

GEM CORTEX AI

– KATALOG DER ANWENDUNGSFÄLLE UND VERTRIEBSSZENARIEN

Grundprinzip von GEM Cortex AI: Wir verkaufen kein allgemeines AI-Versprechen. Bei jedem Szenario zeigen wir, wer die Lösung nutzt, was sich im Arbeitsalltag dieser Person verändert, wo der Nutzen entsteht und welche Kontrollpunkte vorgesehen sind.



Fünf Merkmale eines geeigneten ersten Anwendungsfalls

1. Der Prozess hat einen klar benannten Verantwortlichen und einen Nutzer, der die derzeitige Arbeitsweise beschreiben kann.
2. Es gibt eine wiederkehrende Aufgabe, ein wiederkehrendes Dokument oder eine wiederkehrende Entscheidungsgrundlage mit messbarem Aufkommen.
3. Die Daten sind verfügbar, und die Zugriffsberechtigungen lassen sich rechtmäßig festlegen.
4. AI kann zunächst als Unterstützung oder zur Erstellung von Entwürfen mit menschlicher Freigabe eingesetzt werden; das Risiko ist begrenzt.
5. Vor dem Start lassen sich eine Abnahmemetrik und eine Entscheidung über stop/scale definieren.



Überblick über die wichtigsten Szenarien

1. **AI-Backend zur Vorverarbeitung von Versicherungsdokumenten:** Zeit bis zum fertigen Entwurf. Eine Schadenart, 100-300 Fälle.
2. **Interner Wissensassistent:** Recall@k. Eine Gruppe, 500-2000 Dokumente.
3. **RFP Analyzer:** Recall kritischer Anforderungen. Eine Ausschreibung + Holdout.
4. **Technische Dokumentation in einer eingeschränkten Umgebung:** Anteil korrekter Revisionsstände. Ein Gerätetyp.
5. **CV Converter:** Feldgenauigkeit. 50-150 CVs, 10-20 Felder.
6. **Access & Spend Baseline:** Inventarabdeckung. 1-2 Anbieter, Daten aus 90 Tagen.
7. **Sicherer AI Workspace für Unternehmen:** Aktive Nutzer. Eine Gruppe, drei Aufgaben.
8. **Modernisierung von Legacy-Software:** Abdeckung kritischer Pfade. Ein Modul.
9. **Öffentliche Verwaltung – Klassifizierung von Eingaben:** Klassifizierungsgenauigkeit. Ein Aufgabenbereich.
10. **BI-Fragen in natürlicher Sprache:** Korrektheit der Kennzahl. 10-20 Kennzahlen.
 - **Recall@k:** Prüft, ob die richtige Quelle unter den ersten k Ergebnissen erscheint.
 - **False negative:** Eine vom System übersehene kritische Anforderung oder ein übersehenes Risiko.
 - **Abstention:** Kontrollierter Verzicht auf eine Antwort bei unzureichenden Belegen oder Berechtigungen.
 - **Holdout:** Ein separater Testdatensatz, den das Team bei der Optimierung nicht verwendet hat.
 - **Read-only / draft-only:** Betriebsmodus ohne irreversible Schreibvorgänge oder autonome Entscheidungen.

1. *AI-Backend zur Vorverarbeitung von Versicherungsdokumenten*

Aus Sicht des Nutzers: Der Schadenregulierer öffnet einen Fall in der bestehenden Anwendung. Statt sämtliche Anhänge nacheinander zu lesen, sieht er eine vorab erstellte Zusammenfassung, eine Liste der ermittelten Fakten und Verweise auf konkrete Dokumente und Seiten. Jede Angabe kann er prüfen, ändern oder verwerfen. Das System trifft keine Entscheidungen an seiner Stelle und führt ohne Freigabe keine irreversiblen Schreibvorgänge aus.

