ELEVAGENT AI · COMPLIANCE AUDIT · VERTRAULICH

KI-Risiko-Audit & Strategie-Analyse

Prüfung gemäß EU AI Act, DSGVO & den Standards des Zertifizierten KI-Beauftragten nach ÖVE/ÖNORM EN ISO/IEC 17024

PROJEKT	**SmartInbox – KI-gestützte E-Mail-Verarbeitung und Kundenkommunikation**
AUFTRAGGEBER	**Hartmann & Kern Unternehmensberatung**
BRANCHE	**Unternehmensberatung mit gelegentlichen HR-Beratungsprojekten**
DATUM	**5. April 2026**
REPORT-ID	**EA-202604-0686**

Erstellt von: ElevAgent AI
Anton Bauer, Zertifizierter KI-Beauftragter
office@elevagent.eu

ElevAgent AI

Entscheidungsvorlage für das Management

EMPFOHLENE MANAGEMENT-ENTSCHEIDUNG

Freigabe mit Auflagen

EXECUTIVE DECISION

Bedingte Freigabe ausschließlich für die weitere Prüf- und Auswahlphase. Eine sofortige Vertragsunterzeichnung wird vor Abschluss der datenschutzrechtlichen, vertraglichen und technischen Klärungen nicht empfohlen. Die finale Anbieterentscheidung sollte erst nach Vorlage belastbarer Vertrags- und Sicherheitsunterlagen sowie nach Abschluss der Datenschutz-Vorprüfung getroffen werden. Bei positivem Ergebnis der Prüfungen erscheint eine Freigabe mit Auflagen realistisch.

RISIKOKLASSE

BEGRENZTES RISIKO nach Art. 50 EU AI Act - Begründung: Das System dient der unterstützenden E-Mail-Verarbeitung mit menschlicher Freigabe und trifft keine autonomen Entscheidungen über Personen. Es fallen Transparenzpflichten an, insbesondere hinsichtlich der Kennzeichnung KI-generierter Entwürfe. Eine Hochrisiko-Einstufung wäre nur bei einer Erweiterung auf Personalbewertungen oder Bewerberauswahl im Rahmen der HR-Beratungsprojekte zu prüfen.

HAUPTGRÜNDE

- ✓ Unklare Nutzung von Kundendaten für Training, Feinabstimmung oder sonstige Modellverbesserung durch beide Anbieter, was ein erhebliches Vertraulichkeits- und Datenschutzrisiko darstellt.
- ✓ Systematische Verarbeitung unstrukturierter E-Mails mit vertraulichen Unternehmens-, Finanz- und HR-Daten ohne geprüfte Auftragsverarbeitungsverträge und ohne dokumentierte Datenflussanalyse.
- ✓ Offene Fragen zu Cloud-Nutzung und möglichen Drittlandtransfers beim US-basierten Anbieter sowie fehlende Modelltransparenz und Subdienstleister-Dokumentation beim deutschen Anbieter.

Risiko & Freigabe

GRÖSSTES RESTRISIKO

Das größte verbleibende Risiko besteht in der Fehlzuweisung oder unzulässigen Offenlegung vertraulicher Kundeninformationen innerhalb des Unternehmens oder gegenüber dem Anbieter. Zusätzlich besteht das Risiko fehlerhafter, aber menschlich freigegebener Antwortentwürfe mit falschen Projekt- oder Angebotsinformationen, insbesondere unter Zeitdruck. Beide Risiken könnten zu Reputationsschäden, Vertragsverletzungen gegenüber Großkunden und datenschutzrechtlichen Konsequenzen führen.

FREIGABEBEDINGUNGEN

- ✓ Prüfung und Abschluss eines belastbaren Auftragsverarbeitungsvertrags mit klarer vertraglicher und technischer Regelung zum Ausschluss der Nutzung von Kundendaten für Modellverbesserung, Training oder Feinabstimmung durch den Anbieter.
- ✓ Dokumentierte Klärung der tatsächlichen Datenflüsse, Hosting-Architektur, Unterauftragsverarbeiter und etwaiger Drittlandbezüge einschließlich erforderlicher Transfermechanismen wie Standardvertragsklauseln und Transfer Impact Assessment.
- ✓ Durchführung und Dokumentation einer Datenschutz-Folgenabschätzung, sofern die Vorprüfung deren Erforderlichkeit bestätigt, sowie Implementierung eines dokumentierten Berechtigungs-, Freigabe- und Qualitätskontrollkonzepts vor Produktivbetrieb.

ERFORDERLICHE MANAGEMENT-ENTSCHEIDUNG

Die Geschäftsleitung muss entscheiden, ob die Anbieterauswahl nach Vorlage und Prüfung der Vertrags- und Sicherheitsunterlagen fortgesetzt wird und ob externe Datenschutz- und IT-Sicherheitsunterstützung für die Due-Diligence-Prüfung eingebunden werden soll.

Management Summary

Das Projekt SmartInbox adressiert ein nachvollziehbares Effizienz- und Serviceproblem in der zentralen E-Mail-Bearbeitung und ist als unterstützender Office- und Kommunikations-Use-Case grundsätzlich plausibel. Nach aktueller Einschätzung handelt es sich voraussichtlich um ein System mit begrenztem Risiko nach dem EU AI Act, da keine autonomen Entscheidungen über Personen getroffen werden und eine menschliche Freigabe vorgesehen ist. Die wesentlichen Risiken liegen in der Verarbeitung unstrukturierter und teils vertraulicher Kundeninformationen, möglicher personenbezogener Daten aus HR-Beratungsprojekten, unklarer Anbietertransparenz zum Lern- und Trainingsverhalten sowie noch ungeprüften Vertrags- und Transferfragen. Das DSGVO-Risiko wird als hoch eingestuft, da eine systematische Verarbeitung sensibler Kommunikationsdaten mit Zusammenführung mehrerer Datenquellen erfolgt. Vor einer Beauftragung sind insbesondere Datenschutz, Vertraulichkeit, technische Datenflüsse, Rollenverteilung und Kontrollmechanismen verbindlich zu klären. Eine sofortige Vertragsunterschrift erscheint auf Basis des vorliegenden Inputs nicht belastbar. Es wird empfohlen, die Anbieterauswahl erst nach Abschluss einer strukturierten Due-Diligence-Prüfung und einer Vorprüfung zur Datenschutz-Folgenabschätzung fortzusetzen.

