



WARUM UNTERNEHMEN EINEN PRIVATEN KI-ARBEITSBEREICH BRAUCHEN, BEVOR SENSIBLE DATEN DAS UNTERNEHMEN VERLASSEN

Erstellt von cogify

Implementierung eines IVY-gesteuerten privaten KI-Arbeitsbereichs,
Wartung und Cybersecurity-Beratung



MANAGEMENT-SUMMARY

Künstliche Intelligenz ist bereits Teil des Arbeitsalltags. **Mitarbeitende nutzen öffentliche KI-Tools**, um Dokumente zusammenzufassen, E-Mails zu überarbeiten, Inhalte zu übersetzen, Präsentationen vorzubereiten, Berichte zu analysieren, den Vertrieb zu unterstützen, Verträge zu prüfen, Code zu schreiben und wiederkehrende Aufgaben zu beschleunigen. Das ist nicht automatisch ein Problem. KI kann die Produktivität steigern, manuelle Arbeit reduzieren und Teams dabei helfen, schneller voranzukommen.

Das Problem beginnt dann, wenn Unternehmen nicht wissen, welche KI-Tools von den Mitarbeitenden genutzt werden, welche Daten eingegeben werden, wo diese Daten verarbeitet werden, ob Zugriffsrechte eingehalten werden und ob die **Sicherheitsregeln des Unternehmens** Beachtung finden.

Diese unkontrollierte Nutzung von KI wird oft als **Shadow AI (Schatten-KI)** bezeichnet.

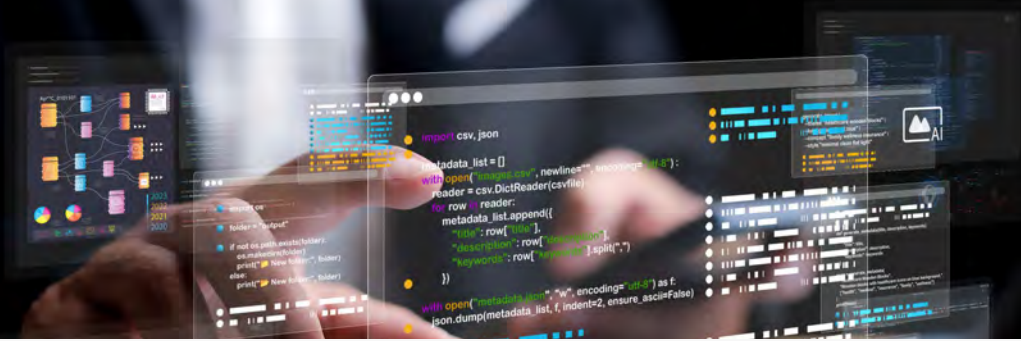
Shadow AI stellt ein **Cybersecurity- und Datenschutzrisiko** dar, da sensible Unternehmensinformationen durch die alltägliche Verwendung der Mitarbeitenden in die Öffentlichkeit geraten:

- Kopieren von Dokumenten in öffentliche KI-Tools
- Hochladen von Dateien zur Zusammenfassung
- Einfügen von Quellcode in Chatbots
- Umformulieren von Text mit **vertrauliche Kundeninhalte**

Für technische Teams kann das Risiko noch gravierender sein. Entwickler und IT-Mitarbeitende fügen beim Debuggen, Dokumentieren oder Verstehen eines Problems oft Fehlerprotokolle (Logs), API-Keys, Access-Tokens, Umgebungsvariablen, Datenbankdetails, Systemarchitekturen sowie technische **Benutzernamen und Passwörter** in öffentliche KI-Tools ein, um schneller eine Lösung zu finden.

Dies geschieht in der Regel nicht in böser Absicht. Die Mitarbeitenden versuchen lediglich, effizienter zu arbeiten. Aus Sicht der Cybersecurity kann das Ergebnis dennoch eine schwerwiegende **Offenlegung von Daten sein**.





Die Lösung besteht nicht darin, KI komplett zu verbieten. Mitarbeitende nutzen KI, weil sie ihnen hilft, schneller zu arbeiten. Wenn Unternehmen keine sichere und nützliche Alternative anbieten, nutzen die Mitarbeitenden öffentliche Tools möglicherweise über private Konten, Browser-Erweiterungen, mobile Apps oder unkontrollierte Workflows weiter. Der bessere Ansatz ist die **sichere KI-Einführung** durch einen privaten KI-Arbeitsbereich.

cogify unterstützt Unternehmen bei der Implementierung, dem Betrieb und der Wartung eines privaten KI-Arbeitsbereichs, in dem Mitarbeitende KI für das Unternehmenswissen, die Dokumentensuche, Zusammenfassungen und die tägliche Arbeit nutzen können – ohne sensible Informationen an öffentliche KI-Tools zu senden. Der Arbeitsbereich kann auf IVY basieren, der privaten KI-Technologieschicht von cogify. Die kundenorientierte Lösung ist einfach: ein **privater KI-Arbeitsbereich**, bei dem die Unternehmensdaten in einer kontrollierten Umgebung innerhalb des Unternehmens verbleiben.

Diese Umgebung kann auf der **Schweizer Infrastruktur** von cogify gehostet oder **direkt auf dem eigenen Server** bzw. in der privaten Umgebung des Kunden installiert werden – je nach den Sicherheits-, Datenschutz- und Betriebsanforderungen des Kunden.

Innerhalb des Arbeitsbereichs werden die Unternehmensdaten durch **separate Workspaces**, rollenbasierte Zugriffsrechte, freigegebene Wissensquellen und klare Einschränkungen kontrolliert. **Mitarbeitende greifen** nur auf die Informationen zu, die sie auch sehen dürfen.

HR-Dokumente können auf den HR-Bereich beschränkt werden. Finanzdokumente können dem Finanzbereich und dem Management vorbehalten bleiben. Kundenprojektdaten können pro Kunde oder Projektteam getrennt werden. Technische Dokumentationen können auf Entwickler und die IT beschränkt werden. Vorstands- und Strategiedokumente können der Führungsebene vorbehalten bleiben.

Dies bietet Unternehmen eine praktische Alternative zur unkontrollierten Nutzung öffentlicher KI. **Die Mitarbeitenden erhalten KI für ihre tägliche Arbeit.**

Das Management behält die Kontrolle über Daten, Zugriffsrechte, Hosting und das Cybersecurity-Risiko. Das Unternehmenswissen bleibt in einer kontrollierten Unternehmensumgebung und verlässt das Unternehmen zu keinem Zeitpunkt.

INHALTSVERZEICHNIS

Seite

1	1. Die neue Realität: KI ist bereits am Arbeitsplatz angekommen
2	2. Was ist Shadow AI?
3	3. Warum Shadow AI ein Cybersecurity-Problem ist
5	4. Warnsignale aus dem Markt
6	5. Warum Shadow AI für Schweizer Unternehmen wichtig ist
7	6. Praktische Schweizer Risikoszenarien
10	7. Warum „KI einfach verbieten“ nicht ausreicht
11	8. Was Unternehmen stattdessen brauchen
12	9. Wie Daten innerhalb des Unternehmens kontrolliert werden
15	10. Wie cogify unterstützt
17	11. KI-Sicherheits-Checkliste für das Management
18	12. Empfohlene erste Schritte
19	Fazit
20	Referenzen & Weiterführende Literatur

1. DIE NEUE REALITÄT: KI IST BEREITS AM ARBEITSPLATZ ANGEKOMMEN



KI IST KEIN ZUKUNFTSTREND MEHR. SIE IST BEREITS TEIL DES TÄGLICHEN GESCHÄFTS.

