

# GEMMA ANALYTICS

## PROJEKTVORSCHLAG

Data Lake & Claude Cowork

---

Erstellt für

**ROTOP Pharmaka GmbH**

26. Mai 2026

7 Tage gültig

---

# Inhaltsverzeichnis

Zusammenfassung .....	3
Ausgangslage .....	3
Unternehmen & Kontext .....	3
Aktuelle Herausforderungen .....	3
Zielsetzung .....	3
Ansatz .....	4
Projektstruktur .....	4
Phase 1 – Scope .....	5
Technische Architektur .....	6
Technisches Setup .....	7
Option A – bevorzugt .....	7
Option B – Fallback .....	7
Tool-Stack .....	7
Datenschutz .....	7
Hinweis Phase 2 .....	8
Zeitplan .....	8
Zeitplanübersicht .....	8
Investition .....	9
Phase 1: Data Lake Foundation (Budget) .....	9
Laufende Toolkosten (Richtwerte) .....	9
Gesamtübersicht .....	10
Explizite Abgrenzung .....	10
Voraussetzungen .....	10
Arbeitsweise .....	11
Rechtlicher Rahmen .....	11
Phase 2 & Ausblick .....	11
Phase 2: Recurring Pipeline ( Q4 2026) .....	11
Mögliche Phase-3-Workstreams .....	11
Über Gemma Analytics .....	13
Kontakt .....	13

## ZUSAMMENFASSUNG

Gemma Analytics schlägt vor, für ROTOP Pharmaka ein zentrales Data Lake auf Basis von Snowflake aufzubauen und über **Claude Cowork** als Self-Service-Interface zugänglich zu machen. Anstelle klassischer BI-Dashboards wird Claude als primäre Schnittstelle zu den Daten eingesetzt – sauber modellierte Daten, zugänglich per natürlicher Sprache.

Das Engagement ist in **drei Phasen** gegliedert:

- **Phase 1 (Data Lake Foundation)** – Aufbau eines arbeitsfähigen Lakehouse auf Basis eines statischen ERP-Snapshots. Funktionierende MVPs für Finance (konsolidierter Management-Report) und Sales (Auftrags-/Kundensichten via Claude). Agentic Workflows für Voice-to-CRM und Auftragseingangs-Automatisierung. Sofortiger Start, ca. 4–6 Wochen Laufzeit.
- **Phase 2 (Recurring Pipeline)** – Nach dem MDG-Carve-out ( Ende September 2026) Migration von statischem Snapshot auf tägliche/inkrementelle Pipeline. Keine Architekturänderung, nur Verbindungsmethode.
- **Phase 3 (Erweiterungen)** – Produktion/Ops-Daten, CDMO-Pipeline, weitere Workstreams. Separat zu planen nach dediziertem Workshop.

**Geschätztes Budget Phase 1: EUR 24.200** (20 Tage).

## AUSGANGSLAGE

### UNTERNEHMEN & KONTEXT

ROTOP Pharmaka GmbH ist ein Dresdner Nuklearmedizin-Pharmaunternehmen mit ca. 200 Mitarbeitern und rund 50 Jahren Unternehmensgeschichte. Das Unternehmen stellt radioaktive Tracer für Krebsdiagnostik und -therapie her. ROTOP ist Teil des GENUI-Portfolios und befindet sich aktuell in der Digitalisierung des Reporting- und Analyse-Stacks. Claude Cowork wurde unternehmensweit in der Woche ab 19. Mai 2026 ausgerollt.

Zwei Rechtseinheiten – ROTOP Pharmaka und ROTOP Pharmazie – werden aktuell getrennt geführt und manuell konsolidiert.

### AKTUELLE HERAUSFORDERUNGEN

- **Monatliches Management-Reporting** dauert aktuell eine Woche manueller Excel-Arbeit (Finance). Ziel: Minuten.
- **Kein CRM** – Außendienst und Vertriebsleitung haben keinen strukturierten Zugriff auf Auftrags-/Kundendaten.
- **Zwei Rechtseinheiten** (Pharmaka + Pharmazie) erfordern konsolidiertes Reporting; derzeit manuell und fehleranfällig.
- **Auftragseingang** aus Email/Fax ist ein 3–4-stündiger manueller Prozess pro Batch.
- **Produktions-/Ops-Daten** liegen unstrukturiert in Dateiverzeichnissen (Excel, MS-Teams-Ordner).

