

ANHANG A

12-Punkte-Audit der menschlichen Schicht

Ein Werkzeug für Werksleiter, CTOs und Compliance-Verantwortliche. Zwölf Fragen, die jedes produktive KI-System einer DACH-Industrieorganisation beantworten können sollte — mit klaren Kriterien für bestanden, kritisch, nicht vorhanden.

Dieses Audit ersetzt keinen technischen Penetrationstest und keine ISO-Zertifizierung. Es prüft die *menschliche Schicht* — die 30 % eines KI-Systems, in denen die Verantwortung sitzt. In den Projekten, die ich gesehen habe, schlagen Vorfälle fast nie an der technischen Schicht ein. Sie schlagen dort ein, wo eine Rolle vakant, ein Mandat unklar oder eine Dokumentation veraltet war. Die zwölf Fragen unten sind aus genau diesen Vorfällen rückwärts konstruiert.

Die Fragen sind in drei Blöcken organisiert, entsprechend den drei Rollen aus Kapitel 1: **Wer trainiert, Wer kontrolliert, Wer haftet**. Pro Frage: *Warum es zählt, Wie prüfen, Was tun, wenn die Antwort Nein ist*. Am Ende eine Auswertungstabelle.

Realistische Dauer für einen ersten Durchgang: 60 bis 90 Minuten pro KI-System mit den richtigen Personen am Tisch (Werksleiter, KI-Operator, IT-Verantwortlicher, Compliance Officer).

. . .

A · TRAINING — WER PRÄGT DIE WIRKLICHKEIT DES MODELLS

A1 Daten-Kurator: Ist die Person namentlich benannt?**WARUM ES ZÄHLT**

Trainingsdaten formen das, was das Modell für Realität halten wird. Ohne benannte Person mit dem expliziten Mandat *diese Datenbasis akzeptieren oder ablehnen* sammelt Data Engineering automatisch — und Anomalien wie ein Streiktag im Datensatz werden zu systematischen Fehlern (siehe Kapitel 6, Fall B).

WIE PRÜFEN

Ein Dokument mit Namen, Position, Mandatsbeschreibung. Nicht *Data Science Team*, sondern eine Person mit Unterschrift und Vertretungsregelung.

BEI NEIN

Innerhalb von 14 Tagen: eine Person zuweisen, Mandat schriftlich definieren (eine A4-Seite reicht), in die Betriebsdokumentation aufnehmen.

A2 Datenherkunft und Lineage: Können Sie für jede Trainingsquelle Provenienz und Stand belegen?**WARUM ES ZÄHLT**

Ein Datensatz aus 2022 fließt ins Modell ein, obwohl die Linie 2025 umgebaut wurde. Niemand widerspricht, weil niemand das Mandat dafür hat. Lineage ohne Eigentümer ist keine Lineage.

WIE PRÜFEN

Eine Quellenübersicht pro Modell: System (z. B. SAP, MES, Lieferantenexport), Zeitraum, letzte Aktualisierung, verantwortliche Person, Hash oder Versionsnummer.

BEI NEIN

Datenkarte je Modell anlegen — Vorlage in Kapitel 7, Schritt 7. Ohne Datenkarte keine Re-Trainings-Freigabe.

A3 Anomalie-Filterung: Wer prüft den Trainingsdatensatz auf Ereignisse, die das Modell *nicht* als normal lernen soll?**WARUM ES ZÄHLT**

Streiktage, Werksumbauten, Lieferantenwechsel, Saisonsondereffekte sind technisch gültige Datenpunkte — und semantisch toxisch für das Lernziel. Diese Filterung ist eine redaktionelle Aufgabe. Sie passiert nicht automatisch.

WIE PRÜFEN

Eine Liste der bekannten *do-not-train*-Ereignisse pro Datensatz, mit Begründung und freigebender Person. Eine Diff-Prüfung vor jedem Re-Training.

BEI NEIN

Vor dem nächsten Re-Training: einen 60-Minuten-Workshop mit Operator, Daten-Kurator und Domänenexperte. Output: erste Liste der Ausschluss-Ereignisse.

A4 Re-Training-Trigger: Existieren schriftliche Kriterien, wann re-trainiert wird — und wer es freigibt?

WARUM ES ZÄHLT

Re-Training nach Bedarf bedeutet in der Betriebsrealität meist *nie* — bis das System merklich erodiert. Ein Modell, das auf der Anfangsbeschleunigung 24 Monate weiterläuft, fährt zwangsläufig in einen Drift, den niemand mehr datiert.

WIE PRÜFEN

Definierte Trigger (zeitlich, performance-basiert, ereignisbezogen) plus eine Freigabeprozedur. Beispiele: Quartals-Review, Drift-Schwelle 4 Prozent über zwei Wochen, signifikante Prozessänderung.