Mitarbeitende nutzen KI-Tools, um Berichte zusammenzufassen, E-Mails zu überarbeiten, Angebote zu entwerfen, Dokumente zu übersetzen, Tabellenkalkulationen zu analysieren, Verträge zu prüfen, technische Inhalte zu erklären, Code zu debuggen, Kundenantworten vorzubereiten, Besprechungsnotizen zu erstellen und Marketing- oder Vertriebsinhalte zu verbessern.

Die meisten Mitarbeitenden tun dies in guter Absicht. Sie wollen dem Unternehmen nicht schaden. Sie versuchen, Zeit zu sparen, die Qualität zu verbessern und den Erwartungen gerecht zu werden.

Aus Sicht der Cybersecurity schafft die unkontrollierte Nutzung von KI jedoch einen gefährlichen blinden Fleck.



Ein Unternehmen weiss oft nicht:

- welche KI-Tools genutzt werden
- ob Mitarbeitende private Konten oder Unternehmenskonten verwenden
- ob vertrauliche Dokumente in öffentliche Tools kopiert werden
- ob Kunden- oder Mitarbeiterdaten geteilt werden
- ob Quellcode, Zugangsdaten oder die technische Architektur offengelegt werden
- wo KI-Prompts und hochgeladene Dateien gespeichert werden
- ob Zugriffsrechte eingehalten werden
- ob KI-generierte Antworten korrekt oder sicher sind
- ob die KI-Nutzung mit den internen Richtlinien übereinstimmt

Die Kernherausforderung ist einfach:

Die Produktivität entwickelt sich schneller als die Governance.

2. WAS IST SHADOW AI?

Shadow AI bezeichnet den Einsatz von Werkzeugen der künstlichen Intelligenz innerhalb einer Organisation ohne die entsprechende Genehmigung, Governance, Transparenz oder Kontrolle durch die IT/Management.

Sie ähnelt der Schatten-IT (Shadow IT), bei der Mitarbeitende nicht genehmigte Tools oder Cloud-Dienste nutzen, um ihre Arbeit schneller zu erledigen. Shadow AI kann jedoch noch sensibler sein, da KI-Tools Daten nicht nur speichern. Sie verarbeiten, fassen zusammen, formulieren um, ziehen Schlüsse und generieren neue Ergebnisse aus den Informationen, die Mitarbeitende bereitstellen.

Das bedeutet, dass ein einfaches Kopieren und Einfügen Informationen offenlegen kann, die normalerweise innerhalb der Unternehmenssysteme geschützt wären.

SENSIBLE INFORMATIONEN, DIE IN UNKONTROLLIERTE KI-TOOLS GELANGEN KÖNNEN:

- Kundendokumente
- Verträge und rechtliche Unterlagen
- Interne Berichte
- Vorstandsdokumente
- Quellcode
- API-Keys und Access-Tokens
- Benutzernamen und Passwörter
- Datenbank-Zugangsdaten
- Umgebungsvariablen
- Fehlerprotokolle und System-Traces
- Interne URLs
- Server- und Cloud-Konfigurationen
- Mitarbeiterdaten / Personendaten
- HR-Dokumente / Personalunterlagen
- Finanzdaten
- Vertriebs-Pipelines / Sales-Pipelines
- Produkt-Roadmaps
- Unternehmensstrategie
- Support-Tickets
- Sicherheitsverfahren / Security-Prozeduren
- Kundenkommunikation
- Projektdokumentation
- Lieferanteninformationen
- Dokumente mit Bezug zu Behörden

In den meisten Fällen werden diese Informationen nicht durch einen Cyberangriff offengelegt, sondern durch das ganz alltägliche Verhalten der Mitarbeitenden. Jemand möchte eine schnellere Zusammenfassung, ein Manager wünscht sich eine fehlerfreie E-Mail, ein Entwickler benötigt Hilfe beim Code, ein Vertriebsmitarbeiter möchte ein Angebot überarbeiten oder ein Finanzmitarbeiter braucht Unterstützung bei der Erklärung einer Tabelle. Die Absicht ist Produktivität.

Das Ergebnis kann dennoch der Abfluss von Daten sein.

3. WARUM SHADOW AI EIN CYBERSECURITY-PROBLEM IST

Shadow AI sollte nicht nur als Innovationsthema behandelt werden. Es ist in erster Linie ein Cybersecurity-Problem.

Traditionelle Cybersecurity schützt Systeme, Dateien, Zugriffsrechte, E-Mails, Netzwerke und Geräte. Öffentliche KI-Tools schaffen jedoch einen neuen Pfad, über den sensible Informationen das Unternehmen verlassen können. Ein vertrauliches Dokument mag im internen System des Unternehmens geschützt sein. Wenn ein Mitarbeiter den Inhalt jedoch in ein externes KI-Tool kopiert, spielt die ursprüngliche Zugriffskontrolle keine Rolle mehr.

Shadow AI birgt mehrere Risiken.

DATENABFLUSS (DATA LEAKAGE)

Vertrauliche Geschäftsdaten können mit einem Tool geteilt werden, das vom Unternehmen weder genehmigt noch überwacht oder vertraglich kontrolliert wird.

VERLUST DER TRANSPARENZ

IT- und Sicherheitsteams wissen unter Umständen nicht, welche Daten von wem, mit welchem Tool und zu welchem Zweck geteilt wurden.



VERLUST DER ZUGRIFFSKONTROLLE

Informationen aus zugriffsbeschränkten Dokumenten können in einen KI-Chat kopiert und anschliessend mit Personen geteilt werden, die eigentlich keinen Zugriff auf den Originalinhalt haben sollten.

COMPLIANCE- UND DATENSCHUTZRISIKEN

Personendaten, Kundendaten, Mitarbeiterdaten oder regulierte Informationen werden möglicherweise ausserhalb genehmigter Umgebungen verarbeitet.



(oo) OFFENLEGUNG VON GEISTIGEM EIGENTUM

Quellcode, technische Architekturen, Produktideen, Forschungsergebnisse, geschützte Algorithmen und geschäftliches Know-how können über Prompts oder hochgeladene Dateien abfließen.

UNZUVERLÄSSIGE ERGEBNISSE

KI-generierte Antworten mögen professionell aussehen, können aber Fehler, unbelegte Behauptungen, unsicheren Code oder falsche Annahmen enthalten.

ERWEITERTE ANGRIFFSFLÄCHE

KI-gestützte Workflows, Chatbots und Agenten können manipuliert werden, wenn sie ohne angemessene Verifizierung, Zugriffskontrolle, Prüfbarkeit und menschliche Aufsicht mit sensiblen Aktionen verknüpft sind.

