
Anlage Künstliche Intelligenz

1 Gegenstand der Anlage

Diese Anlage regelt die besonderen Anforderungen an den Einsatz von KI-Systemen und KI-gestützten Funktionen im Rahmen des Vertrags. Sie ergänzt die übrigen Vertragsbestandteile, insbesondere die Anlagen Informationssicherheit und DORA.

Die dort geregelten Anforderungen an Informationssicherheit, digitale operationale Resilienz, IKT-Risikomanagement, IKT-Drittparteiensteuerung, Prüfungsrechte, Nachweise und Mitwirkungspflichten gelten uneingeschränkt auch für die KI-bezogenen Leistungen des Auftragnehmers.

Der Auftragnehmer stellt die KI-bezogenen Leistungen so bereit, dass der Auftraggeber deren Nutzung, Risiken, Datenverarbeitung, Sicherheitsmaßnahmen und Modellgrundlagen nachvollziehen und prüfen kann.

Soweit Anforderungen dieser Anlage mit Anforderungen aus den Anlagen Informationssicherheit oder DORA zusammentreffen, gelten die Anforderungen kumulativ. Bei inhaltlichen Abweichungen ist diejenige Regelung maßgeblich, die das höhere Schutzniveau für den Auftraggeber gewährleistet, sofern nicht ausdrücklich etwas anderes vereinbart ist.

2 Architekturkonzept

Der Auftragnehmer beschreibt das eingesetzte KI-System einschließlich der für Bereitstellung, Betrieb und Nutzung wesentlichen Systemkomponenten, Modelle, Schnittstellen, Datenverarbeitungsprozesse, Betriebsumgebungen und Drittleistungen in prüffähiger Form.

Mit Angebotseinreichung hat der Auftragnehmer ein Architekturkonzept bereitgestellt. Dieses beschreibt die fachliche, technische, sicherheitsbezogene und betriebliche Zielarchitektur der KI-Plattform. Es erläutert, wie die einzelnen Softwarekomponenten zusammenwirken, wie sie in die Gesamtverarbeitung integriert sind und welche Rechenressourcen für Betrieb, Training, Validierung und Test des KI-Systems eingesetzt werden.

Das Architekturkonzept umfasst eine Darstellung der Systemkomponenten, der Datenflüsse, der Speicherorte, der Schnittstellen, der eingesetzten Modelle, der eingebundenen Drittleistungen, der Authentifizierungs- und Berechtigungskonzepte, der Protokollierung, der Sicherheitsmaßnahmen sowie der Betriebsverantwortlichkeiten.

3 Governance, Risikomanagement und Nachvollziehbarkeit

Der Auftragnehmer legt eine nachvollziehbare Einordnung des KI-Systems nach der Verordnung (EU) 2024/1689 vor. Dabei ist insbesondere darzustellen, ob das KI-System oder einzelne Systemkomponenten als verbotenes KI-System, Hochrisiko-KI-System, KI-System mit Transparenzpflichten, KI-System mit allgemeinem Verwendungszweck oder sonstiges KI-System einzuordnen sind.

Der Auftragnehmer legt außerdem offen, welche Rolle er und die von ihm eingesetzten Dritten im Sinne der KI-Verordnung einnehmen. Dies umfasst insbesondere Rollen als Anbieter, Betreiber, Einführer, Händler, Bevollmächtigter, Produkthersteller oder sonsti-

ger Akteur. Änderungen der Rollenverteilung sind dem Auftraggeber unverzüglich mitzuteilen.

Vor produktiver Nutzung stellt der Auftragnehmer eine dokumentierte Risikobewertung des KI-Systems bereit. Diese enthält die identifizierten KI-spezifischen Risiken, die vorgesehenen Risikominderungsmaßnahmen, die verbleibenden Restrisiken sowie deren Bewertung im Hinblick auf die vereinbarte Zweckbestimmung.