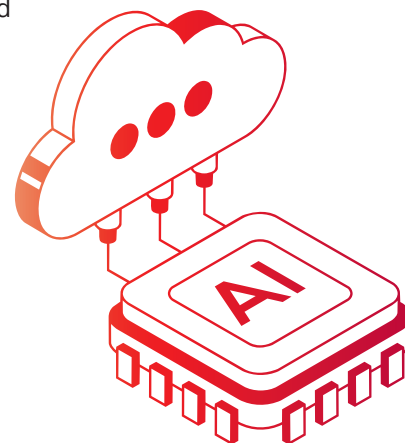
Funktionsweise der Lösung: Nach dem Eingang werden die Dokumente geprüft, mittels OCR und Extraktion in eine maschinenlesbare Form überführt und nach Typ geordnet. Cortex sucht relevante Passagen, erstellt einen strukturierten Entwurf und übergibt ihn über eine API zurück an den Versicherungsworkflow.

Vorteile: Kürzere Suchzeiten; einheitlichere Struktur der Unterlagen; Nachvollziehbarkeit von Aussagen.

Kontrollpunkte: Der Pilot arbeitet im draft-only-Modus; für sensible Daten sind Zweck und Aufbewahrungsdauer definiert.

KPI: Zeit bis zum fertigen Entwurf; Anteil der ohne Änderungen übernommenen Entwürfe; schwerwiegende Fehler.

Erster Pilot: Eine Schadenart, 100-300 Fälle, read-only/draft-only.



2. *Interner Wissensassistent für Mitarbeitende und Neueinsteiger*

Aus Sicht des Nutzers: Der Mitarbeiter meldet sich mit seiner Unternehmensidentität an und fragt nach einer Vorgehensweise. Der Assistent durchsucht ausschließlich Inhalte, auf die der jeweilige Nutzer zugreifen darf, antwortet mit Quellenangabe und bietet an, die Quelle zu öffnen. Bei Unsicherheit verweist er auf den Prozessverantwortlichen.

Funktionsweise der Lösung: Kuratierte Dokumente aus SharePoint oder Confluence werden mit Metadaten und ACL eingelesen. Die Suche berücksichtigt die ACL; anschließend wird eine Antwort mit Quellenangaben erstellt. Aus dem Feedback entsteht ein Backlog zur Korrektur der Quellen.

Vorteile: Schnellere Orientierung ohne Überlastung der Experten; stärkere Nutzung der Dokumentation.

Kontrollpunkte: Die Berechtigungen werden angewendet, bevor Inhalte an das Modell übergeben werden. Jede Antwort enthält eine Quellenangabe.

KPI: Recall@k; Genauigkeit der Quellenangaben; Anteil korrekter Abstentionen.

Erster Pilot: Eine Gruppe, 500-2000 Dokumente, 150+ Fragen.

3. *RFP Analyzer – Matrix zu Konformität, Pflichten und Risiken*

Aus Sicht des Nutzers: Der Manager lädt Dokumente hoch und erhält eine Anforderungsmatrix mit Kategorie, Schweregrad, Quellseite und Antwortvorschlag. Der Jurist konzentriert sich ausschließlich auf die kritischen Punkte.

Funktionsweise der Lösung: Die Pipeline identifiziert die Struktur der Dokumentation, extrahiert Anforderungen, entfernt Duplikate und erstellt eine Compliance-Matrix.

Vorteile: Schnellere Erstellung der Matrix; geringeres Risiko, eine kritische Bedingung zu übersehen; transparente Quellenangabe.

Kontrollpunkte: Kritische Anforderungen müssen eine Quelle und eine manuelle Bestätigung aufweisen. Das System darf kein Angebot einreichen.

KPI: Recall kritischer Anforderungen; false negatives; Zeitaufwand des Prüfers.

Erster Pilot: Eine abgeschlossene Ausschreibung + ein neues Holdout-Dokument.