GESAMTBEWERTUNG

Freigabe mit Auflagen

USE CASE

Das Vorhaben sieht den Einsatz einer KI-Lösung zur automatisierten Verarbeitung eingehender E-Mails vor. Die KI soll E-Mails nach Thema und Dringlichkeit sortieren, intern zuweisen, Antwortentwürfe mit menschlicher Freigabe erstellen sowie Rechnungen und Belege erkennen und Projekten zuordnen. Zusätzlich sollen bei Angebotsanfragen relevante CRM-Daten aus HubSpot für eine Ersteinschätzung herangezogen werden. Die Lösung verarbeitet dabei unstrukturierte Kommunikationsdaten aus Microsoft 365, CRM-Daten und Buchhaltungsdaten aus BMD. Betroffen sind rund 85 aktive Firmenkunden, 12 Mitarbeiter sowie potenziell sensible Daten aus HR-Beratungsprojekten. Die operative Nutzung liegt primär bei der Office-Managerin, den Beratern und der Buchhaltung, wobei keine automatische externe Kommunikation ohne menschliche Freigabe erfolgen soll.

EMPFEHLUNG DES AI OFFICERS

Freigabe mit Auflagen wird empfohlen, jedoch ausschließlich für die weitere Prüfphase. Die Vertragsunterschrift ist vorerst zurückzustellen, bis beide Anbieter anhand eines strukturierten Due-Diligence-Fragenkatalogs geprüft wurden. Ein vermeintlicher Standortvorteil des deutschen Anbieters allein reicht nicht aus, da Modelltransparenz, Datenverwendung und Subdienstleister-Strukturen bei beiden Anbietern entscheidend und bislang unzureichend dokumentiert sind.

EU AI ACT RISIKOKLASSE

BEGRENZTES RISIKO nach Art. 50 EU AI Act - Begründung: Das System dient der unterstützenden E-Mail-Verarbeitung mit menschlicher Freigabe und trifft keine autonomen Entscheidungen über Personen. Es fallen Transparenzpflichten an, insbesondere hinsichtlich der Kennzeichnung KI-generierter Entwürfe. Eine Hochrisiko-Einstufung wäre nur bei einer Erweiterung auf Personalbewertungen oder Bewerberauswahl im Rahmen der HR-Beratungsprojekte zu prüfen.

Risiko-Übersicht

PRÜFBEREICH	BEWERTUNG
EU AI Act Konformität	MITTEL
DSGVO / Datenschutz	HOCH
Ethik & Fairness	MITTEL
Technische Umsetzbarkeit	MITTEL
Wirtschaftlichkeit / ROI	MITTEL

PFLICHTPRÜFUNGEN

DPIA — DATENSCHUTZ-FOLGENABSCHÄTZUNG

EMPFOHLEN — Sollte geprüft werden.

FRIA — GRUNDRECHTE-FOLGENABSCHÄTZUNG

ERFORDERLICH — Bei Hochrisiko-Systemen nach EU AI Act durchzuführen.

XAI — TRANSPARENZ & ERKLÄRBARKEIT

EMPFOHLEN — Grundlegende Transparenz sollte gewährleistet sein.

AI Value Creation Canvas

PROBLEM & LÖSUNG

Das Projekt adressiert die Überlastung der zentralen E-Mail-Bearbeitung und die Abhängigkeit von einer einzelnen Office-Managerin. Ziel ist die Reduktion manueller Sortier-, Zuweisungs- und Entwurfsarbeit um circa 60 Prozent sowie die Senkung der durchschnittlichen Antwortzeit von 8 auf unter 2 Stunden. Gleichzeitig sollen übersehene E-Mails reduziert und die Rechnungszuordnung beschleunigt werden. Der Nutzen liegt in verbesserter Servicequalität, schnellerer Reaktionsfähigkeit und operativer Entlastung.

STRATEGISCHER FIT

Die Teilautomatisierung von Routineaufgaben im E-Mail- und Rechnungseingang fügt sich in eine Strategie zur Steigerung operativer Effizienz und Servicequalität ein. Die Beibehaltung menschlicher Freigaben für ausgehende Kommunikation und kaufmännisch relevante Inhalte sichert die Qualitätskontrolle. Strategisch reduziert das Vorhaben die operative Abhängigkeit von einer einzelnen Person und schafft Kapazitäten für wertschöpfende Beratungstätigkeiten. Der Wettbewerbsvorteil liegt in schnellerer und zuverlässigerer Kundenkommunikation.

DATENSTRATEGIE

Die Datenbasis umfasst E-Mails und Anhänge aus Microsoft 365, CRM-Daten aus HubSpot sowie Buchhaltungsdaten aus BMD. Enthalten sein können Namen, Kontaktdaten, Projektinformationen, Vertragsinhalte, Rechnungsdaten, Bankdaten sowie fallweise personenbezogene Daten aus HR-Beratungsprojekten und vertrauliche Unternehmensunterlagen. Die CRM-Daten sind nach aktuellem Input gut gepflegt und werden quartalsweise aktualisiert. E-Mail-Daten sind hingegen unstrukturiert, sprachlich gemischt und in Qualität und Detailgrad variabel, was Fehlerrisiken bei Klassifikation und Extraktion erhöht. Bias-Risiken bestehen insbesondere bei der Priorisierung, wenn bestimmte Absender, Sprachen oder Formulierungsstile systematisch bevorzugt oder benachteiligt werden. Die Zusammenführung mehrerer Datenquellen erhöht die Komplexität und das Risiko unbeabsichtigter Datenoffenlegung.