OFFENLEGUNG VON ZUGANGSDATEN UND INFRASTRUKTUR

Mitarbeitende können versehentlich technische Benutzernamen, Passwörter, API-Keys, Access-Tokens, Datenbank-Zugangsdaten, Umgebungsvariablen, Server-Logs, interne URLs, Cloud-Konfigurationsdetails oder die Systemarchitektur in öffentliche KI-Tools einfügen.

Das ist besonders gefährlich, da technische Informationen direkt von Angreifern genutzt werden können. Selbst eine kurze Log-Datei oder Fehlermeldung kann Zugangsdaten, Tokens, Systemnamen, Endpoint-URLs oder Details darüber preisgeben, wie die Systeme eines Unternehmens aufgebaut sind und wie auf sie zugegriffen wird.

4. WARNSIGNALE AUS DEM MARKT

Der Zweck von Marktbeispielen besteht nicht darin, zu behaupten, dass jeder KI-Vorfall gleich abläuft. Einige Fälle betreffen den durch Mitarbeitende verursachten Datenabfluss. Andere betreffen die Ausnutzung von KI-Workflows oder allgemeinere KI-Sicherheitsbedenken. Zusammenfassend zeigen sie eine klare Lektion:

Unternehmen benötigen Governance, Sicherheit und kontrollierte Umgebungen für die KI-Nutzung.

Praxisbeispiele aus Unternehmen haben gezeigt, dass Mitarbeitende sensible interne Informationen, einschliesslich Quellcode, in öffentliche KI-Tools eingeben. Beispiele aus dem öffentlichen Sektor verdeutlichen, dass selbst sicherheitsbewusste Umgebungen vor Herausforderungen stehen, wenn sensible Dokumente auf öffentliche KI-Plattformen hochgeladen werden. KI-gestützte Workflows haben zudem gezeigt, dass Automatisierungen, die mit sensiblen Aktionen verknüpft sind, mit starker Verifizierung, Zugriffskontrolle und Aufsicht gestaltet werden müssen.



Die Lehre daraus ist nicht, dass Unternehmen KI meiden sollten. Die Lehre ist, dass KI sicher eingeführt werden muss.

Unternehmen benötigen klare Regeln, genehmigte Tools, kontrollierte Zugriffsrechte, sichere Hosting-Modelle, Richtlinien für Mitarbeitende und kontinuierliche Überprüfungen. Ohne diese Massnahmen erfolgt die KI-Einführung informell, und sensible Daten können das Unternehmen verlassen, noch bevor das Management überhaupt bemerkt, dass ein Problem existiert.

5. SCHWEIZER KONTEXT: WARUM SHADOW AI FÜR SCHWEIZER UNTERNEHMEN WICHTIG IST

Es gibt derzeit keine breit bestätigten öffentlichen Schweizer Fälle, in denen ein Unternehmen spezifisch deshalb gehackt wurde, weil Mitarbeitende KI-Tools übermässig genutzt haben. Dennoch befindet sich die Schweiz bereits in einem realen Cybersecurity-Risikoumfeld. Schweizer Organisationen sehen sich mit Ransomware, Phishing, Datenerpressung, Software-Schwachstellen, Supply-Chain-Angriffen und Risiken des Datenabflusses konfrontiert.

Der Fall Xplain hat zudem gezeigt, wie schwerwiegend eine Datenoffenlegung werden kann, wenn sensible Informationen im Zusammenhang mit Bundesbehörden, kantonalen Behörden und Lieferanten geleakt werden. Dies war zwar kein KI-Vorfall, aber es ist eine eindrückliche Schweizer Mahnung, dass Datenexposition, Lieferantenrisiken und eine schwache Kontrolle über sensible Informationen erhebliche Konsequenzen haben können.

Shadow AI sollte daher als eine zusätzliche Risikoschicht verstanden werden.

Sie ersetzt die traditionellen Cybersecurity-Risiken nicht, sondern schafft einen neuen Weg, wie sensible Informationen durch das alltägliche Verhalten der Mitarbeitenden abfliessen können:

- ⚠ Kopieren und Einfügen von Dokumenten in öffentliche KI-Tools
- ⚠ Hochladen von Dateien zur Zusammenfassung
- ⚠ Übersetzen von vertraulichen Inhalten
- ⚠ Aufforderung an die KI, Kundendaten zu analysieren
- ⚠ Nutzung von KI zur Überprüfung von Quellcode
- ⚠ Einfügen von Finanzdaten in Chat-Prompts
- ⚠ Einfügen von Benutzernamen, Passwörtern, Tokens oder API-Keys in Prompts
- ⚠ Generieren von Ergebnissen aus zugriffsbeschränkten internen Informationen

Für Schweizer KMU, Treuhänder, Verbände, Softwareunternehmen, Organisationen im Gesundheitswesen, Zulieferer des öffentlichen Sektors und Beratungsfirmen ist dies von grosser Bedeutung, da viele Teams täglich mit hochsensiblen Informationen arbeiten.

Ein privater KI-Arbeitsbereich hilft Unternehmen dabei, ihren Mitarbeitenden eine sicherere Alternative zu unkontrollierten öffentlichen KI-Tools zu bieten und gleichzeitig das Unternehmenswissen, die Zugriffsrechte und die Datenhaltung (Data Residency) besser unter Kontrolle zu halten.

6. PRAKTISCHE SCHWEIZER RISIKOSZENARIEN

Die folgenden Szenarien werden nicht als bestätigte, durch KI verursachte Schweizer Hacking-Vorfälle dargestellt. Sie zeigen vielmehr auf, wie Shadow-AI-Risiken in typischen Schweizer Geschäftsumgebungen auftreten können.

TREUHAND- ODER WIRTSCHAFTSPRÜFUNGSGESELLSCHAFT

Ein Treuhand-Team nutzt KI, um Kundendokumente zusammenzufassen, E-Mails zu entwerfen, finanzbezogene Informationen zu erklären oder interne Notizen vorzubereiten.

Ohne einen kontrollierten KI-Arbeitsbereich fügen Mitarbeitende finanzielle Kundendokumente, Lohnabrechnungsdaten, Steuerunterlagen, Rechnungen oder geschützte Personendaten in öffentliche KI-Tools ein.

Ein privater KI-Arbeitsbereich kann das allgemeine Unternehmenswissen, Kundendokumente, Lohndaten, Steuerunterlagen und Managementinformationen in separate, zugriffsbeschränkte Workspaces unterteilen. Nutzende erhalten nur Zugriff auf die Bereiche, die ihrer Rolle und Verantwortung entsprechen. Für besonders strenge Anforderungen an die Vertraulichkeit von Kundendaten kann der Arbeitsbereich auf den eigenen Servern des Unternehmens oder in einer privaten Umgebung installiert werden.

SCHWEIZER KMU MIT KUNDENPROJEKTEN

Ein Schweizer Dienstleistungsunternehmen nutzt KI, um Kunden-E-Mails zu entwerfen, Projektdokumente zusammenzufassen, Angebote vorzubereiten und interne Inhalte zu übersetzen.

Kundennamen, Preise, vertragliche Details, Projektrisiken und die interne Strategie werden in unkontrollierte KI-Tools kopiert.

