



White Paper zur sicheren KI-Nutzung

Promptspace.ch bietet das KI-Governance-Paket für Unternehmen mit sensiblen Daten, welche Standard-LLMs produktiv nutzen und Datenabflüsse vermeiden möchten.

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.



Anhang: Funktionsweise & Sicherheit

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.

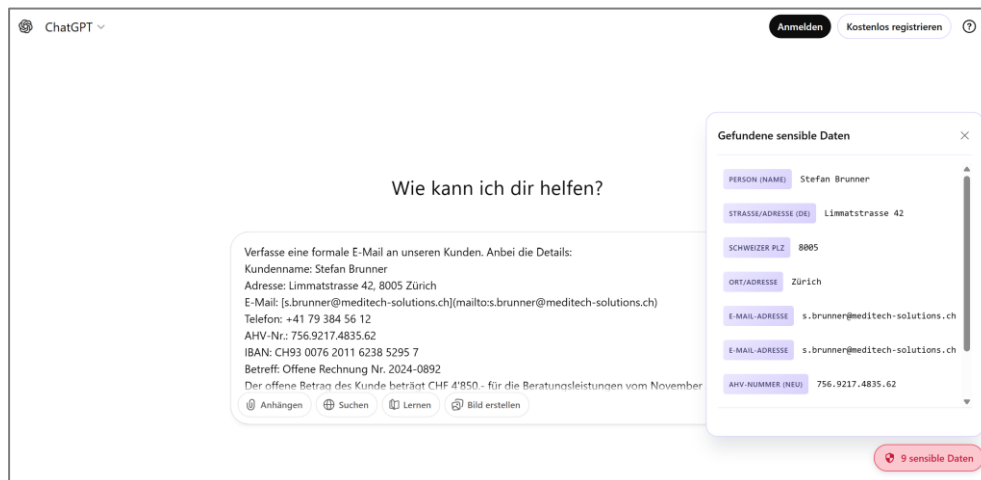


Abb. 1.1: Erkennung sensibler Daten

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.

