ÖFFENTLICHE VERWALTUNG

Generative KI. Sicher einführen.

Ein pragmatisches Vorgehensmodell an der Schnittstelle von IT-Sicherheit, EU AI Act und Fachbereich.

Sicherheit als Handlungsrahmen

Risiken systematisch einordnen, Schutzmaßnahmen technisch verankern und Fachbereiche zur verantwortlichen Nutzung befähigen.

Die Ausgangslage in der Praxis

Fachbereiche wollen Sprachmodelle zur Aktenauswertung, zur Vorbereitung von Bescheiden oder als Wissensmanagement-Assistenten nutzen. Gleichzeitig treffen neue technische Risiken auf hohe Anforderungen an Informationssicherheit und Datenschutz.

Die Vorlage beschreibt ein Spannungsfeld zwischen vollständiger Blockade und unkontrollierter Nutzung privater Konten. Das vierstufige Modell schafft einen verlässlichen Rahmen für die Einführung.

01 Risiken einordnen

02 Architektur absichern

03 Verantwortung regeln

04 Im ISMS verankern

PHASE 01

Einsatzzweck klären.
Risiken sichtbar machen.

Vor Beschaffung und Schnittstellenfreigabe: Anwendungsfälle, Daten und Auswirkungen strukturiert erfassen.

01 Use Cases systematisch erfassen

Daten: Offene Verwaltungsdaten, personenbezogene Informationen oder Berufs- und Amtsgeheimnisse?
Autonomie: Formulierungshilfe oder vorbereitende Entscheidungsgrundlage?
Betroffene: Wer nutzt das System intern, wer ist extern betroffen?

02 Regulatorik und Schutzbedarf prüfen

Abgleich mit Anhang III des EU AI Act, etwa bei Leistungsansprüchen, Personalentscheidungen oder Strafverfolgung. Schutzbedarf für Vertraulichkeit, Integrität und Verfügbarkeit nach BSI-Methodik ermitteln.

Ergebnis: ein zentrales KI-Inventar

Das Ampelmodell der Vorlage übersetzt die Einordnung in einen operativen Freigabeprozess.

GRÜN Standardisierte Freigabe für unkritische Assistenzaufgaben.

GELB Freigabe mit spezifischen technischen Schutzmaßnahmen und Dokumentationsauflagen.

ROT Verbot oder Stopp bis zur formalen Konformitätsbewertung laut Vorgehensmodell.

PHASE 02

Datenflüsse schützen. Zugriffe begrenzen.

Schutz vor Datenabfluss und semantischer Manipulation wird auf Systemebene verankert.

01 Inferenz passend zum Schutzbedarf

Für vertrauliche Vorgänge priorisiert die Vorlage lokale Open-Weight-Modelle auf eigener Hardware oder souveräne Rechenzentren. BSI C5 und ISO 27001 werden als relevante Nachweise genannt.

Bei API-Diensten: vertragliche Vorgaben zu Zero Data Retention und zum Ausschluss von Training mit Behördendaten technisch absichern.

02 Berechtigungen bis ins Retrieval durchsetzen

Das Sprachmodell darf nur Informationen abrufen, die der anfragenden Person zugänglich sind. Berechtigungen aus DMS und Fachverfahren müssen auch im Vektor-Index gelten: Document-Level Security.

03 Semantische Angriffe berücksichtigen

Eingabe- und Ausgabefilter (Guardrails) zur Erkennung indirekter Prompt Injections vorsehen, besonders bei der automatisierten Analyse externer Dokumente wie Bürger-Mails oder PDFs.

Leitprinzip für RAG

Das System darf niemals mehr „wissen“ als die abfragende Person.

PHASEN 03 & 04

Verantwortung regeln. Sicherheit im Betrieb leben.

03 Governance & Mensch im Prozess

Human-in-the-Loop. Im vorgeschlagenen Modell erstellt die KI Entwürfe. Vor Aktenrelevanz oder Weiterleitung an Bürger werden Fakten, Quellen und Zahlen aktiv durch qualifizierte Mitarbeitende geprüft; die Vorlage sieht das Vier-Augen-Prinzip vor.

KI-Kompetenz & Automation Bias. Praxisnahe Schulungen mit Bezug auf Art. 4 EU AI Act: Halluzinationen erkennen, Prompts zur Fehlervermeidung formulieren und unzulässige Dateneingaben vermeiden.

Verantwortlichkeiten. Datenschutz, ISB/CISO und Fachbereichsleitung verbinden. Zuständigkeiten für die laufende Systemüberwachung personell festlegen.

04 Verankerung im behördlichen ISMS

Incident Management

Melde- und Reaktionswege für sensible Fehleingaben und fehlerhafte Bescheidentwürfe definieren.

Änderungsmanagement

Bei Modell-Updates Guardrails und RAG-Pipelines erneut prüfen und auditieren.

Protokollierung & Nachweisführung

Manipulationssichere Logs zu Abfragen und Entscheidungsgrundlagen vorsehen, ohne unzulässige Verhaltenskontrolle.

Sicherheit und KI-Nutzung gemeinsam gestalten

Fachbereich, IT und Recht früh verbinden, um Geschwindigkeit, Datensouveränität und Informationssicherheit zusammenzubringen.

Redaktionell aufbereitete Arbeitsprobe auf Basis der bereitgestellten Vorlage; keine eigenständige rechtliche oder technische Prüfung.