4. *Technische Dokumentation in einer eingeschränkten oder kritischen Umgebung*

Aus Sicht des Nutzers: Der Techniker stellt eine Frage und erhält eine Antwort mit einem präzisen Verweis auf das Handbuch und den Revisionsstand. Fehlt die Quelle, gibt das System keine Antwort und leitet die Frage an einen Fachexperten weiter.

Funktionsweise der Lösung: Die Wissensschicht indexiert Handbücher und Revisionsstände. Die Suche wird durch die Rolle, den Gerätetyp und die Gültigkeit des Dokuments eingeschränkt.

Vorteile: Geringeres Risiko, einen ungültigen Revisionsstand zu verwenden; Sicherung von Know-how.

Kontrollpunkte: Eine Sicherheitsanweisung ohne Quelle wird nicht angezeigt. Der Offline-Modus erfordert einen kontrollierten Aktualisierungsprozess.

KPI: Anteil der Antworten mit dem korrekten Revisionsstand; Zeit bis zum Auffinden.

Erster Pilot: Ein Gerätetyp, ein Dokumentensatz.



5. *CV Converter – standardisiertes Berufsprofil*

Aus Sicht des Nutzers: Der Recruiter lädt einen Lebenslauf hoch und erhält einen strukturierten Profilentwurf in einer Vorlage. Das System hebt fehlende Felder hervor. Eine Person gibt das Ergebnis frei und exportiert es.

Funktionsweise der Lösung: Das Dokument durchläuft OCR, die Extraktion in ein Schema und die Normalisierung. Kritische Felder werden zur Prüfung markiert.

Vorteile: Weniger manuelle Übertragung; höhere Konsistenz; Export in ATS/CRM.

Kontrollpunkte: Minimierung personenbezogener Daten; Verbot der Generierung sensibler Merkmale.

KPI: Feldgenauigkeit; Rate kritischer Fehler; Prüfungsdauer.

Erster Pilot: 50-150 CVs, 10-20 Pflichtfelder.

6. *Access & Spend Baseline*

Aus Sicht des Nutzers: CIO, CISO und die Finanzabteilung erhalten einen Überblick über AI-Dienste, Projekte und Kosten. In der ersten Phase wird nichts abgeschafft; sie dient der Identifizierung von Lücken.

Funktionsweise der Lösung: Eine Read-only-Inventarisierung führt das Verzeichnis, die Projekte bei den Anbietern und die Abrechnungsdaten zusammen.

Vorteile: Einheitliche Zuständigkeit; Erkennung ungenutzter Lizenzen; Grundlage für FinOps.

Kontrollpunkte: Jede Person verwendet ihre eigene Identität. Die Automatisierung hängt von der Admin API ab.

KPI: Inventarabdeckung; verwaiste Projekte; nicht zugeordnete Kosten.

Erster Pilot: Ein Verzeichnis, 1-2 Anbieter, Daten aus 90 Tagen.

7. *Sicherer AI Workspace für Unternehmen*

Aus Sicht des Nutzers: Der Nutzer meldet sich mit seiner Unternehmensidentität an und wählt eine Aufgabe aus. Cortex wählt auf Grundlage der Identität und der Datenklassifizierung den Verarbeitungsweg: no-LLM, RAG oder ein allgemeines Modell.

Funktionsweise der Lösung: Die Eingabe durchläuft die Policy-Prüfung, die Klassifizierung und das Routing. Control misst Antwortzeit, Qualität und Kosten.

Vorteile: Reduzierung von Shadow AI; einheitliche Identität; transparente Messung der Nutzung.

Kontrollpunkte: Der Modellkatalog ist eine Allowlist. Die Nutzeraktivität muss mit einer abgeschlossenen Aufgabe verknüpft sein.

KPI: Aktive Nutzer; Aufgabenabschluss; Kosten pro aktivem Nutzer.

Erster Pilot: Eine Gruppe, drei definierte Aufgaben, ein freigegebenes Modell.