TECHNOLOGIE & INFRASTRUKTUR

Zwei Anbieteroptionen stehen zur Auswahl: Eine individuelle Lösung auf Basis von GPT-4 mit AWS-Hosting laut Angebot in der EU sowie ein deutsches SaaS-Produkt mit eigenen Servern in Deutschland. Beide Lösungen erfordern Schnittstellen zu Microsoft 365, HubSpot und BMD. Belastbare Detailinformationen zu Modellarchitektur, Unterauftragnehmern, Logging, Verschlüsselung und konkreten Sicherheitsmaßnahmen fehlen bei beiden Anbietern. Beim US-Anbieter ist ein möglicher Drittlandtransfer trotz EU-Hosting nicht abschließend geklärt, beim deutschen Anbieter fehlt Transparenz zur eingesetzten Modellarchitektur und zu etwaigen Subdienstleistern.

12 Building Blocks

01 · PROJECT PURPOSE

Das Projekt wird durchgeführt, um die manuelle E-Mail-Bearbeitung im zentralen Postfach signifikant zu reduzieren und die Reaktionsgeschwindigkeit gegenüber Kunden zu verbessern. Das zu lösende Problem ist die Überlastung der Office-Managerin, die als Single Point of Failure für die gesamte eingehende Kommunikation fungiert. Der angestrebte Nutzen umfasst eine Reduktion des manuellen Aufwands um circa 60 Prozent, eine Senkung der Antwortzeit von 8 auf unter 2 Stunden sowie eine Verringerung übersehener oder fehlzugewiesener E-Mails. Für das Unternehmen entsteht operative Resilienz und Effizienz, für die Kunden eine schnellere und zuverlässigere Betreuung. Gesellschaftlich ist der Nutzen begrenzt, jedoch trägt eine verantwortungsvolle Umsetzung zur Demonstration guter KI-Governance in der Beratungsbranche bei.

02 · STRATEGIC VALUE

Das Vorhaben liefert strategischen Mehrwert durch die Verbesserung der Servicequalität und Reaktionsgeschwindigkeit, die in der Unternehmensberatung wesentliche Differenzierungsmerkmale darstellen. Die Reduktion der operativen Abhängigkeit von einer einzelnen Mitarbeiterin erhöht die organisatorische Resilienz und Skalierbarkeit. Durch die Entlastung von Routineaufgaben werden Kapazitäten für wertschöpfende Beratungstätigkeiten freigesetzt. Die Integration von CRM-Daten in die Angebotserstellung kann die Qualität und Geschwindigkeit der Akquise verbessern. Der Wettbewerbsvorteil liegt in der Kombination aus schnellerer Reaktion und konsistenterer Kommunikationsqualität. Der strategische Fit ist gegeben, sofern die Umsetzung die Vertraulichkeitsstandards der Branche wahrt.

03 · GOVERNANCE

Die Gesamtverantwortung für das Vorhaben liegt nach aktuellem Input bei der Geschäftsleitung von Hartmann und Kern. Die operative Nutzung erfolgt durch die Office-Managerin, die Berater und die Buchhaltung in ihren jeweiligen Zuständigkeitsbereichen. Ein formalisierter Kontrollprozess für die KI-Qualität, einschließlich regelmäßiger Stichproben, Eskalationsregeln und Verantwortlichkeiten für Fehlerkorrekturen, fehlt aktuell und muss vor Produktivbetrieb definiert werden. Die Pflicht zur Benennung eines Datenschutzbeauftragten ist angesichts der systematischen Verarbeitung personenbezogener Daten zu prüfen. Freigabe- und Eskalationswege für kritische Fälle, etwa bei vertraulichen Großkundendaten oder erkannten Fehlern, sind noch nicht dokumentiert. Es wird empfohlen, eine verantwortliche Person für die laufende KI-Governance zu benennen.

12 Building Blocks

04 · AI BOUNDARIES

Das System darf E-Mails nach Thema und Dringlichkeit sortieren, intern zuweisen, Rechnungen und Belege erkennen und Projekten zuordnen sowie Antwortentwürfe erstellen. Es darf keine E-Mails automatisch extern versenden, keine Preise oder Konditionen verbindlich festlegen und keine Zahlungen freigeben. Zahlungsfreigaben verbleiben ausschließlich beim Geschäftsführer, Preise und Konditionen bei Angeboten werden manuell geprüft. Menschliche Freigabe ist für jede ausgehende Kommunikation zwingend erforderlich. Bei besonders sensiblen Anhängen, Absendern oder Projekten sollte die Möglichkeit bestehen, diese von der KI-Verarbeitung auszunehmen. Die Grenzen des Systems müssen technisch durchgesetzt und nicht nur organisatorisch vereinbart werden.

05 · DATA QUALITY

Die CRM-Daten in HubSpot sind nach aktuellem Input gut gepflegt und werden quartalsweise aktualisiert, was eine solide Basis für die Angebotsersteinschätzung bietet. Die E-Mail-Daten sind hingegen unstrukturiert, sprachlich gemischt zwischen Deutsch und Englisch und in Qualität und Detailgrad erheblich variabel. Diese Heterogenität erhöht das Risiko von Fehlklassifikationen, falschen Priorisierungen und fehlerhaften Extraktionen. Verzerrungen könnten entstehen, wenn bestimmte Kommunikationsstile, Sprachen oder Absendertypen systematisch anders behandelt werden. Die Qualität der Buchhaltungsdaten aus BMD ist im Input nicht näher beschrieben und sollte vor der Anbindung geprüft werden. Insgesamt ist eine systematische Datenqualitätsprüfung vor Produktivbetrieb erforderlich.

