

GEM Cortex AI

– PRODUKT- UND VERTRIEBSSPEZIFIKATION

Grundprinzip von GEM Cortex AI: Sie erwerben weder ein KI-Modell noch einen Chat. Sie erwerben einen messbar besseren Prozess, einen sicheren Zugang zu KI und die Fähigkeit, Qualität, Kosten und Verantwortlichkeiten im Betrieb zu steuern.



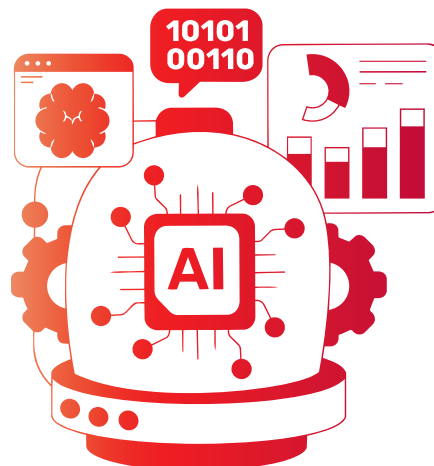
Management Summary zu GEM Cortex AI

GEM Cortex AI ist eine Produktfamilie und ein Implementierungsrahmen für Organisationen, die KI in reale Prozesse, Daten und Systeme einbinden und dabei die Kontrolle über Identitäten, Kosten, Qualität und Verantwortlichkeiten behalten möchten. Ausgangspunkt ist nicht die Frage, welcher Chatbot gekauft werden soll. Ausgangspunkt sind die Aufgabe, die für das Ergebnis verantwortliche Person, die maßgebliche Informationsquelle und eine sichere Übergabestelle, an der die KI ihre Arbeit an einen Menschen oder ein Unternehmenssystem übergibt.

- ✓ Integriert KI in bestehende Anwendungen, Dokumentenflüsse, Wissensquellen und Workflows.
- ✓ Wählt zwischen einem präzisen Verarbeitungspfad ohne LLM, einer Antwort auf Basis unternehmensinterner Quellen, einem allgemeinen Modell und einem kontrollierten Werkzeug.
- ✓ Vereinheitlicht die Unternehmensidentität, Projekte, Modelle, Zugänge und Kostenverantwortung.
- ✓ Misst das akzeptierte Geschäftsergebnis, kritische Fehler, das betriebliche Leistungsniveau und die Kosten je Einheit.
- ✓ Ermöglicht den Einstieg mit einem eng begrenzten Pilotprojekt und den Ausbau einer gemeinsamen Steuerung, ohne vorzugeben, dass jedes Pilotprojekt automatisch ein Produkt ist.

Grundlegende Positionierung

Kontrollierte KI für den realen Betrieb. GEM Cortex AI verbindet Modelle mit Unternehmensdaten, Identität und Prozessen, damit der Kunde KI sicher nutzen, messen und weiterentwickeln kann – ohne Abhängigkeit von einer einzigen Benutzeroberfläche oder einem einzelnen Anwendungsfall.



1. Warum GEM Cortex AI entsteht

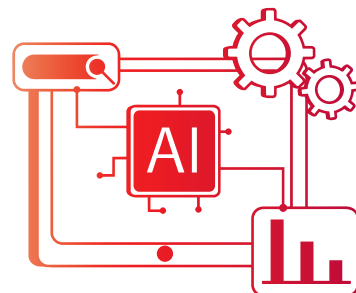
In den meisten Organisationen ist die fehlende Verfügbarkeit eines Modells nicht das Hauptproblem. Modelle sind sowohl als Dienste für Endnutzer als auch über APIs verfügbar. Das Problem entsteht beim Übergang in den Betrieb: Interne Daten unterliegen Berechtigungen und haben Verantwortliche, Anwendungen benötigen stabile Schnittstellen, die Finanzabteilung möchte Kosten einem Ergebnis zuordnen, und die Sicherheitsverantwortlichen müssen wissen, wohin Daten übermittelt werden und wer eine Aktion auslösen kann.