### ZIELSETZUNG

1. **Zentrales Data Lake** – saubere, einheitliche Datenbasis für alle Unternehmensbereiche

2. **Claude Cowork als Interface** – Self-Service-Analysen per natürlicher Sprache, keine klassischen Dashboards
3. **Agentic Workflows** – automatisierte Routineaufgaben (Management-Report, Auftragseingang, Sales-Briefings)
4. **Agile Iteration** – bauen, zeigen, behalten was funktioniert
5. **Schneller erster Wert** – arbeitsfähige MVPs innerhalb weniger Wochen

## ANSATZ

- **Data Lake, nicht Data Warehouse** – Snowflake als zentraler Datenspeicher, optimiert für die Anbindung an Claude Cowork statt für klassisches Reporting
- **Claude als primäre Schnittstelle** – Stakeholder fragen ihre Daten direkt per Claude Cowork ab, unterstützt durch einen sauberen Metadaten-Layer in dbt
- **Vertikale Wertlieferung** – End-to-End-Ergebnisse (Datenquelle → Modell → Claude-Abfrage) statt isolierter Infrastrukturarbeit
- **Iterativ, nicht Wasserfall** – MVPs zuerst, dann schrittweise Verfeinerung basierend auf Stakeholder-Feedback
- **Code-first, KI-gestützt** – alle Pipelines und Modelle versioniert in Git; KI-Tools (Cursor, Claude Code) als Standard für beschleunigte Entwicklung

## PROJEKTSTRUKTUR

### Phase 1: Data Lake Foundation

Aufbau eines arbeitsfähigen Lakehouse auf statischem ERP-Snapshot. Snowflake als Data Lake, dbt Cloud für Transformationen, Claude Cowork via MCP als Interface. MVPs für Finance und Sales. Agentic Workflows für Voice-to-CRM und Auftragseingangs-Automatisierung.

**Start:** Juni 2026

**Dauer:** ca. 4–6 Wochen

**Modell:** Budget (Time & Materials)

### Phase 2: Recurring Pipeline

Migration von statischem Snapshot auf tägliche/inkrementelle Pipeline nach dem MDG-Carve-out. Keine Architekturänderung; Verbindungsmethode wird von einmaligem SSH-Dump auf geplanten Connector umgestellt.

**Start:** nach MDG-Carve-out ( Q4 2026)

**Dauer:** separat zu planen

**Modell:** separat zu vereinbaren

### Phase 3: Erweiterungen

Produktion/Ops-Daten, CDMO-Pipeline, Kalender-/Außendienstanalyse. Erfordert dedizierte Workshops mit den jeweiligen Fachbereichen. Nicht Teil dieser Beauftragung.

**Start:** separat zu planen

**Modell:** separat zu vereinbaren

## PHASE 1 – SCOPE

### Infrastruktur-Setup

- **Snowflake** Data Lake (EU-Region, Trial- oder ROTOP-eigenes Konto)
- **dbt Cloud** (Paid Tier) für Transformationen – inkl. integriertem MCP-Server für Claude-Anbindung
- **GitHub** für Versionskontrolle
- CI/CD-Pipeline für automatisierte Tests und Deployments

### Datenanbindung

- Einmaliger Snapshot von **Business Central** (beide Rechtseinheiten) und **Sage 100** (ROTOP Pharmazie)
- Ca. 6 Monate Historie, eingefroren Ende Mai 2026
- Transfer via SSH (Public-Key-Authentifizierung + IP-Whitelisting) vom ROTOP-Linux-Server
- Laden nach Snowflake

### dbt-Datenmodellierung

Schicht	Inhalt
<b>Staging</b>	Rohdaten bereinigen, typisieren, standardisieren. Eine Base-Tabelle pro ERP-Entität.
<b>Transformation</b>	Geschäftslogik, KPI-Definitionen, Cross-Entity-Joins. Konsolidierte GuV über beide Rechtseinheiten, Sales Cube (Produkt × Kunde × Gebiet), Auftragseingangssicht.

