

Schatten-KI mit ChatGPT, Claude und Gemini als Datenschutzrisiko

Deine Belegschaft nutzt längst KI. Die Frage ist nur, ob du davon erfährst, bevor es die Aufsichtsbehörde tut.

Executive Summary

Irgendwo in deinem Unternehmen sitzt gerade jemand, kopiert eine Kundenreklamation samt Namen, Adresse und Vertragsnummer in ChatGPT und tippt darunter: „Formuliere das bitte freundlicher.“ Das ist kein böser Wille, das ist Effizienz. Und genau das macht **Schatten-KI** so tückisch: Sie wird von deinen fleißigsten Leuten betrieben, nicht von den faulsten.

Security-Experten schlagen Alarm. Bei einer Konferenz in Abu Dhabi warnte Kaspersky-Forscher Vladislav Tushkanov Anfang Oktober, dass Arbeitgeber oft schlicht keine Ahnung haben, welche KI-Systeme genutzt und welche Informationen dort verarbeitet werden. Die Risiken: Daten wandern in fremde Clouds, gehackte Privat-Accounts machen Firmengeheimnisse zur Beute, und über gefälschte KI-Apps und Browser-Erweiterungen kommt Schadsoftware gleich mit ins Haus.

Die Zahlen sind ernüchternd: Laut der aktuellen IBM-Studie zu Datenpannen war Schatten-KI 2026 an **43 Prozent** der Sicherheitsvorfälle in betroffenen Unternehmen beteiligt – ein Jahr zuvor waren es noch 20 Prozent. **68 Prozent** der betroffenen Firmen hatten keinerlei KI-Governance. Die gute Nachricht: Das Problem ist lösbar, und zwar nicht mit Verboten, sondern mit Transparenz und besseren Angeboten.

★ Wichtig: Die Kurzfassung für die Geschäftsführung

1. Schaff Transparenz: Finde heraus, welche KI-Tools wirklich genutzt werden.
2. Biete freigegebene Alternativen an, die mindestens so bequem sind wie die Gratis-Version.
3. Regle Datenklassen, Zugriffe und Verantwortung – auch für KI-Agenten.
4. Schule dein Team. Seit Februar 2025 verlangt der EU AI Act ohnehin KI-Kompetenz.

Worum geht es im Detail?

Schatten-KI ist die logische Fortsetzung der guten alten Schatten-IT. Früher war es die Dropbox, die sich die Vertriebsabteilung „nur mal kurz“ eingerichtet hat, oder die Excel-Datei mit Makros, von der das halbe Controlling abhing. Heute ist es der Browser-Tab mit ChatGPT, Claude oder Gemini, eingeloggt mit dem privaten Google-Konto. Der Unterschied: Eine Dropbox speichert Dateien. Ein Sprachmodell **liest, versteht und verarbeitet** sie – und je nach Tarif und Einstellung landet der Inhalt in Logs, in Aufbewahrungsfristen oder im Trainingsmaterial.

Warum passiert das überhaupt? Weil KI-Tools schlicht zu gut funktionieren. Eine Umfrage unter 1.250 Büroangestellten großer Unternehmen ergab 2026: Zwei Drittel nutzen KI bei der Arbeit, **obwohl sie glauben, dass es verboten ist**. 43 Prozent haben E-Mails ohne Freigabe in KI-Tools kopiert, 34 Prozent Kundendaten. Die Leute brechen die Regeln nicht aus Rebellion, sondern

weil sie Deadlines haben und das offizielle Tool entweder fehlt oder sich bedient wie ein Behördenformular von 1997.

Was im Prompt-Fenster landet – und was es kostet

Links: Anteil der Befragten, die Inhalte ohne IT-Freigabe in KI-Tools geteilt haben (Umfrage 2026, Büroangestellte großer Firmen)
Rechts: Anteil der Sicherheitsvorfälle mit Schatten-KI-Beteiligung bei betroffenen Unternehmen (IBM/Ponemon)