- ✓ **Fragmentierte Nutzung:** Teams erwerben unterschiedliche Lizenzen und API-Zugänge ohne gemeinsame Zuständigkeit. Cortex bietet eine Bestandsübersicht, Identitäten, Anbieter-/Projektregister und einheitliche Regeln.
- ✓ **Pilotprojekt ohne Prozess:** Die Demo liefert Antworten, doch es ist unklar, wer die Ergebnisse freigibt und wohin sie geschrieben werden. Cortex bietet Workflows, Kontrollpunkte, Integrationen und Abnahmeverfahren.
- ✓ **Daten ohne Kontext:** Der Chatbot kennt die Unternehmensrollen, die gültigen Versionsstände und die maßgebliche Informationsquelle nicht. Cortex bietet Knowledge unter Berücksichtigung von Zugriffsberechtigungen, Metadaten und Quellenangaben.
- ✓ **KI-Kosten ohne Wertbeitrag:** Die Rechnung weist Token oder Lizenzen aus, jedoch keine abgeschlossenen Aufgaben. Cortex bietet AI FinOps, Kostenzuordnung und ein Value Ledger.
- ✓ **Abhängigkeit von einem einzigen Anbieter:** UI, Daten, Prompts und Workflows werden an einen einzigen Dienst gebunden. Cortex bietet Adapter, Exporte und einen bewusst gesteuerten Lock-in.
- ✓ **Bedenken hinsichtlich der Autonomie:** Die Organisation befürchtet, dass ein Agent eine unerwünschte Aktion ausführt. Cortex bietet minimale Berechtigungsumfänge, Vorschauen, menschliche Freigaben, Idempotenz und Audit.

1.1 | Was Cortex ist – und was es nicht ist

Cortex kombiniert wiederholt einsetzbare Produktfunktionen, Integrationskomponenten und Dienstleistungen von GEM für Konzeption, Implementierung und Betrieb. Jede Bereitstellung wird entsprechend dem Anwendungsfall konfiguriert. Einige Komponenten können separat bereitgestellt werden; den größten Mehrwert bieten sie, wenn sie Identitäten, Richtlinien, Telemetrie und Evaluierungsverfahren gemeinsam nutzen.

- **Es handelt sich nicht um ein einziges proprietäres Foundation Model von GEM.** Cortex kann je nach Aufgabe und Datenprofil mehrere kommerzielle oder lokale Modelle verwenden.
- **Es ersetzt nicht sämtliche KI-Werkzeuge des Kunden.** Es kann diese um eine gemeinsame Verwaltung ergänzen oder Prozesse abdecken, die eine SaaS-Standardlizenz nicht umfasst.
- **Es handelt sich nicht um einen universellen autonomen Agenten.** Der Autonomiegrad wird entsprechend den Auswirkungen gewählt und kann auf die reine Erstellung von Vorschlägen beschränkt bleiben.
- **Es handelt sich nicht um eine automatische Compliance-Zertifizierung.** Cortex schafft Nachweise und Kontrollen; die Verantwortung und die rechtliche Bewertung verbleiben jedoch bei den jeweils zuständigen Rollen.



2. Idealer Kunde und Qualifizierung der Geschäftschance

Der ideale Kunde für Cortex hat nicht nur das allgemeine Ziel, KI einzusetzen. Er hat einen konkreten Prozess oder ein Steuerungsproblem, eine für das Ergebnis verantwortliche Person und die Bereitschaft, repräsentative Daten sowie die aktuelle Baseline bereitzustellen.

- ✓ **Großunternehmen und regulierte Organisationen:** Secure Workspace, Knowledge, Document Intelligence, Control.
- ✓ **Öffentliche Verwaltung und Organisationen der öffentlichen Daseinsvorsorge:** Klassifizierung von Eingaben, Dokumentenworkflows, Knowledge in einer eingeschränkten Umgebung.
- ✓ **IT- und Shared-Service-Organisationen:** Access & Spend, Service Desk, Code Discovery, internes Knowledge.
- ✓ **Unternehmen mit hohem Dokumentenaufkommen:** Document Intelligence, RAG mit Quellenangaben, Flow und Integrationen.
- ✓ **Organisationen nach der ersten Welle von KI-Pilotprojekten:** Control, Überführung erfolgreicher Anwendungsfälle in eine gemeinsame Plattform und Beendigung schwacher Anwendungsfälle.

