White Paper zur sicheren KI-Nutzung

Promptspace.ch bietet ein KI-Governance-Paket für Unternehmen mit sensiblen Daten, die Standard-LLMs produktiv nutzen und Risiken von Datenabflüssen wirksam reduzieren möchten.

PromptSpaceGuard bildet dabei die Browser-basierte Schutzschicht für die kontrollierte Nutzung externer KI-Systeme. Sensible Inhalte werden lokal erkannt und je nach Regelwerk blockiert oder pseudonymisiert, bevor sie das Unternehmen verlassen. So lassen sich Datenschutz, Nachvollziehbarkeit und produktive KI-Nutzung besser miteinander verbinden.

Ausgangslage

Mit der zunehmenden Nutzung generativer KI im Arbeitsalltag entstehen für Unternehmen neben Effizienzgewinnen auch Risiken für Datenschutz und Datensicherheit. Ohne klare Rahmenbedingungen bleiben zudem wesentliche Produktivitätspotenziale ungenutzt.

Aktuelle Daten (KPMG, Netskope, Brynjolfsson, Forbes) aus 2025/2026 belegen diese Thematik:

- **48 %** der Mitarbeitenden nutzen KI entgegen den Richtlinien im Verborgenen.
- **22 %** der Uploads enthalten dabei hochsensible Daten wie Namen oder Beträge.
- **34 %** an Produktivität gehen durch Einschränkungen der Nutzung verloren.

Lösung

Sensible Daten werden automatisch erkannt und lokal bei den Unternehmen pseudonymisiert, bevor sie deren Hoheitsbereich verlassen. Im Hintergrund wird automatisch die notwendige Governance (Audit-Trails, Compliance-Nachweise) zur Rechtskonformität sichergestellt.

Nutzen

Unser Ansatz schützt sensible Eingaben bereits vor der Übermittlung und ist modellagnostisch einsetzbar. Wir sind davon überzeugt, damit eine besonders wirksame Lösung für die sichere Nutzung gängiger LLMs zu bieten, ohne dass ein ressourcenintensives IT-Projekt lanciert werden muss:

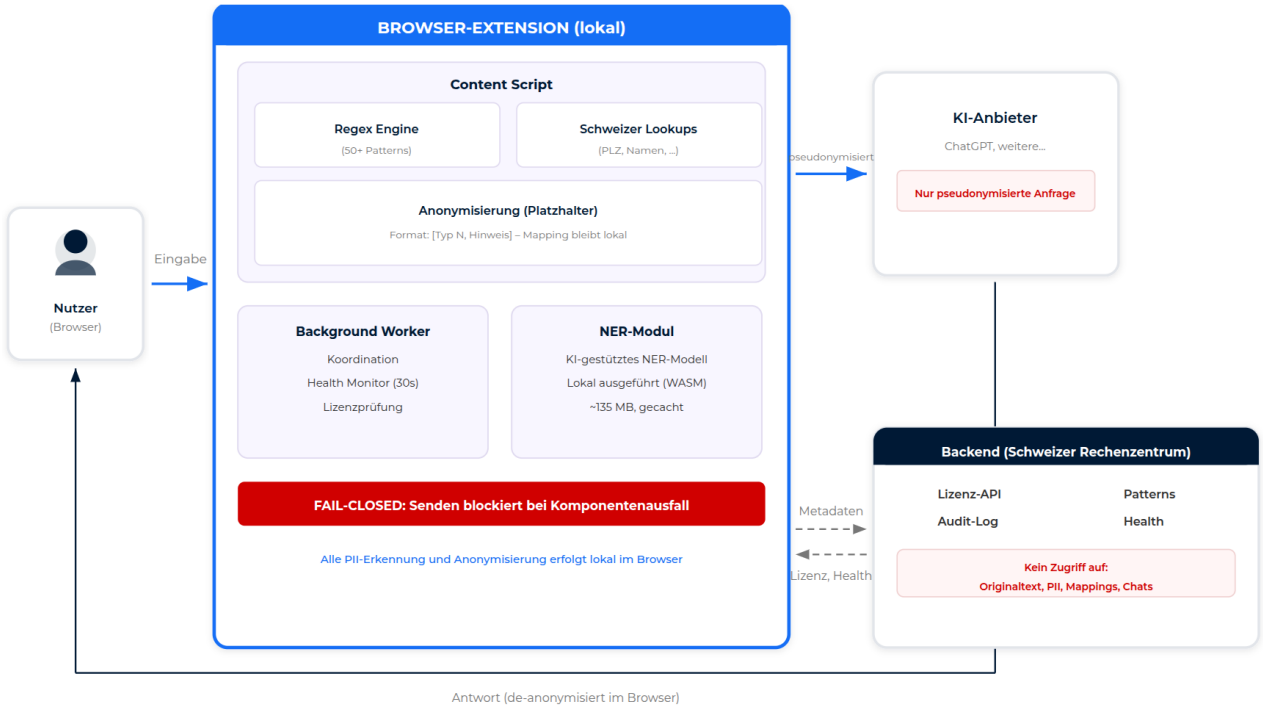
- **Sicherheit:** Sie schaffen die Voraussetzungen für einen sicheren Umgang mit KI. Damit minimieren Sie die ineffiziente Schatten-KI und verhindern den Abfluss sensibler Informationen.
- **Produktivität:** Sie heben das volle Produktivitätspotenzial Ihrer Mitarbeitenden, statt dieses einzuschränken und erhalten ein vollumfängliches KI-Governance Paket.
- **Technologie:** Ihre Mitarbeitenden können in ihren gewohnten LLM-Umgebungen arbeiten und ohne Verzögerung vom technologischen Fortschritt der jeweiligen LLMs profitieren.