- Umfassende Datentests pro Schicht
- Semantische Dokumentation und Spaltenbeschreibungen (Metadaten-Layer für Claude)

### Claude Cowork Integration

- Anbindung des **dbt Cloud MCP-Servers** an Claude Cowork
- Semantischer Metadaten-Layer: Tabellen- und Spaltenbeschreibungen, damit Claude Daten ohne Raten interpretieren kann
- Stakeholder (Dr. Peterseim, Dr. Gottlieb u.a.) fragen das Data Lake direkt per Claude ab

### Finance-MVP

- Automatisierter monatlicher Management-Report: Monatsabschluss-Trigger → Claude generiert GuV, Bilanz, Cashflow aus dem Data Lake → Anomalie-Erkennung
- Konsolidierung über beide Rechtseinheiten (ROTOP Pharmaka + ROTOP Pharmazie)
- Ziel: von einer Woche manueller Arbeit auf Minuten

### Sales-MVP

- Tagesaktuelle Auftrags-/Kunden-/Gebietssicht, abfragbar per Claude
- Top-10-Kunden, Produktentwicklung, Kundenentwicklung nach Gebiet
- Das Data Lake wird zum de-facto CRM-light (da kein CRM vorhanden)

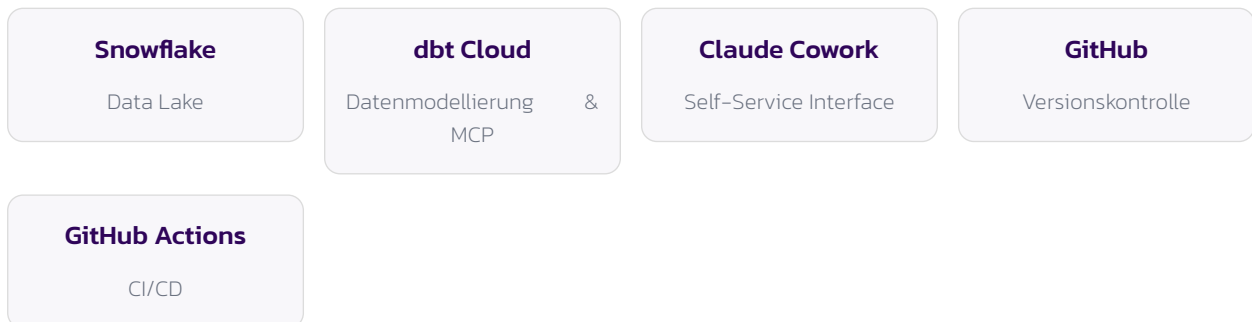
### Voice-to-CRM Workflow

- Außendienst diktiert 10–20 Sekunden Sprachmemo nach Kundenbesuch
- Claude erstellt strukturierten Eintrag: Besuchsqualität, besprochene Themen, Follow-up-Wahrscheinlichkeit
- Daten fließen ins Data Lake und sind für Vertriebsleitung per Claude abrufbar
- Agentic-Workflow-Design und -Implementierung

### Auftragseingangs-Automatisierung

- Email-/Fax-Aufträge → Claude parst und extrahiert Bestelldaten → Sage-kompatible CSV → manuelle Gegenprüfung → Upload
- Aufbauend auf bestehender interner Prototypisierung (Julia, Marketing)
- Ziel: aktueller 3–4-Stunden-Prozess → ca. 2 Minuten

## TECHNISCHE ARCHITEKTUR



- **Snowflake als Data Lake** – EU-Region, skalierbar, getrenntes Compute/Storage
- **dbt Cloud (Paid Tier)** – integrierter MCP-Server für direkte Claude-Anbindung; Scheduling, Monitoring, Web-IDE
- **Claude Cowork als Interface** – Self-Service per natürlicher Sprache statt Dashboards
- **Minimal-Infrastruktur** – kein eigener Orchestrator, keine eigenen Server (außer dem temporären Linux-Server für den initialen Datentransfer)
- **Code-first** – alle Modelle und Pipelines versioniert in Git

## TECHNISCHES SETUP

Dieses Kapitel kann als eigenständiges Dokument an das IT-Team weitergegeben werden.