Ein privater KI-Arbeitsbereich kann nach Abteilung, Kunde, Projekt oder Sensibilitätsstufe unterteilt werden. Der Vertrieb kann mit freigegebenen Angebotsmaterialien arbeiten. Projektteams arbeiten mit ihren eigenen Kundendokumenten. Das Management kann Strategie-, Preis- und Vorstandsdokumente geschützt halten. Dies bietet den Mitarbeitenden einen nützlichen KI-Arbeitsbereich, ohne den Zugriff auf alle Unternehmensinformationen zu öffnen.



SOFTWARE- ODER TECHNOLOGIEUNTERNEHMEN

Entwickler und technische Teams nutzen öffentliche KI-Tools, um Code zu reparieren, Fehler zu analysieren, Architekturen zu überprüfen, Dokumentationen zu schreiben, Testfälle zu generieren oder das Systemverhalten schneller zu verstehen.

Dies kann Quellcode, Systemarchitekturen, API-Strukturen, Logs, Benutzernamen, Passwörter, API-Keys, Tokens, Datenbank-Zugangsdaten, Umgebungsvariablen, interne URLs, Cloud-Konfigurationen oder CI/CD-Informationen offenlegen. In Software- und Technologieunternehmen wird Shadow AI somit zu mehr als einem Dokumentationsrisiko – es wird zu einem Infrastruktur-, Zugriffs- und IP-Risiko (geistiges Eigentum).

Ein privater KI-Arbeitsbereich kann Entwicklern genehmigte technische Dokumentationen, Architekturwissen, Onboarding-Materialien, Codierungsstandards und interne Support-Inhalte zur Verfügung stellen. Der Zugriff kann nach Rolle, Team oder Projekt beschränkt werden.

Ein Entwickler, der an einem bestimmten Produkt arbeitet, benötigt nicht automatisch Zugriff auf alle Kundenumgebungen oder sämtliche Architekturdokumente. Passwörter, API-Keys und Produktions-Zugangsdaten sollten in dafür vorgesehenen Passwortmanagern oder Secret-Management-Systemen verbleiben.

Der KI-Arbeitsbereich unterstützt die technische Arbeit, ersetzt jedoch nicht die sichere Verwaltung von Zugangsdaten.



VERBAND ODER NON-PROFIT-ORGANISATION

Ein Verband verwaltet Mitgliederkarteien, Lieferantenrechnungen, Vorstandsdokumente, Teilnehmerlisten von Veranstaltungen und administrative Dateien. Mitarbeitende nutzen öffentliche KI-Tools, um Protokolle zusammenzufassen, die Mitgliederkommunikation zu übersetzen oder Vorstandsdokumente zu überarbeiten.

Der unkontrollierte Abfluss von Mitgliederdaten oder vertraulichen Vorstandsentscheiden in öffentliche Systeme.

Ein privater KI-Arbeitsbereich unterteilt Informationen in die Bereiche allgemeine Administration, Mitgliederkommunikation, Finanzen, Vorstand und Events. Vorstandsdokumente bleiben den Vorstandsmitgliedern vorbehalten. Finanzdaten bleiben auf den Finanzbereich und das autorisierte Management beschränkt. Event-Dokumente sind nur für die Personen zugänglich, die die Veranstaltung organisieren. Dies hilft kleinen Teams, von KI zu profitieren, ohne alle Dokumente in einer einzigen, offenen Umgebung zu vermischen.



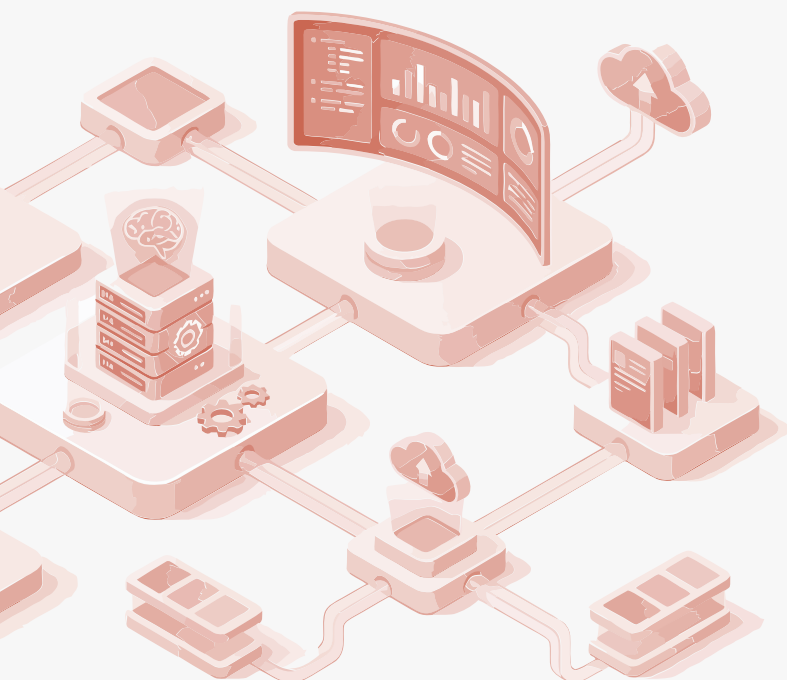
ZULIEFERER DES ÖFFENTLICHEN SEKTORS ODER BEHÖRDENNAHE ORGANISATION

Schweizer Organisationen, die für Behörden tätig sind, verwalten häufig Verträge, Projektdokumentationen, bürgerbezogene Informationen, technische Dokumente oder Betriebsdaten.

Diese Informationen sind zwar nicht immer offiziell klassifiziert, aber dennoch sensibel, vertraglich geschützt, reputationskritisch oder sicherheitsrelevant.

Ein privater KI-Arbeitsbereich kann nach Behörde, Kunde, Projekt, Vertrag oder Sensibilitätsstufe strukturiert werden.

Dies verringert das Risiko, dass behördliche Daten mit allgemeinen Unternehmensinhalten vermischt oder Nutzenden zugänglich gemacht werden, die sie nicht benötigen. Für behördliche oder hochsensible Arbeiten kann der Arbeitsbereich auf den eigenen Servern des Kunden oder einer privaten Infrastruktur installiert werden.



HR-, RECRUITMENT- ODER GESUNDHEITSNAHE ORGANISATION

Teams, die mit gesundheitsbezogenen Daten, HR- oder Rekrutierungsinformationen arbeiten, nutzen KI, um Lebensläufe zusammenzufassen, die Mitarbeiterkommunikation zu verfassen, Berichte zu erstellen oder Interviewnotizen zu strukturieren.

Diese Workflows beinhalten Personendaten, sensible Mitarbeiterinformationen oder vertrauliche Beurteilungen.

Ein privater KI-Arbeitsbereich kann allgemeines HR-Wissen, Rekrutierungsmaterialien, Mitarbeiterdokumentationen, Managementinhalte und Onboarding-Richtlinien voneinander trennen. Recruiter greifen auf kandidatenbezogene Materialien zu. Das HR greift auf Richtlinien- und Prozessdokumente zu. Das Management erhält Zugriff auf vertrauliche Entscheidungsdokumente. Allgemeine Mitarbeitende können nur auf genehmigte HR-Richtlinien, Onboarding-Leitfäden oder interne FAQs zugreifen. Das Ziel ist nicht die Automatisierung sensibler menschlicher Entscheidungen, sondern die Unterstützung sicherer Wissensarbeit, während personenbezogene und interne Informationen aus unkontrollierten KI-Tools ferngehalten werden.