Gerne zeigen wir Ihnen in einem unverbindlichen und vertraulichen **Erstgespräch**, wie sich PromptSpaceGuard in Ihre bestehende KI-Nutzung integrieren lässt.

Anhang: Funktionsweise & Sicherheit

Systemarchitektur von PromptSpaceGuard

PromptSpaceGuard ist kein eigenes KI-Modell, sondern eine vorgeschaltete Schutzschicht für die kontrollierte Nutzung externer KI-Systeme. Sensible Inhalte werden lokal im Browser erkannt und je nach Regelwerk markiert, blockiert oder pseudonymisiert. An den KI-Anbieter wird nur die freigegebene Anfrage weitergegeben; im Backend verarbeitet Promptspace ausschliesslich die für Betrieb, Lizenzierung, Konfiguration und Audit erforderlichen Metadaten.



Datenfluss

- Mitarbeitende geben Inhalte in unterstützte KI-Oberflächen (z.B. ChatGPT) ein.
- Die Extension prüft Eingaben lokal auf definierte sensible Inhalte.
- Je nach Regelwerk werden Inhalte markiert, blockiert oder pseudonymisiert.
- Nur die freigegebene, verarbeitete Anfrage wird an das KI-Modell weitergegeben.
- Die Antwort wird an den Nutzer zurückgegeben.
- Metadaten werden für Administratoren protokolliert.
- Regeln und individuelle Blacklists können zentral über das Dashboard verwaltet werden.

Was Promptspace.ch sieht

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

Grenzen und Annahmen

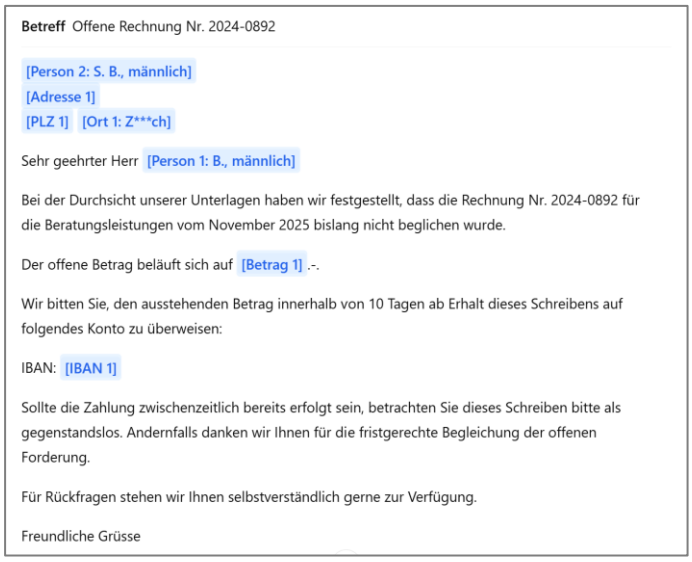
PromptSpaceGuard reduziert Risiken bei der Nutzung externer KI-Systeme erheblich. Seine Wirkung ist am stärksten, wenn technische Schutzmechanismen, organisatorische Vorgaben und Schulung sinnvoll zusammenspielen. Die Wirksamkeit hängt insbesondere davon ab, dass Regeln, Muster und Konfigurationen zum jeweiligen Anwendungsfall passen und die unterstützten Oberflächen korrekt eingebunden sind. Externe KI-Anbieter bleiben eigenständige Dienste mit eigenen Rahmenbedingungen.

