail, nachvollziehbare Anonymisierung
Berufsgeheimnis (Art. 321 StGB)	Geheimhaltungspflicht	Pseudonymisierung vor jeder Übermittlung

Promptspace arbeitet aktiv an branchenüblichen Sicherheitszertifizierungen.

8 ADMIN & GOVERNANCE

- Dashboard mit Statistiken und Audit-Log
- Health Reports und Status-Übersicht
- CSV-Export für Compliance-Nachweise
- Zentrale Pattern-Verwaltung
- Individuelle Admin-Blacklist für eigene Sperrbegriffe
- Sitz-Management (seat-basierte Lizenzierung)
- Schrittweises Rollout (Pilotteam → Organisation)
- Admin/User-Rollentrennung
- «Trotzdem senden» optional deaktivierbar durch Admin