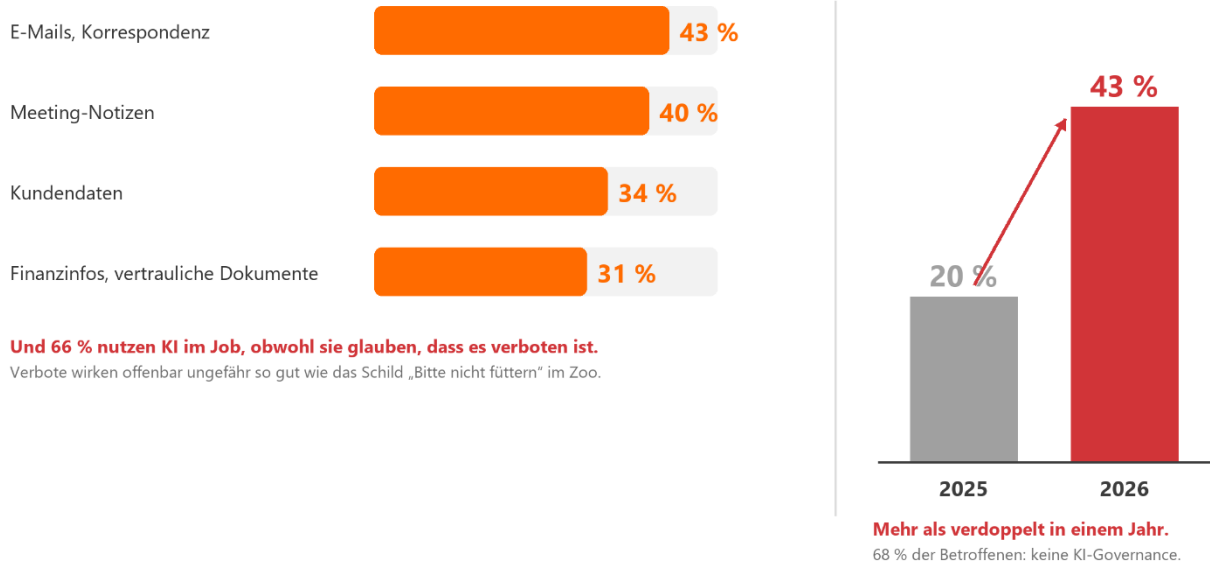


Abbildung 1: Was Beschäftigte ohne Freigabe in KI-Tools teilen – und wie stark der Anteil von Schatten-KI an Vorfällen wächst.

Erstens: der Datenabfluss. Jeder Prompt ist eine Datenübermittlung an einen Dienstleister – meist in den USA. Ohne Auftragsverarbeitungsvertrag nach Art. 28 DSGVO, ohne geprüfte Rechtsgrundlage für den Drittlandtransfer und ohne Löschkonzept. Bei den Verbraucher-Tarifen der großen Anbieter ist die Nutzung von Chats für das Modelltraining je nach Einstellung erlaubt; wer das Opt-out nicht findet, spendet seine Vertragsentwürfe der Wissenschaft. Dazu kommen Aufbewahrungsfristen, Sicherheitsprüfungen durch menschliche Reviewer und gelegentlich gerichtliche Anordnungen, Chatverläufe aufzuheben.

Zweitens: der Account selbst. Ein privater KI-Account hat selten MFA, nie Conditional Access und ist mit derselben E-Mail-Adresse verknüpft, die schon in drei Leaks auftaucht. Wird er gekapert, liegt der komplette Chatverlauf offen: jede Inhaltsliste, jedes Angebot, jede „Fass mir das Board-Protokoll zusammen“-Anfrage. Ein Chatverlauf ist im Grunde das Tagebuch deines Unternehmens, nur ohne Schloss.

Drittens: Schadsoftware. Genau darauf zielen die Experten ab, wenn sie vor eingeschleppter Malware warnen. Anfang 2026 fanden Forscher mindestens 16 bösartige Chrome-Erweiterungen, die sich als praktische ChatGPT-Helfer ausgaben – Ordnerverwaltung, Prompt-Bibliothek, Chat-Suche – und dabei Sitzungs-Token abgriffen. Selbst die offizielle Claude-Erweiterung hatte im Januar eine Lücke, über die Chatverläufe ausgelesen werden konnten; Anthropic hat innerhalb kurzer Zeit gepatcht. Dazu kommt KI-generierter Code, den jemand ungeprüft aus dem Chatfenster in Produktion kopiert. Was soll schon schiefgehen?