Funktion der Pseudonymisierung

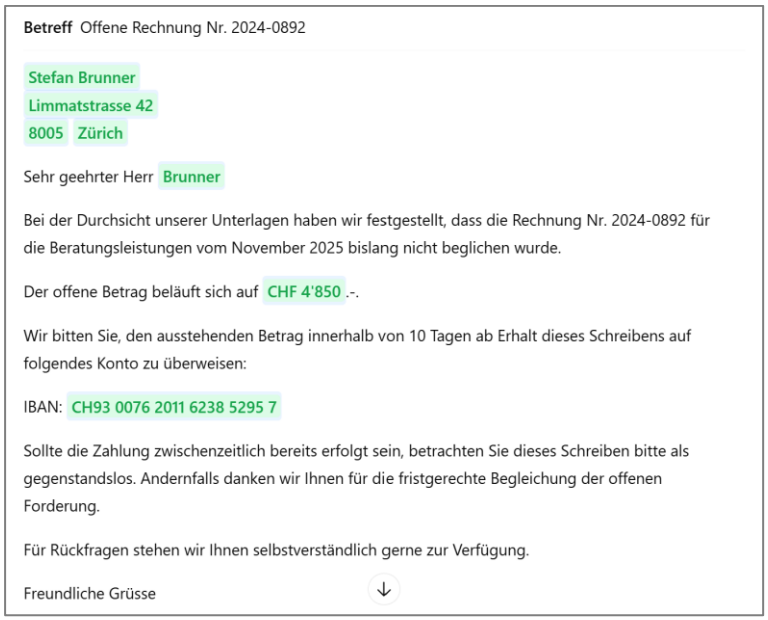
Unsere Lösung PromptSpaceGuard arbeitet als lokale Browser-Erweiterung, wie folgt:

1.1 Erkennung: Sensible Daten (Namen, Orte, E-Mails, Telefonnummern, IBAN, AHV-Nr.) werden direkt im Browser des Mitarbeitenden erkannt.

1.2 Ersetzung und Übermittlung: Sensible Daten werden durch neutrale Platzhalter ersetzt (z. B. [Person A], [Firma B]) und einzig der anonymisierte Text verlässt das Unternehmen.



1.3. Rückübersetzung: Die Antwort der KI wird lokal im Browser wieder mit den Originaldaten angereichert.



Wichtig: Zero-Knowledge, Promptspace.ch oder Dritte haben zu keinem Zeitpunkt Zugriff auf die Dateneingabe. Diese liegen jeweils temporär im Arbeitsspeicher des Nutzers.

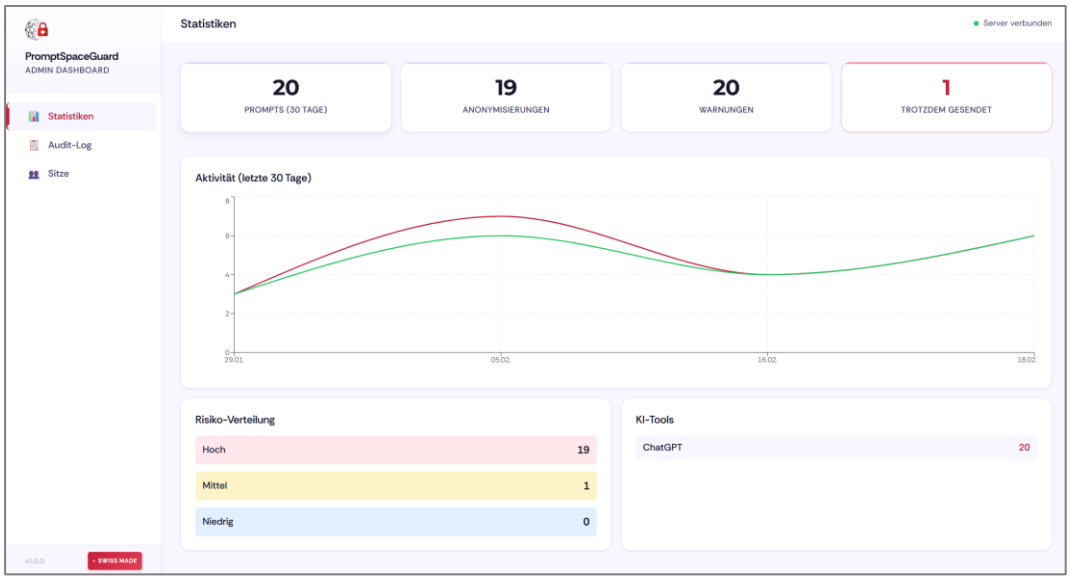
Datenschutz & Compliance (Schweiz & EU)

