

Bereiten Sie Ihren Agenten auf die Prüfung vor.

Eine praktische Nachweis-Checkliste für das Gesundheits- und Finanzwesen. Bestimmen Sie für jede Kontrolle die verantwortliche Person, vorhandene Nachweise und offene Lücken. Erfassen Sie nur Informationen, die Sie weitergeben dürfen.

Agent, Version und Einsatzumfang

Eine ausgefüllte Checkliste ist kein Prüfentscheid und keine Versicherungszusage. Eine unabhängige Prüfung und ein definierter Umfang bleiben erforderlich.

Governance und Umfang

Verantwortung, vorgesehene Nutzung, Ausschlüsse und versioniertes Inventar des geprüften Agenten.

G01 · 0.1

☐ Benannte Verantwortung

Verantwortliche Person, technischen Betrieb und unabhängige Entscheidungsprüfung benennen; wer Kontrollen umsetzt, darf die eigene Arbeit nicht zertifizieren.

Vorzubereitende Nachweise

- Unterzeichnete Rollenmatrix mit Entscheidungsrechten.
- Interessenerklärungen und Kompetenznachweis der prüfenden Person.

Verantwortung / Nachweisreferenz

G02 · 0.1

☐ Vorgesehene Nutzung und Ausschlüsse

Nutzende, Entscheidungen, Kanäle, Daten, Aktionen und Ausschlüsse so beschreiben, dass der Betrieb sie durchsetzen kann.

Vorzubereitende Nachweise

- Freigegebene Einsatzbeschreibung und User Journey.
- Sichtbare Hinweise, Routingregeln und Sperrkonfiguration.

Verantwortung / Nachweisreferenz

G03 · 0.1

☐ Versioniertes Systeminventar

Modell, Prompt, Wissensquellen, Werkzeuge, Regeln und Deployment mit unveränderlichen Versionen an die Prüfung binden.

Vorzubereitende Nachweise

- Signiertes Inventar und Abhängigkeitsliste.
- Hashes oder Release-IDs der geprüften Artefakte.

Verantwortung / Nachweisreferenz

Daten und Vertraulichkeit

Datenflüsse, Minimierung, Zugriff, Aufbewahrung und Schutz vor unbeabsichtigter Offenlegung.

D01 · 0.1

☐ Erfasste und minimierte Datenflüsse

Jede Eingabe, Ableitung, Destination und jeden Zweck erfassen; nicht benötigte Felder entfernen.

Vorzubereitende Nachweise

- Datenflussdiagramm mit deklarierten Standorten.
- Feldweise Zweck- und Minimierungsprüfung.

Verantwortung / Nachweisreferenz

D02 · 0.1

☐ Zugriff und Aufbewahrung

Rollenbasierte Zugriffe, Mandantentrennung sowie dokumentierte Aufbewahrung und Löschung für Prompts, Spuren, Caches und Backups durchsetzen.

Vorzubereitende Nachweise

- Zugriffsmatrix und Konzept der Mandantengrenzen.
- Aufbewahrungsplan mit Löschtests.

Verantwortung / Nachweisreferenz

D03 · 0.1

☐ Schutz vor Offenlegung

Ausgabefilter, Retrieval-Autorisierung und Kontexttrennung so umsetzen, dass nur berechtigte Daten ausgegeben werden.

Vorzubereitende Nachweise

- Autorisierungsregeln für Retrieval.
- Red-Team-Fälle für direkte und indirekte Offenlegung.

Verantwortung / Nachweisreferenz

Integrationssicherheit

Nicht vertrauenswürdige Eingaben, Werkzeugrechte, Geheimnisse, Abhängigkeiten und Mandantengrenzen.

S01 · 0.1

☐ Umgang mit nicht vertrauenswürdigen Eingaben

Externe Inhalte als Daten behandeln, Anweisungshierarchie bewahren und Selbstautorisierung blockieren.

Vorzubereitende Nachweise

- Modell der Vertrauensgrenzen und Prompt-Regeln.
- Injection-Testkorpus mit Sollresultaten.