7. WARUM „KI EINFACH VERBIETEN“ NICHT AUSREICHT

Einige Organisationen versuchen das Risiko zu minimieren, indem sie öffentliche KI-Tools komplett sperren. Dies mag in manchen Fällen kurzfristig notwendig sein, ist aber als langfristige Strategie meist nicht ausreichend.

Mitarbeitende nutzen KI, weil sie ihnen hilft, schneller zu arbeiten, Zeit zu sparen, Routineaufgaben zu reduzieren, die Textqualität zu verbessern, komplexe Informationen zu verstehen und wettbewerbsfähig zu bleiben.

Wenn das Unternehmen keine sichere und nützliche Alternative anbietet, nutzen die Mitarbeitenden KI oft heimlich über private Konten, Browser-Tools, mobile Apps oder unregulierte Plattformen weiter.



Ein vollständiges Verbot kann somit eine falsche Sicherheit suggerieren: Das Management glaubt, die KI-Nutzung sei unter Kontrolle, während die Belegschaft externe Tools im Hintergrund einfach weinternutzt.

Der bessere Ansatz ist die sichere KI-Einführung (Secure AI Adoption).

Das bedeutet, den Mitarbeitenden eine vom Unternehmen genehmigte KI-Umgebung mit klaren Regeln, kontrollierten Wissensquellen, definierten Zugriffsrechten und Cybersecurity-Support zur Verfügung zu stellen.

Das Ziel ist nicht, Menschen von der Nutzung von KI abzuhalten. Das Ziel ist es, ihnen einen sichereren Ort dafür zu bieten.

8. WAS UNTERNEHMEN STATTDESSEN BRAUCHEN

Das Problem mit öffentlichen KI-Tools ist einfach: Mitarbeitende kopieren Unternehmensinformationen oft in Systeme, die das Unternehmen nicht kontrolliert.

cogify unterstützt Unternehmen dabei, dies zu lösen, indem ein privater KI-Arbeitsbereich implementiert wird, bei dem sensible Unternehmensdaten die kontrollierte Umgebung des Unternehmens nicht verlassen müssen. Der Arbeitsbereich kann durch IVY, die private KI-Technologieschicht von cogify, betrieben werden. Die Lösung wird nicht als reiner Verkauf eines KI-Modells positioniert, sondern als sicherer, privater Arbeitsbereich, den die Mitarbeitenden tatsächlich in ihrer täglichen Arbeit nutzen können.

Ein privater KI-arbeitsbereich unterstützt:

- ✔ Die unternehmensweite Wissenssuche
- ✔ Die Zusammenfassung von Dokumenten
- ✔ Interne Frage-Antwort-Systeme
 - Das Entwerfen von E-Mails,
- ✔ Berichten, Angeboten und Präsentationen
- Besprechungsnotizen und die
- ✔ Erstellung von Geschäftsdokumenten
- ✔ Die Unterstützung bei technischen Dokumentationen
- ✔ Das Onboarding und den internen Wissensaustausch
- ✔ Kontrolliertes, KI-unterstütztes Arbeiten über Abteilungen hinweg

Der Arbeitsbereich wird um die reale Umgebung des Kunden herum aufgebaut:

- ✔ ihre Dokumente
- ✔ ihre Wissensquellen
- ✔ ihre Zugriffsrechte
- ✔ ihre Sicherheitsanforderungen
- ✔ ihre Infrastruktur
- ✔ ihre internen Regeln
- ✔ ihre Datenschutzbedürfnisse

Je nach anforderung kann der private KI-arbeitsbereich:

- ✔ Auf der Schweizer Infrastruktur von cogify gehostet werden
- ✔ Auf den eigenen Servern des Kunden installiert werden
- In einer gemeinsam mit dem Kunden
- ✔ definierten privaten Umgebung bereitgestellt werden

Das Ergebnis ist eindeutig: **Die Mitarbeitenden erhalten KI für ihre tägliche Arbeit.** Das Unternehmenswissen bleibt unter der Kontrolle des Unternehmens. Sensible Daten müssen die Unternehmensumgebung nicht verlassen. Das Management gewinnt eine sicherere und besser steuerbare Alternative zu öffentlichen KI-Tools.

9. WIE DATEN INNERHALB DES UNTERNEHMENS KONTROLLIERT WERDEN

Ein privater KI-Arbeitsbereich ist nur dann nützlich, wenn das Unternehmen kontrollieren kann, welche Informationen verfügbar sind, wer darauf zugreifen darf und wo die Daten verarbeitet werden. Deshalb muss der Arbeitsbereich von Anfang an mit klaren Zugriffseinschränkungen konzipiert werden.



KONTROLLIERTES HOSTING ODER KUNDENSEITIGE BEREITSTELLUNG

Der private KI-Arbeitsbereich kann auf der Schweizer Infrastruktur von cogify oder auf dem eigenen Server bzw. in der privaten Umgebung des Kunden betrieben werden. Das bedeutet, dass Unternehmensdokumente, Prompts, Wissensquellen und KI-gestützte Workflows nicht an öffentliche KI-Tools gesendet werden müssen.

Das Bereitstellungsmodell wird basierend auf den Anforderungen des Kunden ausgewählt:

- Von cogify betriebenes Schweizer Hosting
- Installation auf den eigenen Servern des Kunden
- Bereitstellung in einer privaten Umgebung
- Eingeschränktes Setup für sensible Sektoren oder behördliche Arbeiten

Das Ziel ist es, Unternehmensdaten in einer kontrollierten Umgebung zu halten, anstatt sie über unkontrollierte öffentliche KI-Konten zu verstreuen.



SEPARATE WORKSPACES

Ein privater KI-Arbeitsbereich sollte kein offener Sammelplatz für alle Unternehmensdaten sein. Er sollte in separate Workspaces unterteilt werden, die auf der Organisation des Unternehmens, den Sensibilitätsstufen und den realen Arbeitsprozessen basieren.

Beispiele hierfür sind:

- | | |
|--------------------------------------|--|
| • Allgemeiner Unternehmens-Workspace | • Entwicklungs-Workspace |
| • HR-Workspace | • Interner IT-Workspace |
| • Finanz-Workspace | • Management-Workspace |
| • Vertriebs-Workspace | • Vorstands-Workspace |
| • Kundenprojekt-Workspace | • Workspace für den öffentlichen Sektor / Behörden |

Jeder Workspace kann über eigene genehmigte Dokumente, Nutzende und Zugriffsregeln verfügen. Dies verhindert, dass sensible Informationen mit allgemeinem Wissen vermischt werden.

ROLLENBASIERTE ZUGRIFFSRECHTE

Der Zugriff auf den jeweiligen Workspace sollte auf der Rolle des Nutzers und der geschäftlichen Notwendigkeit basieren.