⚠ Warnung: Die Browser-Erweiterung ist das neue Makro

„ChatGPT Sidebar Pro Ultra Free“ mit 4,8 Sternen ist kein Produktivitätstool, sondern im Zweifel ein Datenstaubsauger mit Vollzugriff auf jede geöffnete Webseite – inklusive deines CRM, deiner Personalakte und deines Online-Bankings. Wenn du Erweiterungen nicht per Richtlinie auf eine Positivliste beschränkst, hast du keine Endpoint-Security, sondern ein Gerücht.

Zwei Wege, ein Prompt: wohin die Daten wirklich gehen

Dieselbe Aufgabe – „Fass mir mal diesen Vertrag zusammen“ – mit sehr unterschiedlichem Ausgang

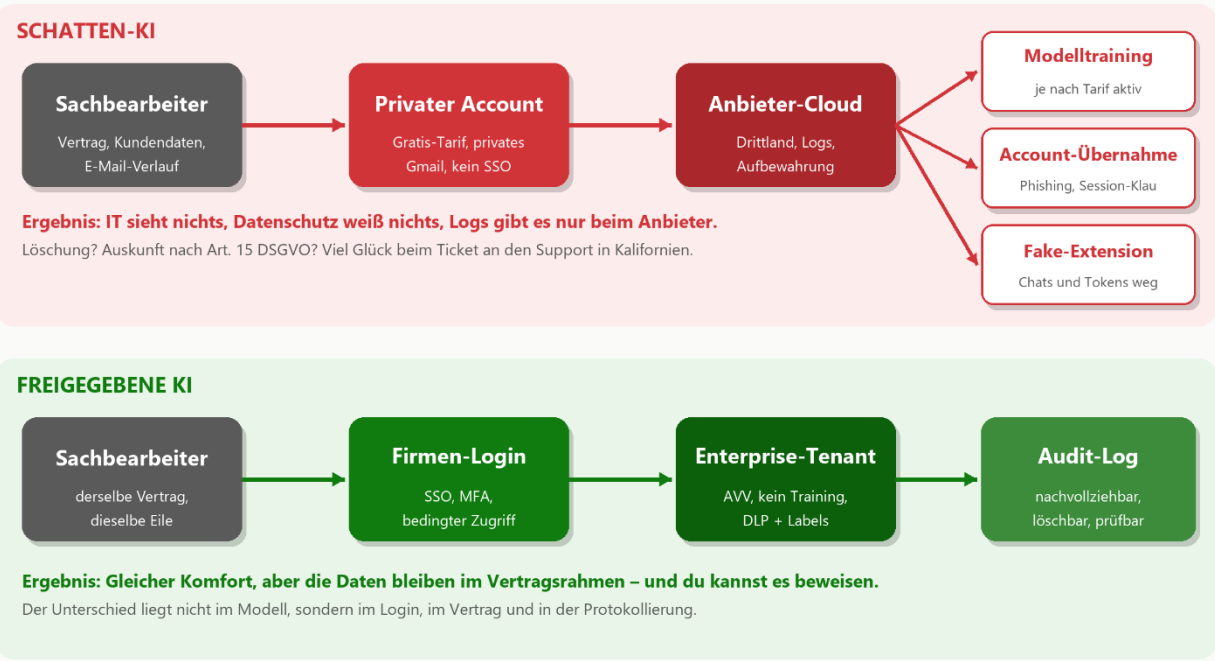


Abbildung 2: Derselbe Prompt, zwei Wege – der Unterschied liegt im Login, im Vertrag und in der Protokollierung.

Die nächste Eskalationsstufe ist schon da: **KI-Agenten**. Die können nicht nur antworten, sondern handeln – E-Mails verschicken, Tickets schließen, Dateien verschieben. Die Regierung der Vereinigten Arabischen Emirate will agentische KI innerhalb von zwei Jahren in der Hälfte ihrer Bundesbehörden einsetzen. Deloitte's KI-Chef Costi Perricos bringt das Problem auf den Punkt: Ein Agent kann nicht zur Verantwortung gezogen werden, also braucht jeder einen menschlichen Verantwortlichen. Ein Schatten-Agent mit Zugriff auf das Postfach eines Abteilungsleiters ist daher kein Tool mehr, sondern ein unbeaufsichtigter Praktikant mit Generalvollmacht.