Verantwortung / Nachweisreferenz

S02 · 0.1

☐ Werkzeuge mit minimalen Rechten

Werkzeuge auf minimale Scopes, Datensätze, Aktionen und Dauer begrenzen; Mandanten- und Umgebungszugänge trennen.

Vorzubereitende Nachweise

- Werkzeugkatalog mit Scopes und Verantwortlichen.
- Credential- und Mandamentrennungs-Konfiguration.

Verantwortung / Nachweisreferenz

S03 · 0.1

☐ Geheimnisse, Abhängigkeiten und Ausführungsgrenzen

Geheimnisse aus Prompts und Code fernhalten, Abhängigkeiten prüfen und Befehle oder Code mit Ressourcenlimiten isolieren.

Vorzubereitende Nachweise

- Nachweise zu Secret-Management und Rotation.
- Abhängigkeitsinventar und Sandbox-Regeln.

Verantwortung / Nachweisreferenz

Zuverlässigkeit

Repräsentative Fälle, unbelegte Ausgaben und Regressionstests für die deklarierte Nutzung.

R01 · 0.1

☐ Repräsentative Leistungsfälle

Versionierten Testsatz für Normal-, Grenz- und Hochrisikofälle des Schweizer Einsatzes samt unterstützten Sprachen definieren.

Vorzubereitende Nachweise

- Begründung, Grösse und Zusammensetzung des Testsatzes.
- Sollresultat und geprüftes Ergebnis je Fall.

Verantwortung / Nachweisreferenz

R02 · 0.1

☐ Umgang mit unbelegten Ausgaben

Bei faktenabhängigen Aufgaben nachvollziehbare Quellen verlangen und schwache oder widersprüchliche Antworten ablehnen oder prüfen lassen.

Vorzubereitende Nachweise

- Grounding-Regeln und Kriterien für Quellenqualität.
- Fälle mit fehlenden, veralteten und widersprüchlichen Belegen.

Verantwortung / Nachweisreferenz

R03 · 0.1

☐ Regressionskontrolle

Risikobasierte Regressionstests vor Freigabe ausführen und Änderungen für Teil- oder Vollprüfung definieren.

Vorzubereitende Nachweise

- Release-Gate und Regressionshistorie.
- Matrix von Änderungen und betroffenen Tests.

Verantwortung / Nachweisreferenz

Menschliche Aufsicht

Freigaben, Eskalation und Stoppmechanismen für folgenreiche Aktionen.

H01 · 0.1

☐ Freigabe folgenreicher Aktionen

Vor Buchungen, externem Versand, Löschungen und klinischem Fallfortschritt eine namentliche Freigabe an genaue Aktion, Ziel und aktuellen Inhalt binden.

Vorzubereitende Nachweise

- Aktionsinventar und Freigaberegeln.
- Anonymisierte Spuren für erlaubte, abgelehnte und abgelaufene Anfragen.

Verantwortung / Nachweisreferenz

H02 · 0.1

☐ Eskalation an eine fachkundige Person

Beobachtbare Auslöser, Ziel, Kontextpaket und Verantwortung für die Antwort festlegen.

Vorzubereitende Nachweise

- Eskalationsmatrix und Serviceverantwortung.
- Beispiele mit Kontext und Lösung.

Verantwortung / Nachweisreferenz

H03 · 0.1

☐ Stopp und Eindämmung

Getestete, agentenunabhängige Mechanismen zum Sperren von Aktionen, Entziehen von Zugängen, Isolieren eines Mandanten und Sichern von Belegen bereitstellen.

Vorzubereitende Nachweise

- Runbook für Notstopp und Eindämmung.
- Übungsnachweis mit Rollen und Ablaufschritten.

Verantwortung / Nachweisreferenz

Transparenz und Auswirkungen

KI-Hinweis, verständliche Entscheidungen und Folgen für betroffene Personen.

T01 · 0.1

☐ Hinweis auf KI-Interaktion

Zum relevanten Zeitpunkt KI, Zweck, wichtigste Datennutzungen und Zugang zu einer Person offenlegen.