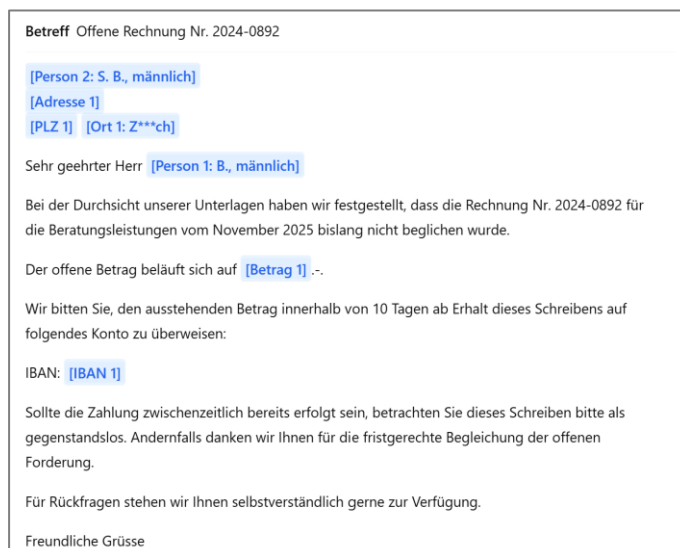


Abb. 1.2: Ersetzung sensibler Daten



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

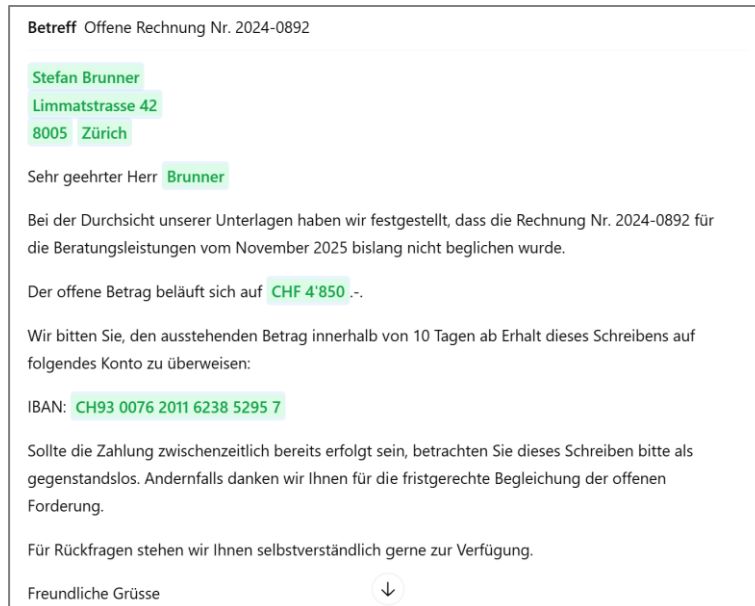


Abb. 1.3: Rückübersetzung sensibler Daten

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:** Alle administrativen Daten (Lizenzen, Logs) liegen in der Schweiz (Bern/Zürich).
- **US-Cloud-Act-Risiko:** Da keine sensiblen Daten im Klartext an US-Server (OpenAI, Microsoft) übermittelt werden, wird das Risiko des Zugriffs durch Anbieter und Behörden auf Mandantendaten deutlich reduziert.
- **DSGVO- und nDSG-Konformität:** Das Verfahren entspricht den Grundsätzen der Datensparsamkeit und Privacy-by-Design.
- **Berufsgeheimnis:** Durch die technische Unmöglichkeit der Einsichtnahme durch Dritte (auch Promptspace.ch) wird die Wahrung des Berufsgeheimnisses unterstützt.



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”)

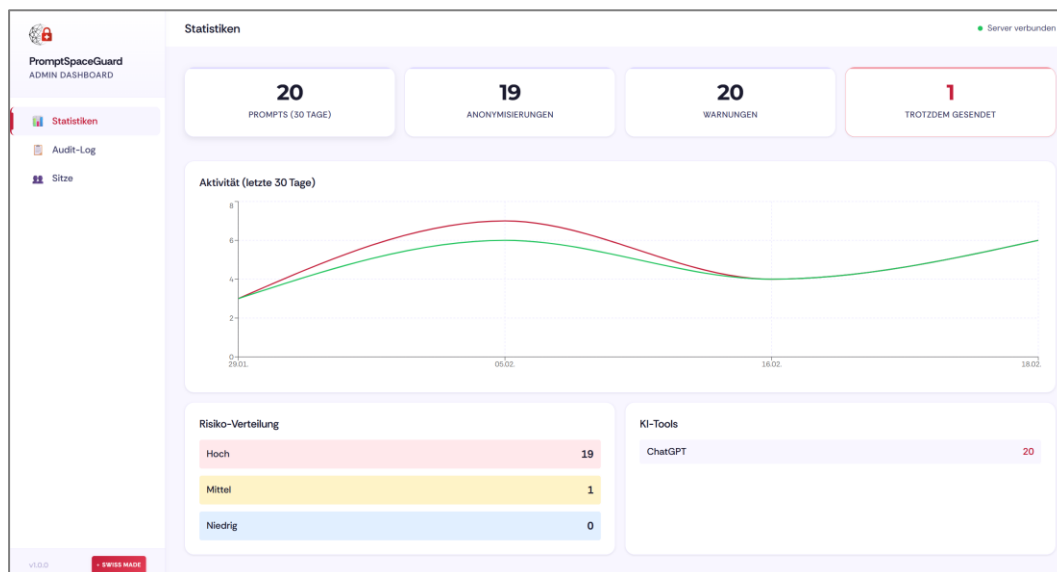


Abb. 3.1: Dashboard für Administrator

ZEITPUNKT	BENUTZER	KI-TOOL	AKTION	RISIKO	PI	BEGRÜNDUNG
18.2.2026, 08:46:00	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	10	-
18.2.2026, 08:45:15	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	10	-
18.2.2026, 08:44:45	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	10	-
18.2.2026, 08:39:20	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	10	-
18.2.2026, 08:37:30	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	1	-
18.2.2026, 08:36:14	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	9	-
16.2.2026, 09:15:08	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	8	-
16.2.2026, 08:03:53	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	MITTEL	1	-
16.2.2026, 08:02:30	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	9	-
16.2.2026, 07:25:02	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	10	-
5.2.2026, 09:59:31	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	9	-
5.2.2026, 09:31:00	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	9	-
5.2.2026, 09:29:12	psg-071c179d-69e9-4029-817a-60f9c28c4bf	ChatGPT	ANONYMSIERT	HOCH	9	-

Abb. 3.2: Audit-Log für Administrator

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