# KalliGPT Minihandbuch – Einzeldatei-Embedding

In der Anfangsphase des Projekts empfiehlt es sich, Embeddings für Anträge einzeln zu erstellen. So behält man die volle Kontrolle über Datenkonsistenz, Uploads und Fehlerbehandlung.

## Aufruf des Scripts

Das Script `embed\_kalli\_antraege\_flexibel\_env.py` kann mit einer JSON-Datei als Argument gestartet werden:  
```bash  
python embed\_kalli\_antraege\_flexibel\_env.py ./docs/1234.json  
```  
Dabei muss die angegebene Datei alle Pflichtfelder enthalten:  
`Titel`, `Datum`, `Thema`, `Drucksache`, `Kurzbeschreibung`

## .env-Datei

Die Zugangsdaten für Supabase und OpenAI liegen in der `.env`-Datei und werden über `load\_dotenv()` geladen.  
Beispiel für `.env`:  
```  
SUPABASE\_URL=https://...supabase.co  
SUPABASE\_SERVICE\_ROLE=...  
OPENAI\_API\_KEY=sk-...  
```

## Fehlerbehandlung

Falls ein Feld `null` oder leer ist, kann es beim Zugriff auf Stringfunktionen zu Fehlern kommen. Daher sollten alle Felder vor dem Embedding auf leere Strings gesetzt werden, wenn keine Inhalte vorhanden sind.  
Beispiel:  
```python  
eintrag["inhalt"] = eintrag.get("inhalt") or ""  
```

## Warum Einzelverarbeitung?

• Mehr Übersicht und Kontrolle  
• Direkte Prüfung des Ergebnisses  
• Schnelleres Debugging bei Fehlermeldungen  
  
Ein Batch-Modus wird sinnvoll, wenn viele Anträge gleichzeitig verarbeitet werden müssen. Für den Einstieg ist jedoch der Einzelmodus deutlich robuster.