2.1 | Merkmale eines geeigneten ersten Anwendungsfalls

- Die Aufgabe wiederholt sich, hat klar erkennbare Eingaben und Ergebnisse und beansprucht bereits heute messbar Zeit oder verursacht Risiken.
- Es gibt eine prozessverantwortliche Person und eine qualifizierte prüfende Person, die ein korrektes Ergebnis bestimmen kann.
- Der Einstieg ist im Read-only-Modus, im reinen Entwurfsmodus oder mit begrenztem Umfang ohne irreversible Auswirkungen möglich.
- Ein repräsentativer historischer Datensatz ist verfügbar und kann rechtmäßig und sicher für die Evaluierung verwendet werden.
- Das Ergebnis lässt sich in eine bestehende Anwendung oder einen bestehenden Workflow einbinden, statt einen weiteren isolierten Posteingang zu schaffen.

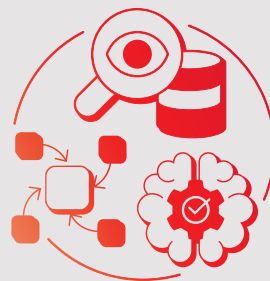
2.2 | Wann eine Geschäftschance nicht weiterverfolgt oder neu formuliert werden sollte

- Der Kunde verlangt eine garantierte hundertprozentige Genauigkeit generativer Ergebnisse ohne Kontrollpunkt.
- Weder die für den Prozess verantwortliche Person noch die maßgebliche Informationsquelle noch das Verfahren zur Erkennung eines korrekten Ergebnisses sind bekannt.
- Das einzige Ziel besteht darin, die Technologie auszuprobieren, ohne eine Nutzenhypothese und ohne eine Entscheidung nach dem Pilotprojekt.
- Eine vorhandene Funktion eines bereits lizenzierten SaaS-Dienstes deckt den Bedarf ausreichend, sicher und kostengünstiger ab.
- Für die benötigten Daten oder die automatische Aktion fehlen ein genehmigter Zweck, eine genehmigte Zugriffsberechtigung oder eine verantwortliche Person.

3. Die Produktfamilie GEM Cortex AI

Die nachstehenden Bezeichnungen beschreiben Funktionsbausteine, die zu einer Lösung kombiniert werden können. Nicht jeder Baustein muss in jedem Projekt ein separat lizenziertes Produkt sein. Für den Kunden ist entscheidender, welches Problem der Baustein löst, welche Systeme er anbindet und welchen Nachweis er liefert.

- ✓ **Cortex Core:** Orchestrierung von Intention, Laufzeitpfad, Modell und Mandantenkontext. Ein zentraler, kontrollierter Zugangspunkt anstelle von Anwendungscode, der fest an ein einziges Modell gebunden ist.
- ✓ **Cortex Access:** SSO, Rollen, Workload-Identitäten, Anbieterkonten, Projekte und Lebenszyklus. Weniger gemeinsam genutzte Schlüssel, verwaiste Zugänge und unklare Zuständigkeiten.
- ✓ **Cortex Knowledge:** Konnektoren, Datenaufnahme, Retrieval unter Berücksichtigung von ACL, Quellenangaben und Lebenszyklus des Index. Antworten auf Basis freigegebener Quellen mit nachvollziehbarer Grundlage.
- ✓ **Cortex Flow:** Workflows, Werkzeuge, Freigaben, menschliche Aufgaben und Prozessstatus. Ein KI-Entwurf wird sicher zu einem Schritt in einem bestehenden Prozess.
- ✓ **Cortex Control:** Richtlinien, KI-Register, Telemetrie, Evaluierung, Budgets und FinOps. Gemeinsame Steuerung von Qualität, Risiken und Kosten über verschiedene Anwendungsfälle hinweg.
- ✓ **Cortex Workspace:** Benutzerkanal für freigegebene Chat-, Wissens- und Arbeitsaufgaben. Eine praxistaugliche Unternehmensalternative zur unkontrollierten KI-Nutzung.
- ✓ **Cortex Document Intelligence:** OCR, Klassifizierung, Extraktion, Normalisierung und strukturierter Entwurf. Weniger manuelle Übertragung und schnellere Dokumentenprozesse bei gleichbleibender Kontrolle.