06 · DATA STRATEGY

Die Verarbeitung erfolgt auf Live-Kommunikationsdaten aus dem operativen Geschäftsbetrieb, was besondere Anforderungen an Zweckbindung und Datenminimierung stellt. Die Datenquellen umfassen Microsoft 365 für E-Mails, HubSpot für CRM-Daten und BMD für Buchhaltungsdaten, wobei die Datenflüsse zwischen diesen Systemen und dem KI-Anbieter noch nicht dokumentiert sind. Eine klare Zweckbindung muss vertraglich und technisch sicherstellen, dass Daten ausschließlich für die definierten Verarbeitungszwecke genutzt werden. Ein Löschkonzept mit definierten Aufbewahrungsfristen ist zwingend erforderlich, insbesondere für E-Mail-Inhalte und Anhänge beim Anbieter. Vertragliche und technische Begrenzungen gegen Modelltraining, Feinabstimmung oder sonstige Sekundärnutzung durch den Anbieter müssen vor Vertragsschluss verbindlich vereinbart werden. Bei Anbieter 1 ist ein möglicher Drittlandtransfer trotz EU-Hosting zu klären, bei Anbieter 2 die Subdienstleister-Struktur.

12 Building Blocks

07 · TECHNOLOGY

Anbieter 1 setzt auf GPT-4 von OpenAI mit AWS-Hosting, das laut Angebot in der EU erfolgt, wobei die genaue Datenverarbeitungsarchitektur und etwaige Rückkanäle zu OpenAI nicht abschließend dokumentiert sind. Anbieter 2 nutzt eigene Modelle auf eigenen Servern in Deutschland, wobei Transparenz zur konkreten Modellarchitektur, Trainingsmethodik und etwaigen Subdienstleistern fehlt. Beide Lösungen erfordern Schnittstellen zu Microsoft 365, HubSpot und BMD, deren technische Absicherung, Authentifizierung und Zugriffssteuerung noch zu spezifizieren sind. Belastbare Informationen zu Logging, Verschlüsselung im Transit und at Rest, Monitoring, Incident Response und Backup-Konzepten liegen bei beiden Anbietern nicht vor. Die technische Umsetzbarkeit erscheint grundsätzlich gegeben, die Bewertung der Sicherheitsarchitektur ist jedoch ohne detaillierte Anbieterunterlagen nicht abschließend möglich. Eine unabhängige technische Prüfung der Datenflüsse und Sicherheitsmaßnahmen wird empfohlen.

08 · STAKEHOLDER MANAGEMENT

Direkt betroffen sind alle 12 Mitarbeiter von Hartmann und Kern, insbesondere die Office-Managerin als Hauptnutzerin, die Berater als Empfänger zugewiesener E-Mails und die Buchhaltung für die Rechnungszuordnung. Extern betroffen sind circa 85 aktive Firmenkunden, deren Kommunikation durch die KI verarbeitet wird, sowie deren Mitarbeiter und Ansprechpartner. Besonders relevant sind sensible Großkunden mit strengen Vertraulichkeitsklauseln, bei denen eine KI-Verarbeitung der Kommunikation vertraglich oder faktisch problematisch sein könnte. Die Geschäftsleitung trägt die Gesamtverantwortung und muss über Freigabe und Auflagen entscheiden. Ein Betriebsrat ist bei der Unternehmensgröße voraussichtlich nicht vorhanden, die Mitarbeiterinformation und -einbindung ist dennoch erforderlich. Die Einbindung eines Datenschutzberaters oder externen Datenschutzbeauftragten wird empfohlen.

09 · AI ETHICS BY DESIGN

Nachvollziehbare Priorisierungsregeln müssen definiert werden, um eine systematische Bevorzugung oder Benachteiligung bestimmter Absender, Themen oder Kommunikationsstile zu vermeiden. Die Vermeidung blinder Übernahme von KI-Vorschlägen erfordert eine bewusste Gestaltung der Freigabeprozesse, die kritische Prüfung fördert statt unterdrückt. Alle KI-generierten Entwürfe sollten klar als solche gekennzeichnet sein, um Transparenz gegenüber internen Nutzern und gegebenenfalls externen Empfängern zu gewährleisten. Ein dokumentiertes Verfahren zur Fehlerkorrektur und zum Feedback an das System ist erforderlich, um kontinuierliche Verbesserung und Lernfähigkeit der Organisation sicherzustellen. Die menschliche Autonomie und Letztentscheidung muss in allen Prozessschritten gewahrt bleiben, insbesondere bei Angeboten und kaufmännisch relevanten Inhalten. Die Würde und Privatsphäre der Kommunikationspartner ist durch Datenminimierung, Zweckbindung und angemessene Zugriffskontrollen zu schützen.

12 Building Blocks

10 · ROI & KPIS

Der erwartete Nutzen liegt in der Reduktion des Zeitaufwands der Office-Managerin, der Senkung der durchschnittlichen Antwortzeit und der Verringerung übersehener oder fehlzugewiesener E-Mails. Relevante KPIs umfassen die Korrekturquote bei Antwortentwürfen, die Genauigkeit der Rechnungszuordnung, die Bearbeitungszeit im Rechnungseingang und die Nutzerakzeptanz bei den Mitarbeitern. Die Wirtschaftlichkeit hängt von den Lizenz- oder Servicekosten der Anbieter, dem Implementierungsaufwand und den laufenden Betriebskosten im Verhältnis zur eingesparten Arbeitszeit ab. Konkrete Kostenzahlen liegen im Input nicht vor, sodass eine belastbare ROI-Berechnung noch aussteht. Risiken für den Business Case bestehen in einer niedrigeren als erwarteten Automatisierungsquote, hohen Korrekturaufwänden oder mangelnder Nutzerakzeptanz. Eine Pilotphase mit definierten Erfolgskriterien und Abbruchbedingungen wird empfohlen, um den Business Case vor einem vollständigen Rollout zu validieren.