Kriterium	Gratis-/Privat-Tarif	Enterprise-Tarif (z. B. ChatGPT Enterprise, Claude Enterprise, Gemini in Workspace, Copilot)
Anmeldung	Private E-Mail, oft ohne MFA	Firmen-SSO, MFA, bedingter Zugriff
Modelltraining	Je nach Einstellung aktiv, Opt-out nötig	Vertraglich ausgeschlossen

Kriterium	Gratis-/Privat-Tarif	Enterprise-Tarif (z. B. ChatGPT Enterprise, Claude Enterprise, Gemini in Workspace, Copilot)
Vertrag	Nutzungsbedingungen, kein AVV	AVV nach Art. 28 DSGVO, Datenresidenz-Optionen
Protokollierung	Nur beim Anbieter, für dich unsichtbar	Audit-Logs, eDiscovery, Aufbewahrungsrichtlinien
Offboarding	Account geht mit dem Mitarbeiter nach Hause	Zugang endet mit dem Firmenkonto
Kosten	0 Euro – du bezahlst mit Daten	Lizenzkosten, dafür planbar und prüfbar

Was sind Chancen? Was sind Risiken?

Fangen wir mit der unbequemen Wahrheit an: Schatten-KI ist auch ein **Geschenk**. Sie ist die ehrlichste Anforderungsanalyse, die du je bekommen wirst. Wenn die Buchhaltung heimlich Rechnungen mit Gemini auswertet, sagt dir das genau, wo der Schmerz sitzt und welcher Prozess nach Automatisierung schreit. Das hätte dir kein Workshop mit Moderationskarten verraten.

Wir hatten einen Mittelständler mit rund 400 Beschäftigten, der ChatGPT 2023 per Firewall gesperrt hatte. Stolz präsentierte die IT-Leitung das Ergebnis: null Zugriffe. Eine anonyme Umfrage ergab dann, dass rund 60 Prozent des Teams KI auf dem privaten Smartphone nutzten – inklusive Fotos von Konstruktionszeichnungen, „weil die Texterkennung so gut ist“. Die Sperre hatte das Problem nicht gelöst, sondern nur aus dem Firmennetz ins Mobilfunknetz verlagert. Dort sieht es garantiert niemand mehr. Glückwunsch.

Chancen	Risiken
Sichtbarer Bedarf: Schatten-Nutzung zeigt, wo KI echten Mehrwert stiftet	Datenabfluss von Kundendaten und Geschäftsgeheimnissen an Dritte
Power-User als Multiplikatoren für den offiziellen Rollout	DSGVO-Verstöße: fehlender AVV, unzulässiger Drittlandtransfer, keine Lösbarkeit
Messbare Produktivität, wenn Tools legal und integriert verfügbar sind	Account-Übernahme mit vollem Zugriff auf den Chatverlauf
Bessere Verhandlungsposition gegenüber Anbietern durch echte Nutzungsdaten	Malware über gefälschte Apps, Erweiterungen und ungeprüften KI-Code
Aufbau von KI-Kompetenz, die der EU AI Act ohnehin verlangt	Fehlentscheidungen durch halluzinierte Ergebnisse ohne Nachvollziehbarkeit

Das Risiko ist dabei nicht nur theoretisch. Der Klassiker bleibt Samsung: 2023 kopierten Ingenieure vertraulichen Quellcode und interne Besprechungsnotizen in ChatGPT, um Fehler zu

finden und Protokolle zu erstellen. Ergebnis: ein konzernweites Verbot und ein Lehrstück, das seither in jeder Security-Präsentation auftaucht. Drei Jahre später ist das Muster dasselbe, nur die Tools sind besser und die Nutzung ist flächendeckend.

1 Fakten: Was eine Datenpanne mit KI-Beteiligung kostet

IBM und Ponemon haben 2026 über 600 betroffene Unternehmen in 16 Ländern untersucht. Vorfälle mit Schatten-KI kosteten im Schnitt **5,39 Millionen US-Dollar**, rund jeder fünfte endete mit einem Bußgeld. Im Nahen Osten lagen die Durchschnittskosten einer Datenpanne bei 8 Millionen US-Dollar, 26 Prozent der böswilligen Vorfälle dort waren KI-gestützt. Nur 38 Prozent der Unternehmen haben strenge Freigabeprozesse für KI-Tools.

