

# Softwareangebot: KI-gestütztes Email-Marketing-Tool

## Inhaltsverzeichnis

Beschreibung.....	2
Kernherausforderungen.....	3
Hohe Varianz.....	3
Bot-Detection.....	3
Crawling-Zeit.....	3
Betriebskosten.....	3
Konzept.....	4
Phase 1: Exploration und Datenbasis.....	4
Phase 2: Überwachung.....	4
Phase 3: Nutzung.....	4
Phase 4: Feedback-Loop & Angebotslogik.....	4
Technische Realisierung.....	4
Angebotsstruktur.....	5
Teil 1: Machbarkeitsstudie (PoC) & Prototyping.....	5
Teil 2: Integration & Skalierung (MVP).....	5
Teil 3: Produktentwicklung & Wartung.....	5
Weiteres Vorgehen.....	5

## Beschreibung

**Zielsetzung:** Transformation einer statischen E-Mail-Liste in ein dynamisches, automatisiertes Vertriebssystem, das Kundenbedarfe periodisch erkennt und personalisierte Erstkontakte vorbereitet.

**Der Effizienz-Hebel:** Automatisierung der zeitintensiven Recherche (Web-Crawling & Analyse). Statt manueller Suche liefert das System täglich 10 unterschrittsreife E-Mail-Entwürfe direkt in Ihren Drafts-Ordner.

**Wettbewerbsvorteil durch lokale KI:** Durch den Einsatz lokaler Sprachmodelle (IBM Granite) werden laufende Token-Kosten eliminiert und höchste Datenschutzstandards gewahrt, da sensible Firmendaten Ihre Infrastruktur nicht verlassen.

**Proaktiver Vertrieb:** Das System reagiert nicht nur auf Änderungen beim Kunden, sondern identifiziert durch semantisches Matching sofort die relevantesten Empfänger aus Ihrer Datenbank (5.000+ Kontakte), sobald Sie ein neues eigenes Angebot definieren.

### **Aus einer bestehenden E-Mail-Liste sollen automatisch:**

- Die Unternehmenswebsite/URL extrahiert werden.
- Aus der Unternehmenswebsite periodisch relevante Informationen extrahiert werden.
- Bei relevanten Updates individualisierte E-Mails an die Empfänger generiert werden, die nach Durchsicht und ggf. Anpassung einfach versendet werden können.
- Täglich maximal 10 (X) E-Mails vorbereitet werden, um die Menge handhaben zu können.
- Alle E-Mails individualisiert für den Empfänger sein und sich nicht wie ein Newsletter lesen.
- Änderungen im eigenen Angebot ebenfalls zu E-Mails führen, die optimalerweise bereits nach relevanten Empfängern vorsortiert sind.

# Kernherausforderungen

Nach einer ersten Analyse ergeben sich im Wesentlichen folgende Herausforderungen bei einer Annahme von 5.000 E-Mail-Adressen:

## Hohe Varianz

Eine der größten Herausforderungen beim Web-Crawling ist die Varianz innerhalb und über Seiten hinweg. 5.000 Seiten bedeuten 5.000 verschiedene Strukturen. Es ist bereits eine Herausforderung, die Struktur einer Zielseite zu erfassen und zu pflegen; bei 5.000 ist dies manuell faktisch nicht möglich. Die einzige Lösung ist die Nutzung eines LLM, da es Code-Strukturen gut „versteht“ und diese für formale Strukturen wie z. B. Scrapy Spiders übersetzen kann.

## Bot-Detection

Trotz guter Ergebnisse ist bei dieser Größenordnung eine Fehlerquote von 0 % nicht zu gewährleisten. Einige Anbieter nutzen Maßnahmen gegen Bots, die es entweder zu überwinden oder zu akzeptieren gilt. Eine Fehlertoleranz muss definiert werden, da die Kosten für eine 100-prozentige Abdeckung den Nutzen übersteigen würden. Eine Alternative sind Anbieter wie crawl4ai, die jedoch pro Request Kosten verursachen und eine Third-Party-Abhängigkeit erzeugen.

## Crawling-Zeit

Bei einer Fetch-Time von ~1 Sekunde pro Seite ergibt sich für 5.000 Webseiten eine Dauer von ca. 1,38 Stunden. Ein paralleler Ansatz über Scrapy ist hier zwingend erforderlich, um eine stabile und skalierbare Grundlage zu schaffen.