11 · SYSTEM PROMPT

Ein System-Prompt ist aktuell nicht definiert und muss vor Produktivbetrieb erstellt, dokumentiert und freigegeben werden. Der Prompt muss Vertraulichkeitsregeln enthalten, die den Umgang mit sensiblen Daten, Geschäftsgeheimnissen und personenbezogenen Informationen steuern. Zulässige Datenquellen, Tonalität und sprachliche Vorgaben für Antwortentwürfe müssen klar definiert werden. Eskalationsregeln für Fälle, in denen das System unsicher ist oder sensible Inhalte erkennt, sind zwingend erforderlich. Der Ausschluss automatischer Zusagen, Preisfestlegungen oder verbindlicher Aussagen muss im Prompt technisch verankert werden. Review-Pflichten für Prompt-Änderungen und ein Versionierungskonzept sollten etabliert werden, um unkontrollierte Verhaltensänderungen des Systems zu vermeiden.

12 · CHANGE MANAGEMENT

Alle betroffenen Mitarbeiter müssen vor Produktivbetrieb zu den neuen Freigabeprozessen, der Erkennung von KI-Fehlern und Halluzinationen sowie zu Vertraulichkeitsanforderungen geschult werden. Die Schulung sollte praktische Szenarien umfassen, in denen fehlerhafte Entwürfe, Fehlzuweisungen und sensible Inhalte erkannt und korrekt behandelt werden. Berechtigungskonzepte und Zugriffsregeln müssen kommuniziert und verstanden werden, insbesondere hinsichtlich projektbezogener Vertraulichkeit. Ausfallverfahren für den Fall eines Systemausfalls oder einer Fehlfunktion müssen definiert und geübt werden. Prozessänderungen, Prompt-Anpassungen und Incident-Meldungen sollten über einen definierten Change-Management-Prozess gesteuert werden. Die Akzeptanz der Lösung sollte regelmäßig erhoben und Feedback der Nutzer systematisch in die Weiterentwicklung einbezogen werden.

EU AI Act & DSGVO

EU AI ACT EINORDNUNG

Nach aktueller Einschätzung fällt das System voraussichtlich in die Kategorie begrenztes Risiko nach Art. 50 EU AI Act, da es als unterstützendes Kommunikations- und Klassifikationssystem mit menschlicher Freigabe konzipiert ist. Eine Einstufung als Hochrisiko-System nach Art. 6 in Verbindung mit Annex III erscheint derzeit nicht gegeben, da keine autonomen Entscheidungen über Personen getroffen werden und keine der in Annex III genannten Anwendungsbereiche unmittelbar betroffen sind. Transparenzpflichten bestehen insbesondere hinsichtlich der Kennzeichnung KI-generierter Inhalte gegenüber internen Nutzern und gegebenenfalls externen Empfängern. Eine Hochrisiko-Prüfung wäre erforderlich, falls das System künftig für Personalbewertungen, Bewerberauswahl oder andere in Annex III genannte Zwecke im Rahmen der HR-Beratungsprojekte eingesetzt werden sollte. Allgemeine Governance-Pflichten wie Dokumentation, Risikomanagement und menschliche Aufsicht sind auch bei begrenztem Risiko als Best Practice zu beachten. Die Einstufung sollte bei Änderungen des Einsatzzwecks oder der Funktionalität erneut geprüft werden.

DSGVO ANALYSE

Die systematische Verarbeitung eingehender E-Mails einschließlich Anhängen, CRM-Daten und Buchhaltungsdaten durch einen externen KI-Anbieter erfordert zwingend einen Auftragsverarbeitungsvertrag nach Art. 28 DSGVO. Als Rechtsgrundlage kommt voraussichtlich Art. 6 Abs. 1 lit. f DSGVO in Betracht, wobei eine sorgfältige Interessenabwägung zu dokumentieren ist, insbesondere hinsichtlich der berechtigten Erwartungen der Kommunikationspartner. Soweit im Rahmen von HR-Beratungsprojekten besondere Kategorien personenbezogener Daten nach Art. 9 DSGVO verarbeitet werden, ist eine gesonderte Rechtsgrundlage erforderlich und die Verarbeitung besonders streng zu prüfen. Informationspflichten nach Art. 13 und 14 DSGVO gegenüber Betroffenen, deren E-Mails verarbeitet werden, sind zu prüfen und umzusetzen. Zweckbindung, Datenminimierung, Speicherbegrenzung und Löschkonzepte müssen vertraglich und technisch sichergestellt werden. Bei Anbieter 1 ist ein möglicher Drittlandtransfer in die USA trotz EU-Hosting zu klären, wobei gegebenenfalls Standardvertragsklauseln und ein Transfer Impact Assessment erforderlich wären.

SANKTIONSRISIKEN

Bei Verstößen gegen die DSGVO drohen Bußgelder von bis zu 20 Millionen Euro oder 4 Prozent des weltweiten Jahresumsatzes nach Art. 83 DSGVO, wobei bei einem Unternehmen dieser Größe die praktische Relevanz eher in behördlichen Anordnungen und Reputationsschäden liegt. Bei Verletzung von Vertraulichkeits- oder Sicherheitszusagen gegenüber Kunden drohen vertragliche Ansprüche, Vertragsstrafen und erhebliche Reputationsschäden, die für eine Unternehmensberatung existenzbedrohend sein können. Sanktionsrisiken nach dem EU AI Act sind bei begrenztem Risiko geringer, können aber bei Verstößen gegen Transparenzpflichten bis zu 15 Millionen Euro oder 3 Prozent des Jahresumsatzes betragen.