### OPTION A – BEVORZUGT

1. Torsten (oder MDG via Ute) exportiert einen einmaligen Dump von Business Central + Sage. Abdeckung: beide ROTOP-Rechtseinheiten. Ca. 6 Monate Historie, eingefroren Ende Mai 2026.
2. Torsten stellt einen Linux-Server bereit (ROTOP-kontrolliert). Hosting-Wahl liegt bei ROTOP – Gemma arbeitet mit jeder Plattform: AWS, GCP, Azure, On-Premise, Hetzner.
3. Der Dump wird auf dem Linux-Server abgelegt.
4. Gemma verbindet sich via SSH (Public-Key-Authentifizierung) + IP-Whitelisting vom Gemma-VPN. Bijan sendet den SSH-Public-Key, sobald der Server bereit ist.
5. ROTOP-Daten verbleiben durchgehend auf ROTOP-kontrollierter Infrastruktur.
6. Gemma lädt die Daten in ein Snowflake Data Lake (Trial-Konto oder ROTOP-eigenes Konto – nach Vereinbarung).
7. dbt Cloud (Paid Tier) führt Transformationen durch und stellt den integrierten MCP-Server bereit.
8. Claude Cowork verbindet sich mit dem dbt Cloud MCP – ROTOP-Mitarbeiter fragen Daten per Claude ab.

### OPTION B – FALLBACK

Wie Option A, mit einer Änderung: Falls der Speicherplatz auf dem Linux-Server begrenzt ist, wird das Snowflake Data Lake in Snowflakes Cloud (EU-Region) gehostet statt auf dem Linux-Server selbst. Zugriffsmuster bleibt gleich – SSH für Dump-Lieferung, Snowflake-Credentials für Query-Zugriff. Alles andere (dbt, GitHub, Claude) identisch.

### TOOL-STACK

Schicht	Tool
<b>Quelldaten</b>	Business Central + Sage 100 (statischer Dump, ca. 6 Monate)
<b>Server / Datentransfer</b>	ROTOP-Linux-Server (On-Prem oder Cloud) + SSH + IP-Whitelist
<b>Data Lake</b>	Snowflake (EU-Region)
<b>Transformation</b>	dbt Cloud (Paid Tier, inkl. MCP-Server)
<b>Versionskontrolle</b>	GitHub
<b>KI-Interface</b>	Claude Cowork (via dbt Cloud MCP)

### DATENSCHUTZ

ROTOP-Daten verbleiben auf ROTOP-eigener Infrastruktur (Linux-Server) für den Dump-/Transferschritt. Die Snowflake-Instanz wird in einer EU-Region gehostet. IP-/Herstellungsprozessdaten und personenbezogene Daten (Personio) sind explizit nicht im Scope.

## HINWEIS PHASE 2

Sobald ROTOP die MDG-Infrastruktur verlässt (erwartet Ende September 2026), wird der statische Dump durch eine tägliche/inkrementelle Pipeline ersetzt, die direkt an ROTOPs eigene BC- und Sage-Datenbanken anbindet. Keine Architekturänderungen erforderlich; nur die Verbindungsmethode ändert sich.

## ZEITPLAN

- **Vorbereitung** Juni 2026, Woche 1

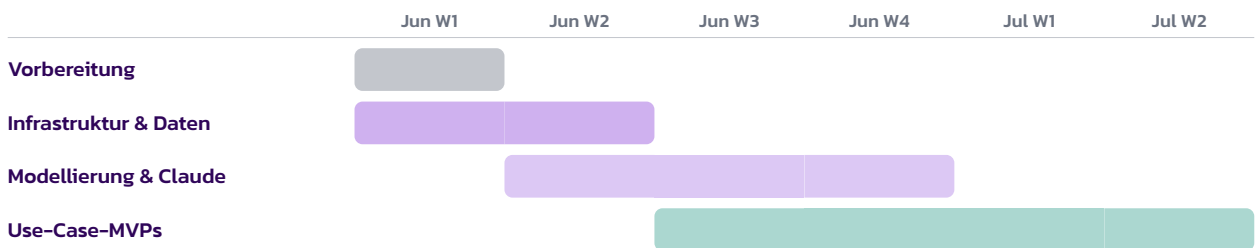
Verträge unterzeichnet. Torsten stellt Linux-Server bereit und koordiniert ERP-Dump. SSH-Zugang eingerichtet. Snowflake- und dbt-Cloud-Konten provisioniert.
- **Infrastruktur & Datenanbindung** Juni 2026, Wochen 1--2

Snowflake-Setup, dbt Cloud, GitHub, CI/CD. ERP-Dump transferiert und in Snowflake geladen. Staging-Layer in dbt.
- **Datenmodellierung & Claude-Anbindung** Juni 2026, Wochen 2--4

Transformations-Layer, konsolidierte Modelle, Metadaten-Layer. MCP-Anbindung an Claude Cowork. Erste Stakeholder-Abfragen möglich.
- **Use-Case-MVPs & Agentic Workflows** Juni/Juli 2026, Wochen 3--6

Finance-MVP (Management-Report, Konsolidierung). Sales-MVP (Kundensichten). Voice-to-CRM Workflow. Auftragseingangs-Automatisierung. Iteration basierend auf Stakeholder-Feedback.