4. Angebotspakete und Einstiegsprodukte

Die Pakete geben dem Vertriebsgespräch einen klaren Einstieg. Jedes Paket hat einen konkreten Kaufanlass, ein erstes greifbares Ergebnis und bewusst definierte Grenzen. Der endgültige Umfang und der Preis werden nach der Qualifizierung von Daten, Integrationen, Volumen und Betriebsmodell festgelegt.

- ✓ **Access & Spend Baseline:** CIO/CISO/CFO haben keinen Überblick darüber, wer für KI-Konten, Projekte und Kosten verantwortlich ist. Erste Lieferung: Bestandsaufnahme im Read-only-Modus, Übersicht der Lücken, 90-Tage-Baseline. Transparenz vor Automatisierung.
- ✓ **Secure AI Workspace:** Das Unternehmen möchte seinen Beschäftigten freigegebene KI mit SSO anbieten. Erste Lieferung: Workspace für eine Gruppe und drei Aufgaben. Kontrollierte Einführung und Audit.
- ✓ **Knowledge Assistant:** Mitarbeitende suchen nach Verfahrensanweisungen und wissen nicht, welche Version gültig ist. Erste Lieferung: kuratierte Wissensdomäne, Quellenangaben, ACL-Tests. Schnelleres Auffinden der Quelle.
- ✓ **Document Intelligence:** Das Team klassifiziert oder überträgt Dokumente manuell. Erste Lieferung: eine Klasse, Ausgabeschema, Entwurf. Weniger mechanische Arbeit.
- ✓ **RFP Analyzer:** Das Angebots-/Einkaufsteam prüft umfangreiche Bedingungen. Erste Lieferung: Matrix der Anforderungen, Quellen, Verantwortlichen und Risiken. Geringeres Risiko, etwas zu übersehen.
- ✓ **Code Discovery & Modernization:** Ein Legacy-System weist verborgene Abhängigkeiten auf. Erste Lieferung: Modulübersicht im Read-only-Modus, Test-Backlog. Verringert die Unsicherheit vor einer Migration.

5. Wann Cortex andere Ansätze ergänzt oder ersetzt

Cortex ist kein Argument gegen fertige KI-Dienste. Eine geeignete Konzeption kann deren Modelle, Schnittstellen und Lizenzen nutzen. Der Mehrwert von Cortex steigt, wenn der Kunde die Anbindung eines eigenen Prozesses oder mehrerer Anbieter, eine gemeinsame Identität, messbare Qualität oder Kostenkontrolle benötigt.

- ✓ **SaaS-Assistent / Copilot:** Stark bei Standardfunktionen. Cortex ist sinnvoll für eigene Workflows, anbieterunabhängige Kontrolle und gemeinsames FinOps. Wir empfehlen Cortex als Ergänzung der Verwaltung, nicht als unbegründeten Ersatz.
- ✓ **Direkte Modell-API:** Stark für ein einzelnes Team und eine klar abgegrenzte Anwendung. Cortex ist sinnvoll für mehrere Anwendungen, bei denen ein Gateway und eine Steuerungsebene benötigt werden.
- ✓ **Individueller Chatbot:** Stark als einfacher Kommunikationskanal. Cortex ist dort sinnvoll, wo ACL, Quellenangaben, Integrationen und Verantwortung für den Betrieb erforderlich sind.
- ✓ **RPA/Workflow:** Stark bei deterministischen Schritten. Cortex ist sinnvoll für unstrukturierte Eingaben, die Sprachverständnis erfordern.
- ✓ **Agentenplattform eines einzelnen Anbieters:** Stark durch umfangreiche Funktionen in einer bestimmten Cloud. Cortex ist sinnvoll für hybride Bereitstellungen.