Haftungsausschluss: Diese Analyse dient der betriebswirtschaftlichen, technischen und ethischen Bewertung. Sie stellt keine Rechtsberatung dar und ersetzt keine juristische Einzelfallprüfung. Optional kann ergänzend eine juristische Detailprüfung durch spezialisierte Rechtsanwälte für KI-, Datenschutz- und Technologierecht gegen Mehrkosten beauftragt werden.

Datenschutz-Folgenabschätzung

Die DPIA nach Art. 35 DSGVO ist bei Verarbeitungen mit hohem Risiko für die Rechte und Freiheiten natürlicher Personen durchzuführen.

ERFORDERLICHKEIT & BEWERTUNG

Eine Datenschutz-Folgenabschätzung ist nach aktueller Einschätzung voraussichtlich erforderlich und eine Vorprüfung ist zwingend durchzuführen. Die systematische Verarbeitung eingehender E-Mails mit potenziell sensiblen personenbezogenen Daten, die Zusammenführung mehrerer Datenquellen und die Übermittlung an einen externen KI-Anbieter erfüllen mit hoher Wahrscheinlichkeit die Kriterien des Art. 35 DSGVO. Insbesondere die Verarbeitung unstrukturierter Kommunikationsdaten, die vertrauliche Unternehmens-, Finanz- und HR-Informationen enthalten können, sowie die systematische Analyse und Kategorisierung dieser Daten sprechen für die Erforderlichkeit. Die Zusammenführung von E-Mail-, CRM- und Buchhaltungsdaten erhöht das Risikoprofil zusätzlich. Bei bestätigtem Erfordernis ist die DPIA vor Produktivbetrieb vollständig durchzuführen und zu dokumentieren. Die finale Entscheidung über die Erforderlichkeit sollte unter Einbeziehung eines Datenschutzberaters oder Datenschutzbeauftragten getroffen werden.

GEPRÜFTE ASPEKTE

- ✓ Systematische Analyse und Kategorisierung eingehender E-Mails und Anhänge durch einen externen KI-Anbieter mit Beschreibung aller Verarbeitungsvorgänge und Zwecke.
- ✓ Verarbeitung unstrukturierter Inhalte mit potenziell sensiblen personenbezogenen oder vertraulichen Daten, einschließlich möglicher besonderer Kategorien aus HR-Beratungsprojekten nach Art. 9 DSGVO.
- ✓ Zusammenführung von E-Mail-, CRM- und Buchhaltungsdaten mit Bewertung der Notwendigkeit und Verhältnismäßigkeit dieser Datenverknüpfung.
- ✓ Risiko unzulässiger Offenlegung durch Fehlrouting, unzureichende Berechtigungen oder technische Schwachstellen in den Schnittstellen.
- ✓ Risiko der Weiterverwendung von Daten durch den Anbieter zu nicht freigegebenen Zwecken wie Modelltraining oder Feinabstimmung.
- ✓ Bewertung technischer und organisatorischer Maßnahmen einschließlich Verschlüsselung, Zugriffssteuerung, Logging, Speicherfristen, Betroffenenrechte und Dokumentation von Restrisiken mit Freigabeprozess.

Grundrechte-Folgenabschätzung

Die FRIA ist bei Hochrisiko-KI-Systemen durch Betreiber strukturiert durchzuführen.

BETROFFENE GRUNDRECHTE

Eine Fundamental Rights Impact Assessment ist nach aktueller Einschätzung nicht zwingend erforderlich, da das System voraussichtlich als begrenztes Risiko einzustufen ist und keine autonomen Entscheidungen über Personen trifft. Dennoch wird eine freiwillige Grundrechtsbetrachtung empfohlen, da die systematische Verarbeitung von Kundenkommunikation Auswirkungen auf Privatsphäre, Vertraulichkeit und informationelle Selbstbestimmung der Kommunikationspartner haben kann. Besondere Aufmerksamkeit verdient der Schutz von Geschäftsgeheimnissen und vertraulichen Kundenunterlagen, die durch die KI-Verarbeitung einem zusätzlichen Offenlegungsrisiko ausgesetzt werden. Die Auswirkungen auf die Arbeitsabläufe und Verantwortlichkeiten der internen Mitarbeiter sollten ebenfalls berücksichtigt werden. Sollte das System künftig auf HR-bezogene Entscheidungsunterstützung erweitert werden, wäre eine FRIA als Pflichtprüfung erneut zu bewerten.

GEPRÜFTE ASPEKTE

- ✓ Schutz von Privatsphäre und Vertraulichkeit der Kommunikationspartner, deren E-Mails ohne deren unmittelbare Kenntnis durch eine KI verarbeitet werden.
- ✓ Vermeidung unangemessener oder intransparenter Priorisierung, die bestimmte Absender, Themen oder Kommunikationsstile systematisch bevorzugen oder benachteiligen könnte.
- ✓ Sicherstellung wirksamer menschlicher Aufsicht über alle KI-generierten Zuweisungen, Entwürfe und Extraktionen mit realer Prüfmöglichkeit.
- ✓ Schutz von Geschäftsgeheimnissen und vertraulichen Kundenunterlagen vor unbeabsichtigter Offenlegung gegenüber dem Anbieter oder durch Fehlrouting.
- ✓ Auswirkungen auf Arbeitsabläufe, Verantwortlichkeiten und Autonomie interner Mitarbeiter, insbesondere der Office-Managerin.
- ✓ Transparenz über den KI-Einsatz gegenüber betroffenen Personen, insbesondere externen Kommunikationspartnern und Kunden.