## ZEITPLANÜBERSICHT



## INVESTITION

Unser Tagessatz beträgt **EUR 1.210** (1 Tag = 8 Stunden).

### PHASE 1: DATA LAKE FOUNDATION (BUDGET)

Leistung	Gesch. Tage
Infrastruktur-Setup (Snowflake, dbt Cloud, GitHub, CI/CD)	1
Datenanbindung (BC + Sage Snapshots → Snowflake)	2
dbt-Modellierung: Staging-Layer	1,5
dbt-Modellierung: Transformations-Layer (GuV, Sales Cube, Konsolidierung)	3
Claude Cowork Integration (MCP-Anbindung, Metadaten-Layer)	1
Finance-MVP (Management-Report, Konsolidierung, Anomalie-Erkennung)	2
Sales-MVP (Auftrags-/Kunden-/Gebietssichten)	1,5
Voice-to-CRM Workflow (Sprachmemo → strukturierter Eintrag)	1,5
Auftragseingang-Automatisierung (Email/Fax → Sage-CSV)	2
Projektmanagement, Workshops, Abstimmung	2,5
Puffer	1,5
<b>Budget</b>	<b>20 Tage × EUR 1.210 = EUR 24.200</b>

Alle Preise netto (zzgl. USt.). Budget-Positionen werden monatlich nach tatsächlichem Aufwand abgerechnet. Nicht verbrauchtes Budget kann für weitere Anwendungsfälle innerhalb von Phase 1 eingesetzt werden. Bei Mehrbedarf kann eine Budgeterweiterung vereinbart werden. Reisekosten werden, falls anfallend, zum Selbstkostenpreis weiterberechnet.

### LAUFENDE TOOLKOSTEN (RICHTWERTE)

Tool	Kosten/Monat	Anmerkung
<b>Snowflake</b>	100–500 EUR	Nutzungsbasiert. Bei statischem Snapshot minimal.
<b>dbt Cloud</b>	100 EUR/Nutzer	Paid Tier, inkl. MCP-Server.

<b>Claude Cowork</b>	bestehendes Abo	Bereits unternehmensweit ausgerollt.
----------------------	-----------------	--------------------------------------

Toolkosten werden direkt von ROTOP getragen und sind nicht im Gemma-Budget enthalten.

## GESAMTÜBERSICHT

<b>Phase 1: Data Lake Foundation (Budget)</b>	EUR 24.200
<b>Budget Phase 1</b>	<b>EUR 24.200</b>

## EXPLIZITE ABGRENZUNG

Folgendes ist **nicht** Bestandteil dieser Beauftragung:

- **IP- und Herstellungsprozessdaten** – aus Compliance-Gründen explizit ausgeschlossen (Cloud-Verarbeitung nicht zulässig)
- **HR-/Lohndaten** (Personio, DATEV) – nicht im Scope
- **GMP-/GDP-nahe Prozessautomatisierung** – regulatorische Klärung ausstehend; nicht in Phase 1
- **Produktion/Ops-Daten** – erfordert dedizierten Workshop mit Hendrike Heinrich und der Betriebsleitung; separat zu planen
- **BI-Dashboards oder Power BI** – explizit nicht Teil der Vision oder des Angebots
- **Phase 2 (Recurring Pipeline)** und **Phase 3 (Erweiterungen)** – separat zu beauftragen

## VORAUSSETZUNGEN

Folgendes sollte vor Engagementstart bereitstehen bzw. parallel vorbereitet werden:

- **Linux-Server** – Torsten Rex stellt einen Linux-Server bereit (ROTOP-kontrolliert) und koordiniert den ERP-Dump mit MDG/Ute
- **ERP-Dump** – einmaliger Export von Business Central (beide Entitäten) und Sage 100 (Pharmazie), ca. 6 Monate Historie
- **SSH-Zugang** – Public-Key-Authentifizierung + IP-Whitelisting für Gemma-VPN
- **Snowflake-Konto** – Trial oder ROTOP-eigenes Konto (nach Vereinbarung)
- **dbt Cloud Konto** – Paid Tier (nach Vereinbarung)
- **GitHub-Zugang** – Repository mit Schreibrechten für Gemma
- **Ansprechpartner** – fachliche Ansprechpartner für KPI-Abstimmung (Finance: Dr. Peterseim, Sales: Dr. Gottlieb, IT: Torsten Rex)