Faire Empfehlung: Wenn die Standardfunktion eines bereits lizenzierten Dienstes den Bedarf, die Anforderungen an den Umgang mit Daten und die wirtschaftlichen Anforderungen erfüllt, muss kein eigener Cortex-Anwendungsfall entwickelt werden. GEM kann bei Auswahl, Integration und Governance unterstützen.

6. Mehrwert für die einzelnen Rollen im Beschaffungsprozess

Dieselbe Plattform hat für jede Rolle eine andere Bedeutung. Die Vertriebskommunikation sollte beim Problem der jeweiligen Rolle ansetzen und erst danach die Technologie erläutern.

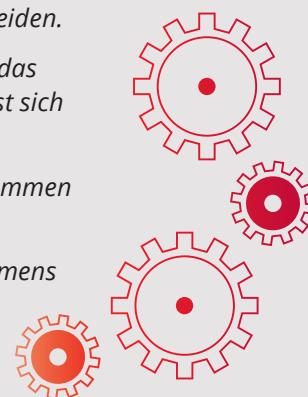
- ✓ **CEO / Unternehmensleitung:** KI-Investitionen werden zu einem gesteuerten Portfolio von Anwendungsfällen mit Verantwortlichen und messbaren Ergebnissen.
Frage: Welche drei Prozesse bieten den größten Mehrwert?
- ✓ **CIO / CTO:** Gemeinsam genutzte Integrations- und Betriebsfunktionen statt einer Reihe isolierter Prototypen.
Frage: Was lässt sich aus dem ersten Anwendungsfall wiederverwenden?
- ✓ **CISO / DPO:** Identitäten, Datenprofile, Berechtigungsumfänge, Audit und kontrollierte Aktionen.
Frage: Welche Daten und Werkzeuge sind für den jeweiligen Zweck zugelassen?
- ✓ **CFO / Einkauf:** Kosten nach Verantwortlichen und Ergebnissen, Vergleich von Lizenzen, API und Betrieb.
Frage: Besteht die Einsparung in freier Kapazität, einer zahlungswirksamen Entlastung oder einem geringeren Risiko?
- ✓ **Prozessverantwortliche:** Schnellere Aufgabenerledigung und weniger Nacharbeit ohne Verlust der Entscheidungsbefugnis.
Frage: Wo soll KI einen Entwurf erstellen, und wo muss ein Mensch die Freigabe erteilen?

7. Wirtschaftlicher Nutzen und TCO

Das wirtschaftliche Modell von Cortex trennt technischen Verbrauch, Prozesseffekt und realisierten Wert. Die Berechnung beginnt mit einer Baseline und endet bei der Frage, wie die Organisation die veränderte Kapazität, den Durchsatz, die Qualität oder das Risikoprofil tatsächlich nutzt.

Arbeitsgleichung: Realisierter jährlicher Wert = genutzter Wert freigesetzter Kapazität + Beitrag eines höheren Durchsatzes + entfallene externe Kosten + erwartete Verlustminderung – wiederkehrende TCO.

- ✓ **Baseline:** Volumen, Zeit pro Fall, Fehlerquote, Nacharbeit, externe Kosten und aktuelle SLA. Ohne Baseline lässt sich der KI-Effekt nicht von einer gewöhnlichen Prozessänderung unterscheiden.
- ✓ **Nutzung und Eignung:** Wie viele Fälle für die Lösung geeignet sind und wie viele Nutzer das Verfahren tatsächlich anwenden. Die unter Laborbedingungen erreichte Genauigkeit lässt sich nicht automatisch auf das gesamte Volumen hochrechnen.
- ✓ **Akzeptanz des Ergebnisses:** Anteil der Entwürfe, die ohne inhaltliche Korrektur angenommen werden. Ein besserer Maßstab als die Zahl der generierten Antworten.
- ✓ **Nutzung freigesetzter Kapazität:** Umverteilung der Zeit zugunsten eines höheren Volumens oder Reduzierung von Überstunden. Eine eingesparte Minute allein ist noch keine zahlungswirksame Einsparung.
- ✓ **Risiko:** Wahrscheinlichkeit und Auswirkungen eines Fehlers vor und nach Anwendung der Kontrollen. Nicht als sichere Einsparung verkaufen; Szenario und Sensitivität angeben.