8. *Modernisierung von Legacy-Software – discovery-first*

Aus Sicht des Nutzers: Der Architekt erhält ein Inventar der Module, Abhängigkeiten und Geschäftsregeln. AI schlägt Tests vor, Änderungen am Code durchlaufen jedoch das übliche Review.

Funktionsweise der Lösung: Eine Read-only-Analyse des Repositorys erstellt eine Abhängigkeitsübersicht. Anschließend werden Migrations-Slices entworfen.

Vorteile: Schnellere Bestandsaufnahme; geringeres Risiko bei der Neuimplementierung; schrittweiser Ausbau der Tests.

Kontrollpunkte: Kein autonomer Commit; Geschäftsregeln werden von einem Menschen freigegeben.

KPI: Abdeckung kritischer Pfade; test pass rate.

Erster Pilot: Ein Modul mit einem Betriebsverantwortlichen.



9. Öffentliche Verwaltung – Klassifizierung von Eingaben und Vorbereitung der Erfassungsdaten

Aus Sicht des Nutzers: Der Sachbearbeiter sieht einen Vorschlag für die Zuordnung zu einem Aufgabenbereich sowie Metadaten. Er prüft und bestätigt die Einträge. Strittige Fälle werden an einen erfahrenen Mitarbeiter weitergeleitet.

Funktionsweise der Lösung: Das Dokument durchläuft OCR, Klassifizierung und Validierung. Flow steuert die menschliche Freigabe und die Integration in die Schriftgutverwaltung.

Vorteile: Schnellere Erfassung; einheitlichere Metadaten; nachvollziehbarer Unterschied zwischen Vorschlag und Entscheidung.

Kontrollpunkte: AI trifft keine Entscheidungen in der Sache. Kritische Fehlzuordnungen werden gemessen.

KPI: Klassifizierungsgenauigkeit; critical misroute rate.

Erster Pilot: Ein Aufgabenbereich, eine anonymisierte Stichprobe.

10. BI-Fragen in natürlicher Sprache mit deterministischer Berechnung

Aus Sicht des Nutzers: Der Manager fragt nach einer Kennzahl. Das Modell ordnet die Frage einer definierten Kennzahl zu. Die Berechnung erfolgt durch die Read-only-Datenschicht.

Funktionsweise der Lösung: Die Anfrage wird im Kontext der Identität ausgewertet. Semantic metric mapping wählt eine freigegebene Abfrage aus.

Vorteile: Schnellere Antworten ohne Analysten; einheitliche Definition der Kennzahlen.

Kontrollpunkte: Allowlist der Kennzahlen; row-level security; das Modell interpretiert ausschließlich das Ergebnis.

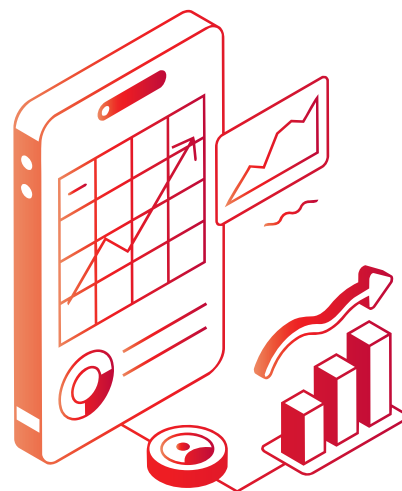
KPI: Korrektheit der Kennzahlauswahl; Übereinstimmung mit der BI-Berechnung.

Erster Pilot: 10-20 Kennzahlen, Read-only-Datenquelle.

Erweiterter Katalog der Vertriebsszenarien (einschließlich detaillierter Erläuterungen)

IT Service Desk

Direct route ohne LLM ermöglicht eine sofortige, kosteneffiziente Reaktion auf bekannte Anliegen wie Kontakte, den Status von Diensten oder genaue Vorgehensweisen. Erst ungelöste oder unklare Anfragen gelangen in die Wissensschicht oder werden an einen menschlichen Service-Mitarbeiter weitergeleitet. Der wesentliche Nutzen besteht in einer messbaren Reduzierung von Routinetickets, kürzeren Lösungszeiten und einer gleichbleibenden Antwortqualität.