## ARBEITSWEISE

Bereich	Ansatz
<b>Kommunikation</b>	Direkter Draht per Slack, Teams oder bevorzugtem Kanal
<b>Sync-Kadenz</b>	Wöchentlicher Call zur Fortschrittsbesprechung und Priorisierung
<b>Tooling</b>	Cursor / Claude Code / KI-gestützte Entwicklung als Standard
<b>Code-Review</b>	Alle Arbeit via Pull Requests mit Review vor Merge
<b>Dokumentation</b>	Alle Modelle und Pipelines dokumentiert und übergabefähig
<b>Ansprechpartner</b>	Torsten Rex (IT), Dr. Peterseim (Finance), Dr. Gottlieb (Sales)

## RECHTLICHER RAHMEN

Das Engagement wird durch folgende Verträge geregelt:

Dokument	Zweck
<b>Vertraulichkeitsvereinbarung (NDA)</b>	Gegenseitige Vertraulichkeit. Bereits unterzeichnet.
<b>Rahmenvertrag (FSA)</b>	Rahmenvereinbarung: Haftung, IP, allgemeine Pflichten.
<b>Auftragsverarbeitungsvertrag (AVV)</b>	Gemäß DSGVO für die Verarbeitung personenbezogener Daten.
<b>Beauftragung</b>	Unterzeichnungsfähiges Dokument mit Scope, Leistungen und Preisen.

**Rechnungsstellung:** Abrechnung monatlich nach tatsächlichem Aufwand. Zahlungsziel: 14 Tage netto.

## PHASE 2 & AUSBLICK

### PHASE 2: RECURRING PIPELINE ( Q4 2026)

Nach dem MDG-Carve-out ( Ende September 2026) gewinnt ROTOP eigene BC- und Sage-Datenbanken. Der statische Snapshot wird durch eine tägliche/inkrementelle Pipeline ersetzt. Technischer Stack unverändert; nur die Verbindungsmethode wird aktualisiert.

### MÖGLICHE PHASE-3-WORKSTREAMS

- **Produktion/Ops-Daten** – OT/Sensor-Daten, GLT, Produktionsaufzeichnungen. Erfordert Workshop mit Hendrike Heinrich und der Betriebsleitung.
- **CDMO-Pipeline** – Funnel-Management, Rohertrag pro CDMO-Kunde.
- **Kalender/Außendienst** – Outlook-Kalender-Zugriff für Besuchsfrequenz-Analyse (vorbehaltlich Datenschutz-Freigabe).
- **Weitere Datenquellen** – zusätzliche ERP-Anbindungen und Quellsysteme nach Bedarf.

## ÜBER GEMMA ANALYTICS

Gemma Analytics ist eine Berliner Datenberatung, spezialisiert auf die Implementierung moderner Datenplattformen. Wir arbeiten mit wachsenden Unternehmen am Aufbau produktionsreifer Dateninfrastruktur – von Data-Lake-Setup und Pipeline-Engineering bis hin zu KI-gestützten Analyseworkflows und Team-Enablement.

Unser Team bringt tiefe Erfahrung über den gesamten modernen Data Stack mit, aus über 70 abgeschlossenen Datenplattform-Projekten. Wir arbeiten code-first und nutzen KI-gestütztes Tooling, um schnell zu liefern, ohne Kompromisse bei Qualität oder Wartbarkeit einzugehen.

Über diese Projekte hinweg haben wir umfangreiche interne Dokumentation, wiederverwendbare Muster und Engineering Best Practices aufgebaut, die in jedes neue Engagement einfließen – und die Entwicklung vom ersten Tag an signifikant beschleunigen.

## KONTAKT



### **Bijan Soltani**

Geschäftsführer

[bijan.soltani@gemmaanalytics.com](mailto:bijan.soltani@gemmaanalytics.com)

+49 179 7438926

# **GEMMA ANALYTICS**

Gemma Analytics GmbH – Chausseestraße 17, 10115 Berlin  
[gemmaanalytics.com](https://gemmaanalytics.com)