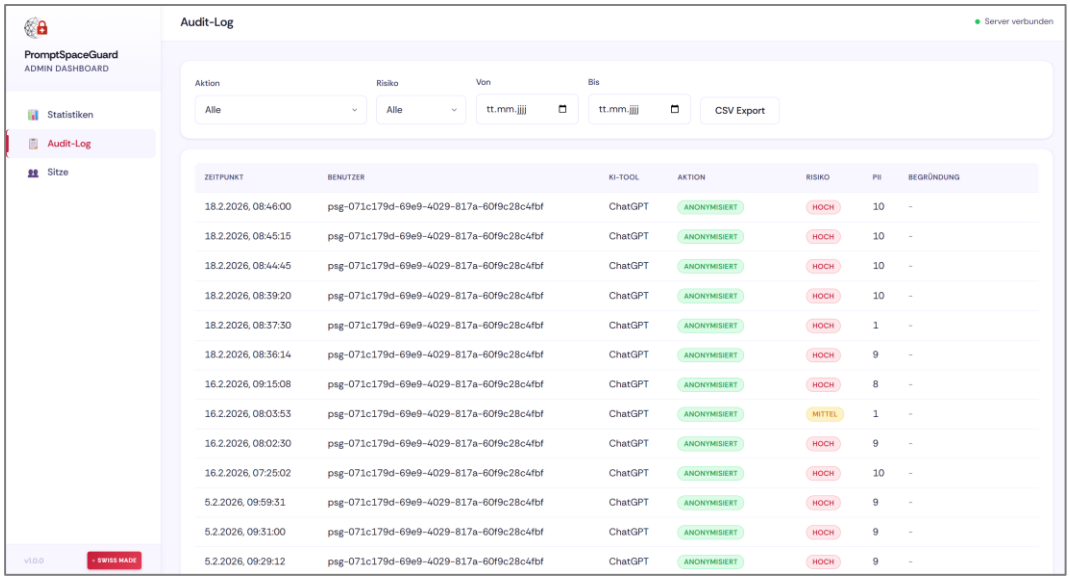
- **Server-Standort:** Administrative Betriebsdaten wie Lizenzen und Audit-Metadaten werden in der Schweiz verarbeitet.
- **Drittstaatenrisiken:** Da sensible Inhalte vor der Übermittlung lokal erkannt und je nach Konfiguration pseudonymisiert werden, wird das Risiko des Zugriffs durch externe Anbieter oder Behörden auf Klartextdaten deutlich reduziert.
- **Datenschutzprinzipien:** Das Verfahren ist auf Datenminimierung, Privacy-by-Design und eine kontrollierte Nutzung externer KI-Systeme ausgelegt.
- **Vertraulichkeit und Berufsgeheimnis:** Da Promptspace keinen Zugriff auf Originaleingaben, erkannte PII-Werte oder lokale Zuordnungen benötigt, kann die Wahrung von Vertraulichkeitspflichten und Berufsgeheimnissen technisch unterstützt werden.

B) Audit-Fähigkeit

Für Compliance-Zwecke wird ein Audit-Log erstellt, das speichert:

- **Wer** hat wann LLMs genutzt?
- **Wie viele** sensible Daten wurden pseudonymisiert?
- **Welche** Art von Daten (z.B. “3x Personenbezug, 1x IBAN”)





Wichtig: Der Inhalt des Prompts wird nicht gespeichert, um die Vertraulichkeit zu wahren.

Weiterführende Informationen zu Datenschutz, Auftragsverarbeitung und Sicherheitsmassnahmen finden Sie in unserem **Trust-Center.** Für eine individuelle Einordnung laden wir Sie gerne zu einem unverbindlichen und vertraulichen **Erstgespräch** ein.