Und dann ist da noch die Haftungsfrage. Wenn ein Vertriebler einen Angebotstext mit halluzinierter Rabattstaffel verschickt, haftet nicht die KI. Wenn eine Personalerin Bewerberdaten in ein privates Tool kippt, ist dein Unternehmen Verantwortlicher im Sinne der DSGVO – nicht OpenAI, nicht Google, nicht Anthropic. Die Aufsichtsbehörde fragt nicht, welches Modell es war. Sie fragt, wo deine Dokumentation ist.

Was müssen wir jetzt schon vorbereiten?

Premchand Kurup, CEO von ParamountAssure, beschreibt die Reihenfolge so: erst die nicht genehmigte Nutzung identifizieren, dann freigegebene Alternativen und klare Zugriffsgrenzen festlegen, und Sicherheit von Anfang an in die Abläufe einbauen. Das ist kein Hexenwerk, aber es braucht Disziplin. Vor allem in der Reihenfolge: Wer mit Verboten anfängt, treibt die Nutzung in den Untergrund und verliert genau die Sichtbarkeit, die er braucht.



Abbildung 3: Ein realistischer 90-Tage-Fahrplan – erst hinsehen, dann anbieten, dann regeln.

Sichtbar machen. Werte Proxy-, Firewall- und DNS-Logs aus. Mit Microsoft Defender for Cloud Apps oder einem anderen CASB siehst du, welche KI-Dienste wie oft aufgerufen werden und wie viel Daten hochgeladen wurden. Inventarisiere Browser-Erweiterungen über Intune oder deine Endpoint-Verwaltung. Und frag die Leute einfach – anonym. Du wirst staunen, wie ehrlich die Antworten sind, wenn niemand Angst vor einer Abmahnung hat.

Alternative liefern. Entscheide dich für mindestens ein Enterprise-Angebot mit AVV und ohne Training auf deinen Daten. Ob das Microsoft 365 Copilot, ChatGPT Enterprise, Claude Enterprise

oder Gemini in Google Workspace ist, hängt von deinem Ökosystem ab. Entscheidend ist: Das freigegebene Tool muss **mindestens so gut** sein wie die Gratis-Version, sonst kehren die Leute nach einer Woche zurück zum Privat-Account. Ein abgespecktes Modell mit drei Klicks mehr ist keine Alternative, sondern ein Rückfall mit Ansage.

✓ Tipp: Die Power-User zuerst

Die Leute, die heute am meisten Schatten-KI betreiben, sind deine besten Pilotnutzer. Gib ihnen als Erste den Enterprise-Zugang und mach sie zu Botschaftern. Wer bisher heimlich Prompts optimiert hat, erklärt den Kollegen gern, wie es richtig geht – und ist plötzlich Teil der Lösung statt Teil des Audit-Befunds.

Leitplanken ziehen. Schreib eine KI-Richtlinie, die auf eine Seite passt. Niemand liest 40 Seiten, und wer sie liest, hat offensichtlich zu wenig zu tun. Definiere Datenklassen: Was darf in welches Tool? Öffentliche Informationen überall, interne Daten nur ins Enterprise-Tool, personenbezogene und streng vertrauliche Daten nur mit Freigabe. Setze Vertraulichkeitsbezeichnungen und DLP-Regeln technisch durch, zum Beispiel mit Microsoft Purview, das Uploads sensibler Inhalte in nicht freigegebene KI-Seiten im Browser blockieren kann. **Erst dann** sperrst du Gratis-Dienste.

Agenten regieren wie Personal. Snowflake-Manager Ahmed Hafez und Deloitte empfehlen, KI-Agenten wie eine digitale Belegschaft zu führen: Jeder Agent bekommt einen menschlichen Owner, minimale Berechtigungen, protokollierte Aktivitäten und einen Ausschalter, von dem jemand weiß, wo er ist. Neue Agenten testest du erst in einer kontrollierten Umgebung, bevor sie an Produktivdaten dürfen. Klingt nach Bürokratie, ist aber deutlich billiger als der Moment, in dem ein Agent fröhlich 2.000 Kunden eine Kündigungsbestätigung schickt.

Befähigen. Seit dem 2. Februar 2025 verpflichtet Art. 4 des EU AI Act Unternehmen, die KI einsetzen, für ausreichende KI-Kompetenz ihres Personals zu sorgen. Nutze das als Rückenwind: Schulungen, die zeigen, was in einen Prompt gehört und was nicht, und warum KI-Ergebnisse geprüft werden müssen. Kaspersky-Forscher Tushkanov empfiehlt genau das: Beschäftigte darin zu schulen, ausschließlich interne KI-Systeme für Firmeninformationen zu verwenden.