Zum Beispiel:

- Alle Mitarbeitenden dürfen auf Unternehmensrichtlinien und genehmigte interne FAQs zugreifen.
- Das HR darf auf HR-Prozesse, Mitarbeiterunterlagen und Rekrutierungsmaterialien zugreifen.
- Die Finanzabteilung darf auf Finanzprozesse und Rechnungsdokumente zugreifen.
- Entwickler dürfen auf technische Dokumentationen und Codierungsstandards zugreifen.
- Projektteams dürfen nur auf ihren eigenen Kundenprojekt-Workspace zugreifen.
- Das Management darf auf Strategien, Vorstandsunterlagen und vertrauliche Entscheidungsdokumente zugreifen.
- Vorstandsmitglieder dürfen nur auf vom Vorstand genehmigte Materialien zugreifen

Das Prinzip ist einfach: Mitarbeitende sollten nur das sehen, was sie sehen dürfen. KI darf nicht zu einer Abkürzung werden, um bestehende Zugriffsrechte zu umgehen.

GENEHMIGTE WISSENSQUELLEN

Der private KI-Arbeitsbereich sollte ausschliesslich verifizierte Wissensquellen des Unternehmens nutzen.

Bevor Dokumente angebunden werden, sollte das Unternehmen entscheiden, welche Dokumente:

- sicher zu verwenden sind
- zugriffsbeschränkt sind
- eine Überprüfung erfordern
- ausgeschlossen werden müssen
- anonymisiert oder bereinigt werden müssen
- Personendaten enthalten
- technische Geheimnisse enthalten
- zu einem bestimmten Kunden oder Projekt gehören

Dies stellt sicher, dass der KI-Arbeitsbereich stets mit den internen Regeln des Unternehmens übereinstimmt.

BERECHTIGUNGSKONFORME ANTWORTEN (PERMISSION-AWARE ANSWERS)

Der Arbeitsbereich muss die Zugriffsrechte bei der Generierung von Antworten strikt berücksichtigen. Ein Nutzer darf nur Antworten erhalten, die auf Dokumenten und Wissensquellen basieren, für die er eine ausdrückliche Zugriffsberechtigung besitzt.

Zum Beispiel:

- Das Marketing sollte keine Antworten erhalten, die auf HR-Gehaltsdokumenten basieren.
- Entwickler sollten keine Antworten erhalten, die auf strategischen Vorstandsdokumenten basieren.
- Ein Projektteam sollte keine Antworten erhalten, die auf vertraulichen Dateien eines anderen Kunden basieren.
- Allgemeine Mitarbeitende sollten keine Antworten erhalten, die auf Finanz- oder Rechtsdokumenten basieren.

Dies ist einer der wichtigsten Unterschiede zwischen einem kontrollierten privaten KI-Arbeitsbereich und der unkontrollierten Nutzung öffentlicher KI. Öffentliche KI-Tools kennen die Zugriffsrechte des Unternehmens nicht. Ein privater KI-Arbeitsbereich kann exakt um diese herum strukturiert werden.

TECHNISCHE GEHEIMNISSE UND ZUGANGSDATEN

Technische Geheimnisse erfordern eine besondere Handhabung. Ein privater KI-Arbeitsbereich darf kein Ort sein, an dem Passwörter, API-Keys, Produktions-Zugangsdaten oder Access-Tokens unkontrolliert gespeichert werden. Diese müssen in dafür vorgesehenen Passwortmanagern oder Secret-Management-Systemen verbleiben.

Für Entwicklungs- und IT-Teams sollte das Unternehmen klare Regeln definieren:

WAS MAN NICHT TUN SOLLTE (DON'T):

-  Passwörter in KI-Prompts einfügen
-  API-Keys in KI-Prompts einfügen
-  Produktions-Zugangsdaten in KI-Prompts einfügen
-  Ungeprüfte Logs hochladen, die Passwörter enthalten könnten
-  Umgebungsvariablen-Dateien (Env-Files) ohne Prüfung hochladen

WAS MAN TUN SOLLTE (DO):

-  Technische Dokumentationen bereinigen, bevor sie mit dem Workspace verknüpft werden
-  Geheimnisse in dafür vorgesehenen Secret-Management-Tools aufbewahren

Der KI-Arbeitsbereich kann die technische Arbeit unterstützen, ersetzt jedoch zu keinem Zeitpunkt ein sicheres Credential-Management.

10. WIE COGIFY UNTERSTÜTZT

cogify unterstützt Unternehmen bei der sicheren Einführung von KI durch die Kombination aus der Implementierung eines privaten KI-Arbeitsbereichs, Cybersecurity-Beratung, Schweizer Hosting und langfristigem technischem Support.

IMPLEMENTIERUNG DES PRIVATEN KI-ARBEITSBEREICHS

cogify konzipiert und implementiert einen privaten KI-Arbeitsbereich basierend auf den Geschäftsanforderungen, der Datensensibilität, der Infrastruktur und den Sicherheitsvorgaben des Kunden.

Der Arbeitsbereich kann durch IVY, die private KI-Technologieschicht von cogify, betrieben werden.

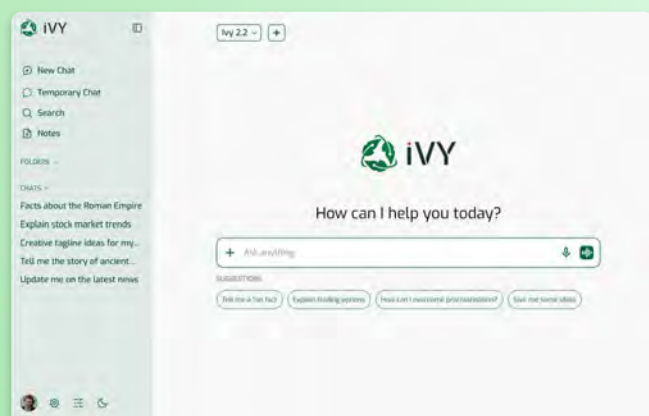
Dies umfasst das Setup, die Konfiguration, die Anbindung von Wissensquellen, die Strukturierung der Workspaces, die Logik der Zugriffsrechte und die Bereitstellungsplanung.

SICHERES HOSTING ODER KUNDENSEITIGE BEREITSTELLUNG

Der private KI-Arbeitsbereich kann auf der Schweizer Infrastruktur von cogify gehostet oder direkt auf dem eigenen Server bzw. in der privaten Umgebung des Kunden installiert werden. Das passende Bereitstellungsmodell wird basierend auf den Datenschutz-, Cybersecurity-, Compliance- und Betriebsanforderungen des Kunden ausgewählt.

KONZEPTION VON WORKSPACES UND ZUGRIFFSRECHTEN

cogify unterstützt Kunden bei der Definition, wie Daten innerhalb des Arbeitsbereichs getrennt werden sollten. Dies umfasst Abteilungsworkspaces, Projektworkspaces, kundenspezifische Workspaces, Managementworkspaces und geschützte Workspaces für sensible Inhalte. Das Ziel ist es, sicherzustellen, dass Nutzende nur auf die Informationen zugreifen, die sie auch verwenden dürfen.