Die Plattform stellt Funktionen bereit, mit denen der Auftraggeber Nutzung, Konfiguration, Modellversionen, Berechtigungen, Eingaben, Ausgaben, Quellenbezüge, relevante Änderungen und sicherheitsrelevante Ereignisse nachvollziehen und auswerten kann.

Die KI-Plattform darf ohne gesonderte schriftliche Freigabe des Auftraggebers und ohne vorherige rechtliche, fachliche und risikobezogene Bewertung nicht für vollautomatisierte Einzelentscheidungen mit rechtlicher oder vergleichbar erheblicher Wirkung, für Profiling natürlicher Personen, für biometrische Identifikation, für Beschäftigtenbewertung, für Kreditwürdigkeitsentscheidungen oder für sonstige Nutzungen eingesetzt werden, die eine Einstufung als Hochrisiko-KI-System im Sinne der KI-Verordnung begründen können.

4 Menschliche Aufsicht

Die Plattform muss berechtigten Nutzern und Administratoren ermöglichen, Ausgaben des KI-Systems zu prüfen, zu verwerfen, zu korrigieren oder nicht zu verwenden. Soweit nach Zweckbestimmung und Risikoeinstufung erforderlich, müssen KI-gestützte Funktionen deaktiviert, unterbrochen oder in einen sicheren Zustand versetzt werden können.

Der Auftragnehmer stellt geeignete Informationen und Funktionen bereit, die Nutzer und Administratoren in die Lage versetzen, Fähigkeiten, Grenzen, Risiken, Anomalien, Fehlfunktionen und unerwartete Leistungen des KI-Systems zu erkennen und angemessen darauf zu reagieren.

5 KI-spezifische Sicherheitsanforderungen

Der Auftragnehmer darf Auftraggeberdaten ausschließlich zur Erfüllung des Vertrags verwenden. Eine Nutzung von Auftraggeberdaten, insbesondere von Prompts, hochgeladenen Dokumenten, erzeugten Inhalten, Metadaten, Protokollen, Ausgaben, Embeddings oder Indizes, für Training, Fine-Tuning, Produktverbesserung, Benchmarking, Modellvalidierung oder sonstige Zwecke außerhalb der Vertragserfüllung ist ohne vorherige ausdrückliche schriftliche Zustimmung des Auftraggebers ausgeschlossen.

Soweit die Plattform retrieval-gestützte Antwortgenerierung, insbesondere Retrieval-Augmented Generation (RAG), einsetzt, muss sie bestehende Berechtigungskonzepte der angebundenen Quellsysteme wirksam berücksichtigen. Nutzer dürfen über die KI-Plattform keine Informationen erhalten, auf die sie im jeweiligen Ursprungssystem nicht zugriffsberechtigt sind.

Das Sicherheitskonzept muss neben allgemeinen IT-Sicherheitsanforderungen auch KI-spezifische Bedrohungen berücksichtigen. Hierzu zählen abhängig vom Einsatzszenario insbesondere Prompt Injection, Jailbreaks, Datenabfluss über Modellantworten, unzulässige Nutzung schutzbedürftiger Daten, Manipulation von Modellen oder Referenzdaten,

Missbrauch von Agenten- oder Plug-in-Funktionen sowie Risiken aus angebundenen Wissensquellen.

Die KI-Plattform muss eine automatische Protokollierung der für Betrieb, Sicherheit, Nachvollziehbarkeit, Fehlersuche, Compliance und Auditierung relevanten Ereignisse ermöglichen. Die Protokollierung muss dem Zweck des KI-Systems, dem Schutzbedarf der verarbeiteten Daten und den vernünftigerweise vorhersehbaren Fehlanwendungen angemessen sein.

Die Protokolle müssen insbesondere geeignet sein, sicherheitsrelevante Ereignisse, wesentliche Änderungen, ungewöhnliches Systemverhalten, Fehlfunktionen, missbräuchliche Nutzung, Änderungen von Modellversionen, Administratorhandlungen, Datenzugriffe sowie relevante Eingaben und Ausgaben nachvollziehbar zu machen.