BEI NEIN

Drei Trigger definieren — einen zeitlichen, einen metrischen, einen ereignisbezogenen — und in den nächsten Quartalstermin der Governance aufnehmen.

. . .

B · KONTROLLE — WER EINGREIFT, WENN DAS SYSTEM ESKALIERT

B1 Operator-Rolle: Ist die KI-Operator-Position namentlich besetzt — mit Vertretung?

WARUM ES ZÄHLT

Die Rolle, in keiner HR-Standardkategorie passgenau abgebildet, ist die einzige Stelle mit operativem Zugriff auf das laufende System. Vakant heißt: Eskalationen landen bei jemandem ohne Mandat oder ohne Wissen.

WIE PRÜFEN

Stellenbeschreibung mit den vier Funktionen aus Kapitel 5 (Prompt-Architektur, Daten-Kuration, Eskalationsvermittlung, Compliance-Übersetzung). Hauptperson + Vertretung. Vergütung im Korridor 95 000–125 000 EUR brutto für die Senior-Variante (siehe Kapitel 9).

BEI NEIN

Akut: jemanden mit dem Mandat *vorläufig* belegen. Mittelfristig: Stellenbeschreibung schreiben (Übung in Kapitel 9). Wenn das schwer fällt, fehlt die Klarheit über die Rolle.

B2 Eskalationsschwellen: Sind die Schwellen numerisch *und* verbal dokumentiert?**WARUM ES ZÄHLT**

Eskaliere bei Konfidenz unter 85 % klingt präzise. In der Nachtschicht, wo der Schichtmeister diesen Wert zum ersten Mal sieht, bedeutet er nichts (siehe Kapitel 6, Fall A). Numerische Schwellen brauchen eine Übersetzung in *was bedeutet das praktisch*.

WIE PRÜFEN

Pro Schwelle: Wert, verbale Erklärung (typischerweise tritt das auf bei ... prüfen Sie zuerst ...), erwartete Reaktionszeit, Pfad. Verfügbar dort, wo der Operator sie braucht — nicht in einem ungelesenen SharePoint-Ordner.

BEI NEIN

Drei häufigste Eskalationen identifizieren, mit dem Operator zusammen die verbale Übersetzung schreiben, in die Systemoberfläche einbauen.

B3 Override-Logging: Wird jede menschliche Aufhebung einer Systementscheidung mit Person, Zeitstempel und Begründung protokolliert?**WARUM ES ZÄHLT**

Override ohne Log ist Schatten-Operation. Der EU AI Act für Hochrisiko-Systeme verlangt automatisches Logging und mindestens fünf Jahre Aufbewahrung. Override-Muster sind außerdem die wertvollste Quelle für Modell-Verbesserung.

WIE PRÜFEN

Stichprobe: Können Sie für die letzten 30 Tage alle Overrides auflisten — wer, wann, warum? Wer prüft diese Logs in welcher Frequenz?

BEI NEIN

Sofort: Override-Endpoint instrumentieren (Pflichtfelder im UI). Mittelfristig: monatliches Override-Review als Bestandteil der Governance.

B4 24/7-Mandat: Existiert für jede Eskalationsstufe ein Entscheidungsträger außerhalb der Bürozeiten?**WARUM ES ZÄHLT**

Critical-Alerts treten typischerweise nicht zwischen 9 und 17 Uhr ein. Eine reparierbare Situation um 21:30 wird zu einem 78 000-Euro-Schaden, weil niemand explizit befugt ist, den Stopp zu autorisieren (Kapitel 6, Fall A).

WIE PRÜFEN

Eskalationsplan mit Tag- und Nacht-/Wochenend-Spalten. Konkrete Personen, Telefonnummern, Stellvertretung. Letzter Eskalations-Drill außerhalb der Kernzeit dokumentiert.

BEI NEIN

Innerhalb einer Woche: Bereitschaftsplan formalisieren. Innerhalb eines Monats: Drill durchführen, Zeit nehmen, Schwachstellen festhalten.

. . .

C · HAFTUNG — WER IM SCHADENSFALL UNTERSCHRIEBEN HAT

C1 EU-AI-Act-Klassifikation: Ist jeder KI-Workload schriftlich klassifiziert, mit Begründung und Unterschrift?

WARUM ES ZÄHLT

Übergangsfristen für Hochrisiko-Kategorien laufen im August 2026 aus. Sanktionen reichen bis 30 Mio. EUR oder 6 % des globalen Jahresumsatzes. Eine Klassifikation, die niemand unterschrieben hat, ist keine Klassifikation.