Ethische Analyse

Der Einsatz greift in sensible Kundenkommunikation ein und berührt damit Vertraulichkeit, Vertrauen und die berechtigten Erwartungen der Kommunikationspartner an den Umgang mit ihren Informationen. Ethisch relevant ist die Frage, ob Kunden und Kommunikationspartner darüber informiert werden, dass ihre E-Mails durch eine KI verarbeitet werden, und ob dies ihren Erwartungen an eine Unternehmensberatung entspricht. Die menschliche Letztkontrolle ist zwar vorgesehen, birgt jedoch das Risiko eines Automatisierungs-Bias, bei dem Mitarbeiter unter Zeitdruck KI-Vorschläge unkritisch übernehmen. Die Priorisierung von E-Mails durch die KI könnte zu einer systematischen Benachteiligung bestimmter Absender oder Themen führen, ohne dass dies bewusst wahrgenommen wird. Die Verlässlichkeit der KI-Entwürfe ist kritisch, da fehlerhafte Informationen in Angeboten oder Kundenkommunikation das Vertrauen nachhaltig beschädigen können. Machtasymmetrien bestehen insbesondere gegenüber den Kommunikationspartnern, die keine Kontrolle über die KI-Verarbeitung ihrer Nachrichten haben. Unbeabsichtigte Folgen könnten in einer schleichenden Entwertung menschlicher Urteilsfähigkeit und einer übermäßigen Abhängigkeit von der KI-Lösung liegen.

TRANSPARENZ & ERKLÄRBARKEIT (XAI)

Für interne Nutzer muss nachvollziehbar sein, warum eine E-Mail als dringlich eingestuft, einem bestimmten Berater zugewiesen oder mit bestimmten CRM-Daten verknüpft wurde. Bei Antwortentwürfen muss erkennbar sein, auf welchen Quellen und Datenpunkten diese beruhen und wo das System Unsicherheiten oder fehlende Informationen identifiziert hat. Die Priorisierungslogik sollte dokumentiert und für die Nutzer einsehbar sein, um systematische Verzerrungen erkennen zu können. Bei der Rechnungszuordnung sollte die Zuordnungslogik transparent dargestellt werden, damit Fehler schnell identifiziert und korrigiert werden können. Gegenüber externen Kommunikationspartnern ist zu prüfen, ob und in welcher Form über den KI-Einsatz informiert werden muss. Die Dokumentation der Entscheidungslogik dient auch der internen Qualitätssicherung und der Nachweisfähigkeit gegenüber Aufsichtsbehörden.

GUARDRAILS & KONTROLLMECHANISMEN

Kein automatischer Versand externer Kommunikation ist als technische Sperre zu implementieren, nicht nur als organisatorische Regel. Eine verbindliche menschliche Freigabe für alle Antwortentwürfe muss prozessual und technisch sichergestellt werden, wobei die Freigabe eine echte inhaltliche Prüfung erfordert und nicht zu einem reinen Klick-Prozess degradiert werden darf. Ein Rollen- und Berechtigungskonzept muss sicherstellen, dass Mitarbeiter nur auf E-Mails und Daten zugreifen können, die ihrem Zuständigkeitsbereich entsprechen, insbesondere bei projektbezogener Vertraulichkeit. Protokollierung und regelmäßige Stichprobenkontrollen von Zuweisungen, Extraktionen und Entwürfen sind erforderlich, um Fehler und Verzerrungen systematisch zu erkennen. Vertragliche und technische Begrenzungen gegen Training, Logging und Speicherung durch den Anbieter über den vereinbarten Zweck hinaus müssen durchgesetzt werden. Die Möglichkeit, besonders sensible Absender, Projekte oder Anhänge von der KI-Verarbeitung auszunehmen, muss als Konfigurationsoption verfügbar sein. Ein Fallback-Prozess für den manuellen Betrieb bei Systemausfall muss definiert und getestet werden. Eskalationswege für erkannte Fehler, Datenschutzvorfälle oder unerwartetes Systemverhalten müssen dokumentiert und allen Nutzern bekannt sein. Regelmäßige Reviews der Systemleistung, der Priorisierungslogik und der Fehlerquoten sollten mindestens quartalsweise erfolgen. Zugriffsprotokolle und Audit-Trails müssen für Nachweiszwecke gegenüber Aufsichtsbehörden und Kunden verfügbar sein.

Empfohlene Maßnahmen

PRIORITÄT 1 Sofort (0–7 Tage)

1. Auftragsverarbeitungsverträge, Leistungsbeschreibungen, technische und organisatorische Maßnahmen sowie Datenflussdokumentationen beider Anbieter anfordern und durch einen Datenschutzberater prüfen lassen.
2. Schriftliche Bestätigung beider Anbieter einholen, dass Kundendaten nicht für Training, Feinabstimmung oder sonstige Modellverbesserung verwendet werden, und diese Zusage vertraglich absichern.
3. Prüfen und dokumentieren, welche Daten tatsächlich an den jeweiligen Anbieter übermittelt werden, einschließlich vollständiger E-Mail-Inhalte, Anhänge, Metadaten und CRM-Zugriffe.
4. Sofortige Vertragsunterzeichnung aussetzen bis zum Abschluss der Due-Diligence-Prüfung und Vorlage belastbarer Anbieterunterlagen.

PRIORITÄT 2 Kurzfristig (7–30 Tage)

5. Vorprüfung zur Datenschutz-Folgenabschätzung durchführen und bei bestätigtem Erfordernis die vollständige DPIA vor Produktivbetrieb abschließen und dokumentieren.
6. Berechtigungs- und Freigabekonzept für E-Mail-Routing, Antwortentwürfe, CRM-Zugriffe und Rechnungszuordnung definieren und mit der Geschäftsleitung abstimmen.
7. Testphase mit realitätsnahen, aber kontrollierten Daten aufsetzen, einschließlich dokumentierter Qualitätsmessung anhand definierter KPIs und Abbruchkriterien.
8. System-Prompt mit Vertraulichkeitsregeln, Eskalationslogik, Tonalitätsvorgaben und Ausschluss verbindlicher Zusagen erstellen und freigeben.