Einkauf

Das System extrahiert automatisch wesentliche Geschäftsbedingungen, Preise, Fristen und Garantien aus unterschiedlichen Angeboten und vergleicht sie in einer übersichtlichen Gap-Matrix. Der Einkäufer erhält damit eine vergleichbare Struktur für schnelle Entscheidungen. Die endgültige Lieferantenauswahl und die Preisentscheidung erfolgen weiterhin im Rahmen des freigegebenen Prozesses und nach menschlichem Ermessen. Der Mehrwert liegt in einer drastischen Zeitverkürzung bei gleichbleibend hohem Kontrollniveau.

Recht und Verträge

Cortex identifiziert und extrahiert rechtliche Klauseln aus Vertragstexten und vergleicht sie mit dem internen juristischen Playbook. Dabei kennzeichnet es potenzielle Abweichungen unmittelbar mit Quellenangaben. Es handelt sich nicht um eine automatisierte rechtliche Stellungnahme, sondern um eine effiziente Arbeitsgrundlage, die Juristen ermöglicht, sich auf kritische und risikobehaftete Passagen zu konzentrieren. Das Ergebnis ist eine deutlich schnellere Vertragsprüfung und eine geringere Wahrscheinlichkeit, eine ungeeignete Klausel zu übersehen.

Produktion und Instandhaltung

Die Lösung führt frei formulierte Berichte aus Protokollen, Vorfallmeldungen und technischen Handbüchern zusammen und strukturiert sie zu einer einheitlichen technischen Zusammenfassung. Dies ermöglicht Technikern und Instandhaltern eine schnellere Ursachenanalyse von Störungen (RCA) und eine effizientere Reparaturplanung. Ein kritischer Kontrollpunkt ist stets der korrekte Versions- und Revisionsstand der Dokumentation, damit keine veralteten oder fehlerhaften Verfahren angewendet werden.

Vertrieb

Vor einem Verkaufsgespräch sammelt und konsolidiert Cortex automatisch Informationen aus dem CRM-System, internen Dokumenten, offenen Aufgaben und weiteren freigegebenen Quellen in einem übersichtlichen Briefing. Der Vertriebsmitarbeiter erhält dadurch ein vollständiges Bild des Kunden und der aktuell offenen Themen. Dies verkürzt die Vorbereitungszeit und erhöht die Qualität und Professionalität des Gesprächs selbst.

Finanzen

Kategorisierung von Ausnahmefällen vor und bereitet eine strukturierte Grundlage für das Freigabeverfahren vor. Sämtliche Buchungen, Freigaben und Zahlungen verbleiben jedoch im deterministischen System mit strengen Berechtigungen für Finanzvorgänge. Der Nutzen besteht in einer schnelleren Verarbeitung, einer höheren Konsistenz der Buchführung und einer einfacheren Nachvollziehbarkeit für Prüfungszwecke.

Entwicklung

Ein mit Jira integrierter Agent kann auf Grundlage einer Anforderung oder des Quellcodes erste Entwürfe für Analysen, Testfälle und Änderungen erstellen. Die Autonomie endet jedoch beim Entwurf – jeder Commit, jeder Merge und jedes Deployment in die Produktionsumgebung bleiben in der Hand des Entwicklers und durchlaufen das übliche Code Review und den CI/CD-Prozess. Dies beschleunigt die Erstellung von Testsuiten, ohne die Qualität oder Sicherheit des Codes zu gefährden.