## Betriebskosten

Am Beispiel von Webseiten wie telekom.de ergeben sich für den HTML-Source-Code ca. 90.000–95.000 Input-Token pro Seite. Basierend auf aktuellen Preisen (z. B. GPT-4o mini) ergäben sich für ein einmaliges Crawling von 5.000 Seiten erhebliche Kosten:

Adressen	Input-Token	Output-Token	Kosten (GPT-5 mini)	Kosten (GPT-5.2)
500	46.250.000	5.000.000	~ 21,56 \$	~ 150,94\$
5000	462.500.000	50.000.000	~ <b>210,00 \$</b>	~ <b>1.509,38 \$</b>

*Hinweis: Schätzungen basieren auf Durchschnittswerten. Durch Code-Bereinigung können 10–30 % eingespart werden.*

**Lösung:** Ich empfehle den Betrieb einer **eigenen LLM-Instanz** (z. B. Granite 1B) auf einem On-Demand-GPU-Anbieter. Auf einer **NVIDIA RTX A4500** (20GB VRAM) lassen sich zwei Instanzen parallel ausführen.

- **Kosten bei 5.000 Adressen:** ca. 4,00 \$(statt 210\$ – 1.500 \$).

# Konzept

## Phase 1: Exploration und Datenbasis

Die Empfängerliste wird algorithmisch durchsucht, um URLs zu erfassen. Für jeden Empfänger wird ein Profil in einer SQL-Datenbank erstellt (Name, Position, Unternehmensbeschreibung). Mittels BeautifulSoup oder httpx werden Quelldaten geladen, bereinigt und per LLM in Scrapy-Schemata übersetzt.

## Phase 2: Überwachung

Die Webseiten werden periodisch auf neue Inhalte geprüft. Bei Überschreitung eines Schwellenwertes für Änderungen wird die Scrapy-Struktur aktualisiert. Bei relevanten inhaltlichen Updates wird automatisch ein E-Mail-Entwurf generiert.

## Phase 3: Nutzung

Integration in den bestehenden E-Mail-Client. Täglich werden die Entwürfe per IMAP in den „Drafts“-Ordner des Nutzers übertragen. Der Nutzer muss die E-Mail nur noch prüfen und versenden.

## Phase 4: Feedback-Loop & Angebotslogik

Sobald ein neues eigenes Angebot definiert wird, gleicht das System dies per semantischem Matching (Vektorsuche) mit der Datenbank ab und erstellt eine „Hot-List“ der relevantesten Leads.

# Technische Realisierung

- **Backend:** Python mit Scrapy und PostgreSQL.
- **KI-Infrastruktur:** Lokale Instanz IBM Granite 1B/H auf NVIDIA RTX A4500.
- **Vektor-Speicher:** Qdrant für semantische Suche und Matching.
- **Integration:** IMAP/SMTP zur Anbindung an den E-Mail-Workflow.

# Angebotsstruktur

## Teil 1: Machbarkeitsstudie (PoC) & Prototyping

Validierung der KI-Pipeline und Erkennungsrate bei 50–100 heterogenen Domains.

- **Ziel:** Nachweis der Qualität. Bei Misserfolg wird die Entwicklung eingestellt.
- **Aufwand:** 80 – 100 Stunden.
- **Investition:** 7.200 € – 9.000 € (Festpreis möglich).

## Teil 2: Integration & Skalierung (MVP)

Aufbau des produktiven Systems für alle Empfänger inklusive Qdrant-Anbindung und IMAP-Integration.

- **Aufwand:** 150 – 200 Stunden.
- **Investition:** 13.500 € – 18.000 € (Abrechnung nach Meilensteinen).

## Teil 3: Produktentwicklung & Wartung

Optionale Erweiterung des MVP zu einer marktreifen Softwarelösung. Aufwand nach Absprache.

*\*Alle Preise zzgl. MwSt. basierend auf einem reduzierten strategischen Stundensatz von 90 €/h für dieses Pilotprojekt. Betriebskosten für die Infrastruktur sind vom Kunden zu tragen.*

## Weiteres Vorgehen

Als nächsten Schritt würde ich gerne gemeinsam mit Ihnen eine Liste exemplarischer E-Mails zusammenstellen, um die ersten Tests durchzuführen. Gerne können wir einen Termin vereinbaren, um Details zu vertiefen.

Mit freundlichen Grüßen



Henrik Lorenzen