PRIORITÄT 3 Mittelfristig (30–90 Tage)

9. Regelmäßiges Qualitätsmonitoring für Fehlrouting, Entwurfsfehler, Priorisierungsverzerrungen und Rechnungszuordnung etablieren, mindestens quartalsweise mit dokumentierten Ergebnissen.
10. Schulung aller betroffenen Mitarbeiter zu Vertraulichkeit, Freigabeprozessen, Erkennung von Halluzinationen und Fehlern, Berechtigungen und Ausfallverfahren durchführen und dokumentieren.
11. Informations- und Transparenzpflichten gegenüber Betroffenen und Kunden auf Basis der finalen Verarbeitungsarchitektur prüfen und umsetzen, einschließlich Anpassung der Datenschutzhinweise.
12. Vertraulichkeitsklauseln mit sensiblen Großkunden auf Vereinbarkeit mit der KI-Verarbeitung prüfen und gegebenenfalls Ausnahmen oder Einwilligungen einholen.

Empfehlungen

ZUSAMMENFASSUNG

Die sofortige Vertragsunterzeichnung ist auszusetzen, bis belastbare Unterlagen und Antworten beider Anbieter vorliegen. Im ersten Schritt sind Auftragsverarbeitungsverträge, technische Dokumentationen und schriftliche Bestätigungen zur Datenverwendung einzuholen und durch einen Datenschutzberater zu prüfen. Parallel sollte die Vorprüfung zur Datenschutz-Folgenabschätzung eingeleitet werden. Nach Abschluss dieser Prüfungen kann eine fundierte Anbieterentscheidung getroffen und bei positivem Ergebnis eine Pilotphase mit kontrollierten Daten und definierten Erfolgskriterien aufgesetzt werden. Die Geschäftsleitung sollte zeitnah über die Einbindung externer Datenschutz- und IT-Sicherheitsunterstützung entscheiden, um die Prüfung fachlich abzusichern.

Klärungsbedarf

RECHTLICHE FRAGEN

1. Welche konkreten Regelungen enthalten die Auftragsverarbeitungsverträge beider Anbieter zu Zweckbindung, Unterauftragsverarbeitern, Löschfristen und Ausschluss der Datenverwendung für Training oder Modellverbesserung?
2. Bestehen bei Anbieter 1 oder dessen Unterauftragnehmern, insbesondere OpenAI und AWS, Drittlandtransfers in die USA, und auf welcher Rechtsgrundlage erfolgen diese, einschließlich Standardvertragsklauseln und Transfer Impact Assessment?
3. Wer haftet rechtlich bei fehlerhaften KI-Antworten, Fehlzuweisungen oder Datenschutzvorfällen, die zu Schäden bei Kunden oder Betroffenen führen, und wie ist die Haftungsverteilung vertraglich geregelt?
4. Ist angesichts der systematischen Verarbeitung personenbezogener Daten die Benennung eines Datenschutzbeauftragten nach Art. 37 DSGVO erforderlich?

TECHNISCHE FRAGEN

1. Welche konkreten Modelle, Subdienstleister, Hosting-Komponenten und Datenverarbeitungspfade werden von Anbieter 2 eingesetzt, und wie ist die Modellarchitektur dokumentiert?
2. Werden vollständige E-Mails einschließlich aller Anhänge an den Anbieter übermittelt und dort verarbeitet, oder erfolgt eine Vorfilterung, Anonymisierung oder Beschränkung auf Metadaten und ausgewählte Inhalte?
3. Gibt es bei beiden Anbietern Konfigurationsmöglichkeiten, um sensible Absender, Projekte, Anhangtypen oder Schlüsselwörter von der KI-Verarbeitung auszunehmen?
4. Welche Logging-, Monitoring-, Verschlüsselungs- und Incident-Response-Mechanismen sind bei beiden Anbietern implementiert und wie werden diese dokumentiert?

ORGANISATORISCHE FRAGEN

1. Wie wird das Risiko einer systematischen Verzerrung bei der Priorisierung von E-Mails organisatorisch überwacht und durch welche Kontrollmechanismen werden Benachteiligungen erkannt?
2. Wie wird verhindert, dass Mitarbeiter Antwortentwürfe unter Zeitdruck ungeprüft freigeben, und welche prozessualen Sicherungen stellen eine echte inhaltliche Prüfung sicher?
3. Wie wird bei Ausfall der KI-Lösung der manuelle Betrieb sichergestellt, und ist ein dokumentiertes Ausfallverfahren mit klaren Verantwortlichkeiten vorhanden?
4. Wie werden bestehende Vertraulichkeitsvereinbarungen mit sensiblen Großkunden auf Vereinbarkeit mit der KI-gestützten E-Mail-Verarbeitung geprüft, und wer verantwortet diese Prüfung?

Rechtliche Hinweise

Haftungsausschluss: Diese Analyse dient der betriebswirtschaftlichen, technischen und ethischen Bewertung. Sie stellt keine Rechtsberatung dar und ersetzt keine juristische Einzelfallprüfung. Optional kann ergänzend eine juristische Detailprüfung mit separatem rechtlichen Dokument durch spezialisierte Rechtsanwälte für KI-, Datenschutz- und Technologierecht gegen Mehrkosten beauftragt werden.

Vertraulichkeit: Dieses Dokument enthält vertrauliche Informationen und ist ausschließlich für den Empfänger bestimmt.

Anton Bauer

Zertifizierter KI-Beauftragter
ElevAgent AI

ElevAgent AI

ElevAgent AI · office@elevagent.eu · www.elevagent.eu

Server: Frankfurt (EU) · DSGVO-konform · Zertifizierter KI-Beauftragter nach ÖVE/ÖNORM EN ISO/IEC 17024