

KI-Systeme kontrollierbar einführen

Beratungsangebot für Medizinische Dienste, Sozialversicherungsträger und Gesundheitsdienstleister

Ausgangslage

Immer mehr öffentliche Einrichtungen führen Assistenzsysteme auf Basis großer Sprachmodelle ein: zur Recherche in internen Wissensbeständen, zur Vorbereitung von Vorgängen, zur Textentlastung im Begutachtungsprozess. Der Nutzen ist erheblich. Zugleich entsteht eine Systemklasse, deren Ausgaben nicht deterministisch sind und die über Wissensanbindung, Schnittstellen und Werkzeuge auf Daten und Funktionen zugreift, die zuvor klar abgegrenzt waren.

Wo besonders schutzbedürftige Sozial- und Gesundheitsdaten verarbeitet werden und wo Ergebnisse in Begutachtungen einfließen, die Leistungsentscheidungen tragen, genügt es nicht, die Einführung rechtlich zu dokumentieren. Es muss technisch nachweisbar sein, worauf ein System zugreift, was es selbst auslösen darf, wie eine Ausgabe zustande kam und wie sich der Betrieb im Zweifelsfall anhalten lässt.

Unsere Abgrenzung

Wir prüfen nicht nur, was dokumentiert werden muss, sondern ob sich das KI-System technisch und organisatorisch tatsächlich kontrollieren lässt.

Die juristische Bewertung des AI Act bieten Kanzleien und Datenschutzberatungen an. Unser Beitrag liegt davor und danach: in der Architektur, im Berechtigungsmodell, in der Protokollierung und im praktischen Nachweis, dass die vorgesehenen Schutzmechanismen dem Angriff standhalten.

Was wir untersuchen

Angriffsfläche – direkte und indirekte Prompt Injection, Anbindung externer Wissensquellen und Vektordatenbanken, unbeabsichtigter Datenabfluss, Absicherung von Modell-, Plugin- und Schnittstellenintegrationen.

Handlungsfähigkeit – welche Aktionen das System selbst auslösen darf, welche Identitäten und Berechtigungen dabei wirken, wie Handlungsspielräume technisch begrenzt sind und wo die Grenze zwischen Unterstützung, Empfehlung und Entscheidung verläuft.

Nachvollziehbarkeit – revisionsfähige Protokollierung von Eingaben, Ausgaben und Systemaktionen, wirksame menschliche Kontrolle, Eskalations- und Abschaltverfahren, Rekonstruierbarkeit fehlerhafter oder manipulierter Ergebnisse.

Governance – Inventar der eingesetzten und geplanten Systeme, Rollen und Freigabeprozesse, Abhängigkeiten von Anbietern, Modellen und Cloud-Diensten sowie die Einordnung in die Risikokategorien des AI Act.

Drei Varianten

Variante	Inhalt	Aufwand	Ergebnis
1 Scoping-Workshop	Aufnahme von Anwendungsfall, Datenquellen, Architektur und Freigabewegen. Ein Tag vor Ort, mit Vorbereitung und Auswertung.	ca. 3 Personen -tage	Kompaktes Ergebnispapier mit den wesentlichen Risikopunkten und einer Empfehlung zum weiteren Vorgehen.
2 AI Security Readiness Assessment	Systematische Prüfung der vier oben genannten Bereiche entlang der	8-14 Personen -tage	Managementbewertung, Risikoregister, priorisierte

Variante	Inhalt	Aufwand	Ergebnis
	bestehenden oder geplanten Architektur.		Maßnahmenliste und Umsetzungsroadmap.
3 Begleitung der Einführung	Mitwirkung von der Architektur über das Sicherheits- und Governance-Konzept und die Komponentenauswahl bis zur Betriebsfreigabe. Optional AI Red Teaming in abgeschotteter Testumgebung.	nach Zuschnitt	Umsetzungsfähige Konzepte, geprüfte Komponenten und eine belastbare Grundlage für die Freigabeentscheidung.

Ein AI Red Teaming findet ausschließlich in einer abgegrenzten Testumgebung mit synthetischen oder ausdrücklich freigegebenen Testdaten statt – nicht mit realen Sozial- oder Versichertendaten.

Regulatorischer Rahmen

Mit dem 2. August 2026 hat der EU AI Act eine weitere Anwendungsstufe erreicht; unter anderem gelten seither die Transparenzpflichten nach Art. 50. Die wesentlichen Pflichten für Hochrisiko-Systeme nach Anhang III wurden durch die Änderungsverordnung vom Juli 2026 auf den 2. Dezember 2027 verschoben. Dieses Zeitfenster ist kein Grund abzuwarten, sondern die Gelegenheit, Architektur und Kontrollmechanismen strukturiert aufzubauen, statt sie später unter Zeitdruck nachzurüsten.

Ob ein konkretes System als hochriskant einzustufen ist, hängt von Zweck, Entscheidungsnähe und Eingriffstiefe ab. Anhang III Nr. 5 Buchstabe a erfasst Systeme, die durch oder im Auftrag öffentlicher Stellen den Zugang zu wesentlichen Unterstützungs- und Gesundheitsleistungen bewerten. Wer sich auf eine Ausnahme nach Art. 6 Abs. 3 beruft, muss diese Bewertung dokumentieren. Hinzu kommen der Sozialdatenschutz nach § 35 SGB I und §§ 67 ff. SGB X sowie die Zweckbindung der an den Medizinischen Dienst übermittelten Daten.

Nächster Schritt

Wir schlagen den Scoping-Workshop als Einstieg vor. Er ist im Umfang überschaubar, erfordert keine Vorarbeit auf Ihrer Seite außer der Bereitstellung vorhandener Unterlagen und liefert eine belastbare Grundlage für die Entscheidung, ob und in welcher Tiefe eine weitere Begleitung sinnvoll ist.