⚠ Warnung: Die Sperre allein ist Sicherheitstheater

Wer ChatGPT an der Firewall blockiert und sich dann zurücklehnt, hat das Problem auf private Smartphones verlagert – wo es garantiert niemand mehr protokolliert. Ein Verbot ohne Alternative ist das Compliance-Gegenstück zu einem Rauchverbot ohne Raucherecke: Am Ende qualmt es auf der Toilette, und der Rauchmelder ist abgeklebt.

Maßnahme	Verantwortlich	Aufwand	Wirkung
Schatten-KI-Inventur über CASB und Logs	IT-Security	niedrig	hoch
Enterprise-KI mit AVV bereitstellen	IT + Einkauf + Datenschutz	mittel	sehr hoch
KI-Richtlinie (eine Seite) und Datenklassen	Datenschutz + Betriebsrat	niedrig	hoch

Maßnahme	Verantwortlich	Aufwand	Wirkung
DLP und Vertraulichkeitsbezeichnungen	IT-Security	mittel	hoch
Positivliste für Browser-Erweiterungen	Endpoint-Team	niedrig	hoch
Agenten-Register mit Owner und Ausschalter	Fachbereiche + IT	mittel	mittel bis hoch
KI-Kompetenz-Schulung nach Art. 4 AI Act	HR + Fachbereiche	mittel	hoch

★ Wichtig: Betriebsrat früh einbinden

Sobald du Nutzung von KI-Diensten auswertest oder KI-Tools einführest, ist in Deutschland regelmäßig die Mitbestimmung nach § 87 Abs. 1 Nr. 6 BetrVG berührt. Hol den Betriebsrat an den Tisch, bevor du die erste Log-Auswertung startest. Das spart Monate – und erspart dir eine Einigungsstelle, die zufällig genau dann tagt, wenn der Rollout starten soll.

Häufig gestellte Fragen

Was versteht man unter Schatten-KI?

Schatten-KI bezeichnet den Einsatz von KI-Tools wie ChatGPT, Claude oder Gemini im Arbeitsalltag, ohne dass die IT oder der Datenschutz sie freigegeben hat. Typisch ist die Nutzung über private Accounts und kostenlose Tarife, bei denen das Unternehmen weder Verträge noch Kontrolle über die eingegebenen Daten hat.

Darf ich personenbezogene Daten in ChatGPT eingeben?

In einen privaten oder kostenlosen Account grundsätzlich nicht, weil dafür kein Auftragsverarbeitungsvertrag nach Art. 28 DSGVO besteht und die Übermittlung in die USA nicht abgesichert ist. Mit einem Enterprise-Tarif samt AVV, ausgeschlossenen Modelltraining und dokumentierter Rechtsgrundlage kann es zulässig sein, sofern die interne Richtlinie es erlaubt.

Wie erkenne ich, welche KI-Tools im Unternehmen genutzt werden?

Am zuverlässigsten über die Auswertung von Proxy-, Firewall- und DNS-Logs mit einem Cloud Access Security Broker wie Microsoft Defender for Cloud Apps, ergänzt um ein Inventar der installierten Browser-Erweiterungen. Eine anonyme Umfrage zeigt zusätzlich die Nutzung auf privaten Geräten, die technisch nicht sichtbar ist.

Reicht es, ChatGPT und andere KI-Dienste per Firewall zu sperren?

Nein, eine reine Sperre verlagert die Nutzung meist auf private Smartphones und macht sie damit vollständig unsichtbar. Wirksam ist erst die Kombination aus einer gleichwertigen, freigegebenen Alternative, klaren Regeln und technischen Schutzmaßnahmen wie DLP.

Was verlangt der EU AI Act von Unternehmen, die KI-Tools einsetzen?

Seit dem 2. Februar 2025 müssen Unternehmen nach Art. 4 AI Act sicherstellen, dass ihr Personal über ausreichende KI-Kompetenz verfügt. Das umfasst praktisch Schulungen zum sicheren Umgang mit KI, zu ihren Grenzen und zu den Risiken bei der Eingabe sensibler Daten.