Vorzubereitende Nachweise

- Oberflächentexte für jeden unterstützten Kanal.
- Prüfung von Zeitpunkt und Barrierefreiheit.

Verantwortung / Nachweisreferenz

T02 · 0.1

☐ Verständliche Ergebnisse

Wesentliche Grundlage, Unsicherheit, nächsten Schritt und Prüfweg zielgruppengerecht darstellen.

Vorzubereitende Nachweise

- Ergebnisvorlagen und Erklärstandard.
- Geprüfte Beispiele für verschiedene Zielgruppen und Sprachen.

Verantwortung / Nachweisreferenz

T03 · 0.1

☐ Folgenabschätzung und betroffene Gruppen

Betroffene Gruppen, plausible Schäden, Schwere, Wahrscheinlichkeit und Minderungen vor risikoreichem Einsatz bestimmen und bei Änderungen neu prüfen.

Vorzubereitende Nachweise

- Folgenabschätzung mit Konsultationsbeiträgen.
- Risikoverantwortliche, Minderungen und Akzeptanzentscheide.

Verantwortung / Nachweisreferenz

Betrieb

Manipulationsnachweisbare Spuren, Vorfallbehandlung und kontrollierte Änderungen nach Freigabe.

O01 · 0.1

☐ Manipulationsnachweisbare Ereignisspuren

Identität, Version, Eingabereferenzen, Regeln, Tool-Aufrufe, Freigaben und Resultate unveränderlich oder mit Änderungsnachweis protokollieren.

Vorzubereitende Nachweise

- Ereignisschema, Integritätskonzept und Zugriffsregeln.
- Durchgehende Beispielspur über Agent, Freigabe und Tool.

Verantwortung / Nachweisreferenz

O02 · 0.1

☐ Vorfallreaktion bei Agentenfehlern

Agentenspezifische Schweregrade, Eindämmung, Beweissicherung, Meldeprüfung und Lernverantwortung festlegen.

Vorzubereitende Nachweise

- Vorfall-Playbook und Kontaktbaum.
- Übungs- oder Realfall mit Folgemassnahmen.

Verantwortung / Nachweisreferenz

O03 · 0.1

☐ Änderung und Quartalsprüfung

Nach diesem Schemaentwurf Belege mindestens quartalsweise sowie sofort nach wesentlichen Änderungen oder Vorfällen prüfen.

Vorzubereitende Nachweise

- Prüfkalender und Kriterien für wesentliche Änderungen.
- Quartalsprotokolle, Release-Historie und offene Ausnahmen.

Verantwortung / Nachweisreferenz

Nachweise

Herkunft, Reproduzierbarkeit, Prüfgrenzen und Restrisiko.

E01 · 0.1

☐ Herkunft und Integrität der Nachweise

Artefakte mit Quelle, erhebender Person, Zeit, Umgebung, Version und Integritätsreferenz kennzeichnen und unnötige Daten schwärzen.

Vorzubereitende Nachweise

- Nachweisregister mit Übernahmekette.
- Originalartefakte oder Hashes in zugriffsgeschütztem Speicher.

Verantwortung / Nachweisreferenz

E02 · 0.1

☐ Reproduzierbare Verifikation

Voraussetzungen, Eingaben, Ablauf, Soll, Ist und prüfende Person so dokumentieren, dass eine qualifizierte zweite Person wiederholen kann.

Vorzubereitende Nachweise

- Ausführbare Tests oder genaue Protokolle.
- Umgebungsmanifest und aufbewahrte Resultate.

Verantwortung / Nachweisreferenz

E03 · 0.1

☐ Umfang und Restrisiko

Version, Umgebung, Einsatz, Ausschlüsse, fehlgeschlagene oder ungetestete Bereiche, Restrisiken und Prüfauslöser beim Ergebnis veröffentlichen.

Vorzubereitende Nachweise

- Signierter Abschnitt zu Umfang und Restrisiko.
- Konsistenzprüfung zwischen Bericht, Register und öffentlichen Aussagen.

Verantwortung / Nachweisreferenz