CYBERSECURITY-BERATUNG

cogify hilft Unternehmen, das Risiko von Shadow AI durch gezielte Cybersecurity-Beratung, Zugriffsüberprüfungen, die Erfassung sensibler Datenflüsse (Sensitive Data Mapping), Unterstützung bei Richtlinien, Leitfäden für Mitarbeitende und Sicherheitsüberprüfungen von KI-Workflows zu reduzieren.

Dies beinhaltet Regeln für vertrauliche Daten, Kundeninformationen, HR-Daten, Finanzdaten, Quellcode, Zugangsdaten, Logs und technische Geheimnisse.



WARTUNG UND KONTINUIERLICHE VERBESSERUNG

Eine sichere KI-Einführung endet nicht mit der Erstinstallation. cogify kann den laufenden Betrieb des privaten KI-Arbeitsbereichs durch folgende Massnahmen unterstützen:

- Technische Updates und Sicherheits-Updates
- Anpassungen von Zugriffsrechten
- Support für neue Nutzende oder Abteilungen
- Überprüfung der angebundenen Wissensquellen
- Verbesserung der Dokumentenstruktur
- Auswertung von User-Feedback
- Optimierung von KI-Workflows
- Periodische Cybersecurity-Überprüfungen

Da sich KI-Anwendungsfälle schnell weiterentwickeln, neue Abteilungen Zugriff anfordern, neue Dokumente erstellt werden, neue Risiken auftauchen und sich interne Richtlinien ändern, sorgt dieser kontinuierliche Support dafür, dass der Arbeitsbereich nützlich, sicher und an der geschäftlichen Realität ausgerichtet bleibt.



11. KI-SICHERHEITS-CHECKLISTE FÜR DAS MANAGEMENT

Unternehmen sollten sich die folgenden Fragen stellen:

- Wissen wir, welche KI-Tools die Mitarbeitenden nutzen?
- Nutzen Mitarbeitende private KI-Konten für die Arbeit?
- Wissen wir, welche Daten in KI-Prompts eingefügt werden?
- Haben wir klare Regeln für Kundendaten, HR-Daten, Finanzdaten und Quellcode?
- Haben wir Regeln für Benutzernamen, Passwörter, API-Keys, Tokens und Logs?
- Verfügen wir über einen offiziell genehmigten KI-Arbeitsbereich?
- Ist dieser Arbeitsbereich in geschützte und voneinander getrennte Bereiche unterteilt?
- Werden die Wissensquellen des Unternehmens durch Zugriffsrechte kontrolliert?
- Können Nutzende nur auf Workspaces zugreifen, die für ihre Rolle relevant sind?
- Wissen wir genau, wo unsere KI-bezogenen Daten gehostet werden?
- Schulung wir unsere Mitarbeitenden im sicheren Umgang mit KI?
- Überwachen und überprüfen wir KI-Risiken regelmässig?
- Werden KI-generierte Ergebnisse geprüft, bevor sie in sensiblen Workflows verwendet werden?
- Wissen wir, welche KI-Tools mit unseren internen Unternehmenssystemen verknüpft sind?
- Haben wir eine klare Verantwortlichkeit (Ownership) für die KI-Governance definiert?
- Haben wir Wartung und Support für unseren KI-Arbeitsbereich sichergestellt?

Wenn Sie mehrere dieser Fragen mit „Nein“ beantworten müssen, besteht in Ihrem Unternehmen höchstwahrscheinlich bereits ein Shadow-AI-Risiko.

12. EMPFOHLENE ERSTE SCHRITTE

Für Unternehmen, die besorgt über die Risiken von Shadow AI sind, empfiehlt cogify die folgenden ersten Schritte:

SCHRITT 1: AKTUELLE KI-NUTZUNG ERMITTELN

Verstehen Sie, welche KI-Tools die Mitarbeitenden bereits nutzen, welche Abteilungen sie einsetzen und zu welchem Zweck.

SCHRITT 2: SENSIBLE DATEN KLASSIFIZIEREN

Definieren Sie, welche Daten niemals in öffentliche oder unkontrollierte KI-Tools eingegeben werden dürfen. Dies sollte Kundendaten, Mitarbeiterdaten, Finanzdaten, Quellcode, Zugangsdaten, Benutzernamen, Passwörter, API-Keys, Access-Tokens, Logs und technische Konfigurationsdetails umfassen.

SCHRITT 3: ZUGRIFFSRECHTE ÜBERPRÜFEN

Prüfen Sie, wer auf welche internen Wissensquellen zugreifen können sollte.

SCHRITT 4: EINSCHRÄNKUNGEN FÜR DEN ARBEITSBEREICH KONZIPIEREN

Definieren Sie, welche Workspaces benötigt werden. Zum Beispiel :

- z. B. Allgemein
- Finanzen
- Kundenprojekte
- Vorstand
- HR
- Entwicklung
- Management

Legen Sie für jeden Workspace fest, wer darauf zugreifen darf, welche Dokumente enthalten sind und welche Daten strikt gesperrt bleiben müssen.

SCHRITT 5: DAS RICHTIGE BEREITSTELLUNGSMODELL WÄHLEN

Entscheiden Sie, ob der private KI-Arbeitsbereich auf der Schweizer Infrastruktur von cogify, auf den eigenen Servern des Kunden oder in einer anderen privaten Umgebung betrieben werden soll.

SCHRITT 6: IMPLEMENTIEREN, SCHULEN UND STEUERN

Richten Sie den Arbeitsbereich ein, binden Sie die genehmigten Wissensquellen an, konfigurieren Sie die Zugriffsrechte, schulen Sie die Nutzenden und erstellen Sie Governance-Regeln für die KI-Nutzung.

SCHRITT 7: KONTINUIERLICH WARTEN UND VERBESSERN

Überprüfen Sie die Nutzung, optimieren Sie die Wissensquellen, aktualisieren Sie die Zugriffsrechte, unterstützen Sie die Mitarbeitenden und passen Sie den Arbeitsbereich an, wenn sich die geschäftlichen und sicherheitstechnischen Anforderungen weiterentwickeln.

FAZIT

KI ist bereits am Arbeitsplatz angekommen. Die Frage ist nicht mehr, ob Mitarbeitende KI nutzen werden, sondern ob das Unternehmen ihnen eine sichere und kontrollierte Umgebung dafür bieten kann. Die unkontrollierte Nutzung von KI führt zu Cybersecurity-, Datenschutz- und Compliance-Risiken. KI allein zu verbieten, ist keine realistische langfristige Lösung. Unternehmen benötigen eine sichere KI-Einführung.

cogify unterstützt Unternehmen bei der Implementierung, dem Betrieb und der Wartung privater KI-Arbeitsbereiche, die die tägliche Arbeit effektiv unterstützen und gleichzeitig die Abhängigkeit von unkontrollierten öffentlichen KI-Tools verringern. Der Arbeitsbereich kann durch IVY, die private KI-Technologieschicht von cogify, betrieben werden. Die kundenorientierte Lösung ist einfach: ein privater KI-Arbeitsbereich, bei dem die Daten die kontrollierte Unternehmensumgebung nicht verlassen müssen.