Sprachassistent

Die Spracherkennungsschicht wandelt eine gesprochene Anfrage in strukturierten Text um, den Cortex anschließend auf den wissensbasierten oder prozessbezogenen Verarbeitungsweg leitet. Erreicht das System nicht den erforderlichen Konfidenzgrad, übergibt es das Gespräch automatisch und nahtlos an einen menschlichen Service-Mitarbeiter (Handoff). Der Mehrwert entsteht durch die Verfügbarkeit eines Supports rund um die Uhr mit einer klaren und sicheren Übergabe komplexerer Fälle an einen Menschen.

Meeting operations

Aus dem Transkript oder der Aufzeichnung einer Besprechung extrahiert Cortex die wesentlichen Entscheidungen, Maßnahmen, verantwortlichen Personen und Fristen und erstellt einen Protokollentwurf. Die endgültige Veröffentlichung oder die Anlage einer Aufgabe im System erfolgt erst nach Freigabe durch die Besprechungsteilnehmer. Dies reduziert den Verwaltungsaufwand erheblich, verbessert die Handlungsfähigkeit des Teams und stellt sicher, dass tatsächlich nur freigegebene Verpflichtungen erfasst werden.

AI-Governance-Katalog

Mithilfe von Access und Control erstellen wir ein laufend aktualisiertes Verzeichnis aller AI-Anwendungsfälle, Modelle, Datensätze, Tools, Verantwortlichen und zugeordneten Risiken. Dieser Katalog verbindet den technischen Betrieb mit der Entscheidungsverantwortung und dem Kostenmanagement. Ein Pilot in einer Organisationseinheit mit einem begrenzten Portfolio überprüft die Abdeckung und etabliert Prozesse, die zu einer nachhaltigen und regulierten AI-Transformation der gesamten Organisation führen.

Verbindliche Bestandteile des Pilot Charters

Pilot Charter ist ein kurzes Steuerungsdokument, das sowohl den Kunden als auch den Anbieter davor schützt, eine technische Demo als Nachweis des Nutzens im Produktivbetrieb zu betrachten. Ein Pilot ist nicht deshalb erfolgreich, weil das Modell eine beeindruckende Antwort erzeugt hat. Er ist dann erfolgreich, wenn er anhand repräsentativer Daten die im Voraus festgelegten Anforderungen an Qualität, Sicherheit und Business outcome erfüllt.

- | | |
|---|--|
| ✓ Business outcome: Wer welche Aufgabe besser abschließt. | ✓ Acceptance: Metrik, Schwellenwert, Toleranz gegenüber kritischen Fehlern. |
| ✓ Baseline: Volumen, Zeit, Fehlerquote, Nacharbeit. | ✓ Control points: Stellen mit menschlicher Freigabe und Rollback. |
| ✓ In / out of scope: Präzise Abgrenzung der Daten, Prozesse und Verbote. | ✓ TCO: Implementierung, Plattform, Verbrauch, Betrieb. |
| ✓ Data & authority: Verantwortlicher, Rechtsgrundlage, Berechtigungen, Aufbewahrungsdauer. | ✓ Betriebsverantwortung: RACI. |
| ✓ Eval dataset: Repräsentative Fälle und Grenzfälle + Holdout. | ✓ Decision gate: Datum und Personen für die Entscheidung über stop / iterate / scale. |

Qualifizierungsfragen für das Vertriebsgespräch

- Wer führt die Aufgabe derzeit aus, und wer trägt die Verantwortung?
- Wie viele Fälle fallen an, und wo geht Zeit verloren?
- Wie sieht ein schwerwiegender Fehler aus, und welche Toleranz gilt?
- Welche Daten und Systeme werden benötigt, und wer darf sie zugänglich machen?
- Für welchen Teil darf AI einen Vorschlag erstellen, und welchen Teil darf sie nach Freigabe ausführen?
- Welche Voraussetzungen müssen erfüllt sein, damit der Kunde in den Produktivbetrieb investiert? Was würde dagegen zu einem Abbruch führen?