Kennzahlen je Einheit: Kosten je akzeptiertem Entwurf statt je Token; Kosten je Dokument mit korrekt extrahierten Daten; Kosten je aktivem Nutzer, der eine Aufgabe abgeschlossen hat; Differenz zwischen interner Telemetrie und Rechnung.

8. Geschäfts-, Implementierungs- und Betriebsmodell

Das Geschäftsmodell folgt dem schrittweisen Abbau von Unsicherheit. Der Kunde erwirbt keine unbestimmte Transformation, sondern eine durch Daten gestützte Entscheidung.

- ✓ **Qualification workshop:** Problem, Sponsor, Daten, Risiko, Alternative. Ergebnis: Opportunity brief.
- ✓ **Discovery & Pilot Design:** Baseline, Lösungsskizze, Evaluierung, Sicherheit, TCO. Ergebnis: Pilot Charter.
- ✓ **Pilot:** Begrenzte Implementierung, historische Daten, Entwurfs-/Read-only-Modus. Ergebnis: Go / adjust / stop report.
- ✓ **Limited production:** Reale Identitäten und reales Volumen, Integration, Support. Ergebnis: Production readiness.
- ✓ **Scale & Managed Service:** Weitere Teams/Anwendungsfälle, Kapazität, FinOps, SLA.

Preiskomponenten: Einmalige Implementierung (Discovery, Konzeption, Tests), Plattformdienst (Core/Access/Control), Verbrauch (Modell/API, Speicher, Rechenleistung), Betrieb und Support, kontinuierliche Qualitätssicherung und Governance, Änderungsanforderungen

9. Access & Spend – eigenständiges Einstiegsangebot

Ein geeigneter Einstieg auch für Kunden, die noch keine eigene KI-Anwendung entwickeln möchten. Schafft eine gemeinsame Datengrundlage für CIO, CISO, Finanzabteilung und Einkauf. Zeigt auf, welche Kosten auf Nutzerlizenzen und welche auf API-Verbrauch entfallen, wer für die Projekte verantwortlich ist und wo der Bezug zum Geschäftsergebnis fehlt.

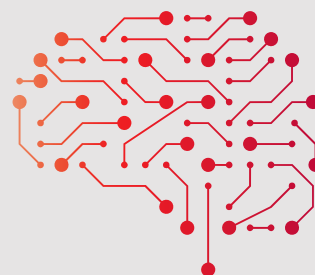
- ✓ **Baseline:** Erfassung von Verzeichnis-, Anbieterprojekt-, Nutzungs- und Abrechnungsdaten im Read-only-Modus. Bestandsübersicht der Verantwortlichen, Zugänge und nicht zugeordneten Kosten.
- ✓ **Control Pilot:** JML-Prozess, Rezertifizierung, Tags, Budgets und Showback für einen begrenzten Umfang. Erster kontrollierter Betrieb.
- ✓ **Governed Scale:** Mehrere Anbieter/Teams, Regeln für Modelle, Chargeback und Value Ledger.

Realistische Grenzen: Cortex konsolidiert keine persönlichen Passwörter und umgeht keine Lizenzbedingungen. Die automatische Erstellung/Entfernung eines Kontos ist nur dort möglich, wo der Anbieter eine unterstützte API/SCIM-Schnittstelle bereitstellt. Ein hartes Budgetlimit gilt nur für den Verbrauch, der über das Cortex-Gateway läuft.

10. Message House und Vertriebsnarrativ

Kernbotschaft: GEM Cortex AI – kontrollierte KI für den realen Betrieb. Wir machen aus einem Modell einen sicher integrierten Dienst: mit Unternehmensidentität, Quellen, auf die der Zugriff zulässig ist, messbarer Qualität, Kostenkontrolle und einem klar definierten Punkt, an dem der Mensch entscheidet.