Unternehmensinformationen können in geschützte Workspaces aufgeteilt werden. Der Zugriff kann nach Rolle, Abteilung, Projekt oder Sensibilitätsstufe definiert werden. HR-Dokumente bleiben beim HR. Finanzdokumente bleiben bei den Finanzen. Kundendaten verbleiben in kundenspezifischen Workspaces. Technische Dokumentationen bleiben bei autorisierten Entwicklern und der IT. Management- und Vorstandsmaterialien bleiben der Führungsebene vorbehalten. So können Unternehmen ihren Mitarbeitenden KI zur Verfügung stellen, ohne die Kontrolle über das Unternehmenswissen zu verlieren.

Kombiniert mit Cybersecurity-Beratung, Governance-Support und langfristiger Wartung bietet ein privater KI-Arbeitsbereich Organisationen einen praktischen Weg, um das Risiko von Shadow AI zu reduzieren, sensible Informationen zu schützen und ein starkes Fundament für eine verantwortungsvolle KI-Nutzung aufzubauen.

BEREIT, IHREN MITARBEITENDEN KI ZU GEBEN, OHNE DIE KONTROLLE ÜBER DIE UNTERNEHMENSDATEN ZU VERLIEREN?

Sprechen Sie mit cogify über die Implementierung eines privaten KI-Arbeitsbereichs, Wartung und Cybersecurity-Beratung. Wir helfen Ihnen zu verstehen, wie Ihre Mitarbeitenden KI heute nutzen, wo sensible Informationen abfließen könnten und wie ein privater KI-Arbeitsbereich Ihrem Unternehmen eine sicherere Alternative zu öffentlichen KI-Tools bieten kann.



Fordern Sie eine KI-Sicherheitsberatung an

WICHTIGER RECHTLICHER HINWEIS (DISCLAIMER)

Die Schweizer Anwendungsfälle in diesem Whitepaper sind praxisnahe Geschäftsszenarien, die auf typischen Workflows in Schweizer KMU, Treuhandunternehmen, Verbänden, Technologieunternehmen, Organisationen im Gesundheitswesen und behördennahen Umgebungen basieren. Sie werden nicht als bestätigte, durch KI verursachte Schweizer Hacking-Vorfälle dargestellt.

Die bestätigten Schweizer Referenzen dienen dazu, das allgemeine Cybersecurity-Umfeld in der Schweiz aufzuzeigen, insbesondere in Bezug auf Datenoffenlegung, Ransomware, Phishing, Lieferanten- und Governance-Risiken. Shadow AI sollte als eine neue Risikoschicht verstanden werden, die den Datenabfluss erhöhen kann, wenn Unternehmen keine sicheren, genehmigten KI-Workflows bereitstellen.

Das Bereitstellungsmodell für einen privaten KI-Arbeitsbereich sollte basierend auf den individuellen Sicherheits-, Datenschutz-, Compliance- und Betriebsanforderungen des jeweiligen Kunden definiert werden. Für einige Kunden kann eine von cogify betriebene Schweizer Infrastruktur geeignet sein. Für andere wird eine Installation auf den eigenen Servern des Kunden oder in einer privaten Umgebung bevorzugt.

REFERENZEN UND WEITERFÜHRENDE LITERATUR

1. NATIONALES ZENTRUM FÜR CYBERSICHERHEIT: HALBJAHRESBERICHT 2025/1

Das NCSC verzeichnete im ersten Halbjahr 2025 insgesamt 35'727 Cyberaktivitätsmeldungen und hob Ransomware, Datenerpressung, Phishing, DDoS-Angriffe, Software-Schwachstellen und Angriffe auf die Lieferkette als relevante Cyberbedrohungen für die Schweiz hervor.

Link: <https://www.ncsc.admin.ch/ncsc/de/home/dokumentation/berichte/lageberichte/halbjahresbericht-2025-1.html>

2. NATIONALES ZENTRUM FÜR CYBERSICHERHEIT: BERICHT ZUR DATENANALYSE DES HACKERANGRIFFS AUF XPLAIN

Der NCSC-Bericht erläutert die Analyse von Daten, die nach dem Hackerangriff auf Xplain, einen wichtigen Anbieter von IT-Dienstleistungen für Bundes- und Kantonsbehörden, im Darknet veröffentlicht wurden. Dies ist kein KI-Vorfall, aber ein relevantes Schweizer Beispiel dafür, warum die Offenlegung sensibler Daten, Lieferantenrisiken und Technologieumgebungen von Drittanbietern von zentraler Bedeutung sind.

Link: <https://www.ncsc.admin.ch/ncsc/de/home/dokumentation/berichte/fachberichte/bericht-datenanalyse-xplain.html>

3. OWASP TOP 10 FÜR LARGE LANGUAGE MODEL (LLM)-ANWENDUNGEN

Das GenAI Security Project von OWASP identifiziert und dokumentiert Sicherheits- und Zuverlässigkeitsrisiken im Zusammenhang mit generativen KI-Technologien, einschliesslich grosser Sprachmodelle (LLMs) und KI-gesteuerter Anwendungen. Zu den relevanten Risiken gehören Prompt Injection, die Offenlegung sensibler Informationen, unsichere Ausgabebehandlung und übermässige Handlungsautonomie (Excessive Agency).

Link: <https://owasp.org/www-project-top-10-for-large-language-model-applications/>

4. CISA-FALL: HOCHLADEN VON DOKUMENTEN IN DAS ÖFFENTLICHE CHATGPT

ITPro berichtete, dass die kommissarische Leitung der US-Cybersecurity and Infrastructure Security Agency (CISA) sensible Dokumente der Einstufung „For Official Use Only“ in eine öffentliche Version von ChatGPT hochgeladen hat, was erhebliche Bedenken hinsichtlich Shadow AI und Datenkontrolle aufwarf.

Link: <https://www.itpro.com/security/data-protection/cisas-interim-chief-uploaded-sensitive-documents-to-a-public-version-of-chatgpt-security-experts-explain-why-you-should-never-do-that>

5. SCHWEIZER BUNDESRAT: AUSRICHTUNG DER SCHWEIZER KI-REGULIERUNG

Der Bundesrat gab bekannt, dass die Schweiz beabsichtigt, die KI-Konvention des Europarats zu ratifizieren und diese über einen sektorspezifischen Ansatz umzusetzen, bei Bedarf unterstützt durch nicht rechtlich bindende Massnahmen.

Link: <https://www.admin.ch/de/nsb?id=104110>

KI IST NÜTZLICH. UNKONTROLLIERTE KI IST RISKANT.

Cogify hilft Unternehmen beim Übergang von Schatten-KI zu einer sicheren KI-Adoption durch die Bereitstellung privater KI-Arbeitsbereiche, Cybersicherheitsberatung sowie langfristige Wartung und technischen Support.

Die Mitarbeitenden können KI für ihre tägliche Arbeit nutzen.

Das Wissen des Unternehmens bleibt in einer kontrollierten Umgebung.

Die Daten sind dank eingeschränkter Arbeitsbereiche voneinander isoliert.

Die Geschäftsleitung behält eine verstärkte Kontrolle über Zugriffsrechte, Datenspeicherung und Cybersicherheitsrisiken.