WIE PRÜFEN

Pro Workload: Klassifikation (Hochrisiko / nicht hochrisiko / nicht abgedeckt), Begründung mit Verweis auf die einschlägige Kategorie, Datum, unterschreibende Person mit Funktion.

BEI NEIN

Liste aller produktiven KI-Workloads erstellen, drei Wochen für die Klassifikation einplanen, externen Compliance-Check für die Hochrisiko-Kandidaten budgetieren.

C2 Provider/Deployer-Mapping: Existiert pro Modul eine Verantwortungsmatrix, die Provider, Deployer, Versicherungsdeckung benennt?

WARUM ES ZÄHLT

Haftung ist selten monolithisch. Modell, Infrastruktur, Integration und Betrieb tragen je eigene Provider-Deployer-Pflichten. Eine Klage findet die Lücke, in der niemand unterschrieben hat.

WIE PRÜFEN

Eine A4-Seite pro Workload (siehe Kapitel 4, *einseitiges Verantwortungsdiagramm*): Provider, Deployer, KI-Operator, Compliance Officer, Versicherungspolicen-Nummer, Stand.

BEI NEIN

Eine Seite pro System ist eine 30-Minuten-Übung mit den richtigen Personen. Ohne diese Seite ist eine eigene Compliance-Architektur nicht vorhanden.

C3 Vendor-Souveränität: Existiert für jeden externen KI- oder Cloud-Anbieter ein Notfallplan für Zugangsbeschränkungen?

WARUM ES ZÄHLT

Sovereign-Cloud-Etiketten verbergen häufig US-Mutterkonzerne mit Cloud-Act- oder ITAR-Verpflichtungen. Eine regionale API-Beschränkung kann ein Real-Time-System binnen Sekunden in einen *degraded mode* kippen (Kapitel 6, Fall D).

WIE PRÜFEN

Pro Anbieter: Jurisdiktion der Mutterstruktur, Vertragsklausel zu Zugangseinschränkungen, dokumentierter Degraded Mode für 4 h / 24 h / 72 h, Zeit bis zum Anbieterwechsel.

BEI NEIN

Die kritischste KI-Abhängigkeit identifizieren, Worst-Case-Szenario beschreiben, einen ersten Degraded Mode entwerfen — auch wenn er unvollständig ist. Unvollständig schlägt nicht vorhanden.

C4 Versicherungsdeckung: Deckt Ihre Police explizit Schäden durch fehlerhafte KI-Entscheidungen?

WARUM ES ZÄHLT

Standardpolicen für Betriebshaftpflicht und Cyberrisiken decken KI-induzierte Schäden meist nicht oder nur in engen Bereichen ab. Im Schadensfall steht der Selbstbehalt nicht in einer juristischen, sondern in einer operativen Größenordnung.

WIE PRÜFEN

Schriftliche Bestätigung des Versicherers für jeden Hochrisiko-Workload: was ist gedeckt, mit welchen Selbstbehalten, mit welchen Ausschlüssen. Datum der letzten Überprüfung.

BEI NEIN

Termin mit Versicherer innerhalb von 30 Tagen. Bei spezialisierten KI-Policen: zwei bis drei Vergleichsangebote vor Vertragsverlängerung einholen.

AUSWERTUNG

PUNKTE ZÄHLEN

Pro Frage: **2 Punkte** wenn vollständig erfüllt mit schriftlichem Beleg. **1 Punkt** wenn teilweise (z. B. Person benannt, aber Vertretung fehlt). **0 Punkte** wenn nicht vorhanden oder nur informell. Maximum: 24.

SCORE	LAGE	EMPFEHLUNG
20–24	Audit-fest	Jährliche Re-Validierung. Externe Stichprobe zur Bestätigung.
14–19	Lückenhaft	Drei größte Lücken priorisieren, in 90 Tagen schließen. Quartals-Governance einrichten.
8–13	Kritisch	Sofortmaßnahmen für Block C (Haftung). Externe Compliance-Begleitung dringend empfohlen.
0–7	Eskalations-Risiko mit GPU	Produktive KI-Systeme stoppen oder degradieren bis grundlegende Rollen besetzt sind. Geschäftsführung informieren.

Ein niedriger Score ist kein moralisches Urteil. Er ist eine Bestandsaufnahme — und zugleich der Aktionsplan für die nächsten 90 Tage. Die meisten DACH-Industrieorganisationen liegen beim ersten Durchgang im Bereich 8–13. Das ist kein Anlass zur Beunruhigung. Es ist der Anfang.

Das Audit als ausfüllbares Arbeitsblatt mit zusätzlichen Vorlagen
(Verantwortungsmatrix, Eskalationsplan, Datenkarte) — auf Anfrage: alpitype.de/buch