- ✓ **Säule 1 – Integration:** KI wird in bestehende Workflows, APIs und die Datenebene eingebunden. Das Ergebnis wird zu einem kontrollierten Prozessschritt.
- ✓ **Säule 2 – Kontrolle:** Identitäten, Richtlinien, Quellenangaben, Freigaben, Audit und Werkzeuge mit Einschränkungen. Die Organisation weiß, wer was genutzt hat.
- ✓ **Säule 3 – Wahlfreiheit:** Direct/no-LLM, RAG, ein allgemeines Modell, eine lokale Laufzeitumgebung oder ein Werkzeug je nach Aufgabe.
- ✓ **Säule 4 – Wirtschaftlichkeit:** Verbrauch mit Bezug zum Anwendungsfall und zum akzeptierten Ergebnis. CIO und Finanzabteilung entscheiden über die Skalierung.



10.1 | Dreißigsekündiger Pitch

GEM Cortex AI unterstützt Unternehmen und öffentliche Organisationen dabei, KI aus isolierten Experimenten in einen kontrollierten Betrieb zu überführen. Es verbindet Modelle mit der Unternehmensidentität, Wissensbeständen, für die eine Zugriffsberechtigung besteht, und bestehenden Prozessen. Je nach Aufgabe wählt es einen präzisen Verarbeitungspfad ohne LLM, eine Antwort mit Quellenangaben oder einen kontrollierten Workflow-Schritt. Gleichzeitig misst es Qualität, Kosten und Akzeptanz des Ergebnisses. Wir beginnen mit einem eng begrenzten Pilotprojekt, einer klaren Baseline und einer Entscheidung über go, adjust oder stop.



10.2 | Zweiminütiger Pitch

Der bloße Zugang zu einem Modell ist heute kein Wettbewerbsvorteil. Die Herausforderung besteht darin, sicherzustellen, dass KI nur die richtigen Daten einsehen darf, das geeignete Werkzeug verwendet, das Ergebnis in einen realen Prozess einfließt und die Organisation sowohl Qualität als auch Wirtschaftlichkeit nachweisen kann. GEM Cortex AI schließt diese Lücke mit einer gemeinsamen Ebene aus Core, Access, Knowledge, Flow und Control. Der Kunde muss Identitäten, Audit, Modelladapter und Kostenmessung nicht für jeden Anwendungsfall neu aufbauen. Im ersten Projekt liefern wir ausschließlich die Komponenten, die für das konkrete Ergebnis erforderlich sind; erfolgreiche Bestandteile lassen sich in weiteren Domänen wiederverwenden. Bestätigt sich die Hypothese im Pilotprojekt nicht, verfügt das Projekt über einen vorab definierten Abbruchpunkt.