6 KI-spezifische Änderungen

Der Auftragnehmer stellt ein Änderungs- und Versionierungskonzept bereit. Dieses beschreibt, wie Änderungen an der KI-Plattform, an Modellen, Trainings-, Validierungs- oder Testdaten, Systemkomponenten, Schnittstellen, Sicherheitsmaßnahmen, Protokollierungsfunktionen und Betriebsprozessen gesteuert, dokumentiert, getestet und dem Auftraggeber angezeigt werden.

Änderungen, die Zweckbestimmung, Modellgrundlage, Datenverarbeitung, Sicherheitsarchitektur, Protokollierung, Subdienstleister, Speicherorte, menschliche Aufsicht oder Risikoeinstufung betreffen, bedürfen der vorherigen schriftlichen Zustimmung des Auftraggebers. Sonstige Änderungen mit Auswirkungen auf Betrieb, Nutzung, Dokumentation oder Nachvollziehbarkeit sind dem Auftraggeber rechtzeitig vor Umsetzung mitzuteilen.

Wesentliche Änderungen dürfen erst umgesetzt werden, nachdem deren Auswirkungen auf Sicherheit, Datenschutz, Informationssicherheit, Transparenz, menschliche Aufsicht, Leistungsfähigkeit und Risikoeinstufung bewertet und dokumentiert wurden.

7 Rechte an Daten

Alle Rechte an Daten des Auftraggebers, einschließlich daraus abgeleiteter, bereinigter, angereicherter, aggregierter oder sonst veränderter Fassungen, verbleiben beim Auftraggeber, soweit nicht ausdrücklich abweichend vereinbart. Der Auftragnehmer ist nicht berechtigt, Daten des Auftraggebers für andere Zwecke als die Erfüllung des Vertrags zu nutzen.

Sämtliche im Rahmen der Nutzung der KI-Plattform erzeugten Ausgaben, Arbeitsergebnisse, Konfigurationen, Prompts, Prompt-Bibliotheken und kundenspezifischen Workflows stehen dem Auftraggeber zu, soweit gesetzlich zulässig und nicht Rechte Dritter entgegenstehen. Der Auftragnehmer erhält hieran keine über die Vertragserfüllung hinausgehenden Nutzungsrechte.

Auf Verlangen des Auftraggebers sowie spätestens bei Vertragsende hat der Auftragnehmer sämtliche Auftraggeberdaten einschließlich Prompts, Ausgaben, hochgeladener Dokumente, Metadaten, Protokolle, Konfigurationen, kundenspezifischer Prompts und sonstiger kundenbezogener Artefakte in einem gängigen maschinenlesbaren Format

herauszugeben oder nach Wahl des Auftraggebers nachweisbar zu löschen.

8 Technische Dokumentation

Die Lieferung der KI-Plattform umfasst die Übergabe einer vollständigen technischen Dokumentation und einer Betriebs- und Nutzungsdokumentation in deutscher Sprache. Die Unterlagen müssen so beschaffen sein, dass der Auftraggeber oder ein von ihm beauftragter sachverständiger Dritter die Konformität der Plattform mit den vertraglichen Anforderungen, den technischen Spezifikationen, den Sicherheitsanforderungen und den Anforderungen an Transparenz, menschliche Aufsicht, Robustheit und Cybersicherheit bewerten kann.

9 KI-spezifische Vorfälle

Der Auftragnehmer informiert den Auftraggeber unverzüglich über KI-bezogene Sicherheitsvorfälle. Dies gilt insbesondere bei unbefugtem Zugriff auf Auftraggeberdaten, Datenabfluss über Modellantworten, fehlerhafter Berechtigungsvererbung, Manipulation von Modellen oder Daten, schwerwiegenden Fehlfunktionen, nicht autorisierten Modelländerungen oder sonstigen Ereignissen, die die sichere oder vertragsgemäße Nutzung der KI-Plattform beeinträchtigen können.

Die Melde-, Mitwirkungs- und Nachweispflichten aus den Anlagen Informationssicherheit und DORA bleiben unberührt.