11. Qualifizierungsfragen und Umgang mit Einwänden

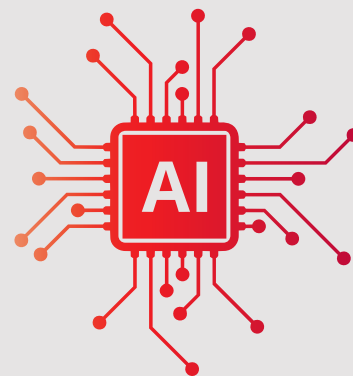
11.1 | Typische Einwände

- **„Wir haben bereits Microsoft Copilot oder einen anderen Enterprise-Assistenten.“** Das kann für Standardaufgaben die richtige Wahl sein. Wir prüfen mögliche Lücken: eigene Prozesse, weitere Modelle, Wissen unter Berücksichtigung von Zugriffsberechtigungen und gemeinsames FinOps. Besteht keine Lücke, empfehlen wir Cortex nicht.
- **„Uns genügt es, die API des Modells aufzurufen.“** Für einen klar abgegrenzten Prototyp ist das ausreichend. Cortex wird wirtschaftlich sinnvoll, wenn Identitäten, Evaluierung, Audit und Budget über mehrere Anwendungen hinweg gemeinsam genutzt werden sollen.
- **„Wir möchten ein einziges Modell, und zwar das beste.“** Welches Modell das beste ist, hängt von Sprache, Aufgabe und den Rahmenbedingungen der Datenverarbeitung ab. Register und Evaluierungen ermöglichen es, ein Standardmodell auszuwählen und eine Änderung sicher zu bewerten.
- **„Wir benötigen 100% Genauigkeit.“** Generative Ergebnisse bieten keine solche allgemeine Garantie. Kritische Felder lassen sich durch ein Schema und Validierung begrenzen; risikobehaftete Entscheidungen verbleiben beim Menschen.
- **„Wir verlangen Zero Data Retention.“** Wir prüfen das konkrete Produkt, die Region und die vertraglichen Rahmenbedingungen. ZDR lässt sich nicht aus dem Namen des Anbieters ableiten.
- **„Alles muss On-Premises laufen.“** Das ist für geeignete Modelle möglich, erfordert jedoch Hardware, MLOps und eine Prüfung der Lizenzbedingungen. Wir vergleichen die TCO hybrider und lokaler Bereitstellungen.
- **„MCP löst die Integrationsfragen.“** MCP kann den Schnittstellenvertrag für Quellen und Werkzeuge standardisieren, ersetzt jedoch weder Datenmapping noch Berechtigungen noch Sicherheitsrichtlinien.
- **„Das Pilotprojekt hat funktioniert, warum ist es nicht sofort im Produktivbetrieb?“** Ein Pilotprojekt überprüft eine Hypothese. Der Produktivbetrieb benötigt reale Identitäten, SLO, Incident-Management, Wiederherstellung und vertragliche Verantwortlichkeiten.
- **„Wie garantieren Sie den ROI?“** Garantiert werden können das Messverfahren und die bereitgestellten Funktionen, nicht jedoch die künftige Nutzung der Kapazität durch die Organisation.

12. Beschaffungs- und Nachweispaket

Ein Unternehmenskunde erwirbt nicht nur Funktionalität. Er benötigt den Nachweis, dass die Lösung sicher übernommen, betrieben, kontrolliert und bei Bedarf wieder aufgegeben werden kann. Presales erstellt ein strukturiertes Nachweispaket, dessen Umfang der jeweiligen Projektphase entspricht.

- ✓ **Produkt und Umfang:** Anwendungsfall, einbezogene/ausgeschlossene Bestandteile, Kanäle, Daten, Modelle, Werkzeuge.
- ✓ **Architektur:** Logisches Diagramm und Bereitstellungsdiagramm, Laufzeitpfade.
- ✓ **Daten und Datenschutz:** Datenfluss, Klassifizierung, Region, Aufbewahrung, Löschung.
- ✓ **Sicherheit:** Bedrohungsmodell, Identitäten, Tests und Schnittstellen für Sicherheitsvorfälle.
- ✓ **Qualität:** Evaluierungsdatensatz, Kennzahlen, kritische Fehler und Freigabekriterien für Releases.
- ✓ **Betrieb:** SLO/SLA, Kapazität, RACI, Sicherung/Wiederherstellung.
- ✓ **Wirtschaftlichkeit:** Implementierung, Plattform, Verbrauch, Value Ledger.
- ✓ **Exit:** Exporte, Übergabe der Konfiguration, bekannter Lock-in.



13. Wie ein qualifizierter Vertriebsdialog beginnt

Der wirkungsvollste erste Schritt ist keine allgemeine Präsentation, sondern ein kurzes Qualifizierungsgespräch zu einem konkreten Prozess oder einer Übersicht der KI-Kosten. GEM beschreibt gemeinsam mit dem Kunden den Ist-Zustand, legt einen sicheren Rahmen für das Pilotprojekt fest und bereitet Nachweise für die Entscheidung vor.

- 1 | **Wir bestimmen die anstehende Entscheidung:** Eine Aufgabe oder ein Steuerungsproblem.
- 2 | **Wir ermitteln die Baseline:** Volumen, Zeit, Qualität, Nacharbeit.
- 3 | **Wir konzipieren das erste Pilotprojekt:** Begrenzter Umfang, Evaluierung, Daten, Kontrollpunkte.
- 4 | **Wir begründen go/adjust/stop:** Ergebnis im Vergleich zur Baseline, Sicherheit und Kosten je Einheit.

Empfohlene Handlungsaufforderung: Beginnen wir mit einem Prozess oder einer 90-Tage-Übersicht der KI-Zugänge und -Kosten. Wir gehen mit einer konkreten Hypothese in die Discovery-Phase und verlassen sie mit Entscheidungsgrundlagen.