

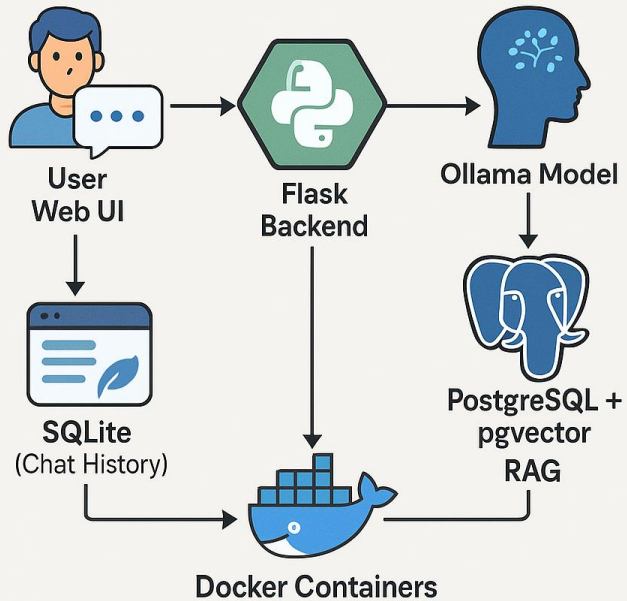
Link-Cloud ERP AI Assistant

Mimari Akışlar ve Fine-tuning
Süreçleri

Oluşturulmuş Akış Şemaları ile

System Architecture Overview

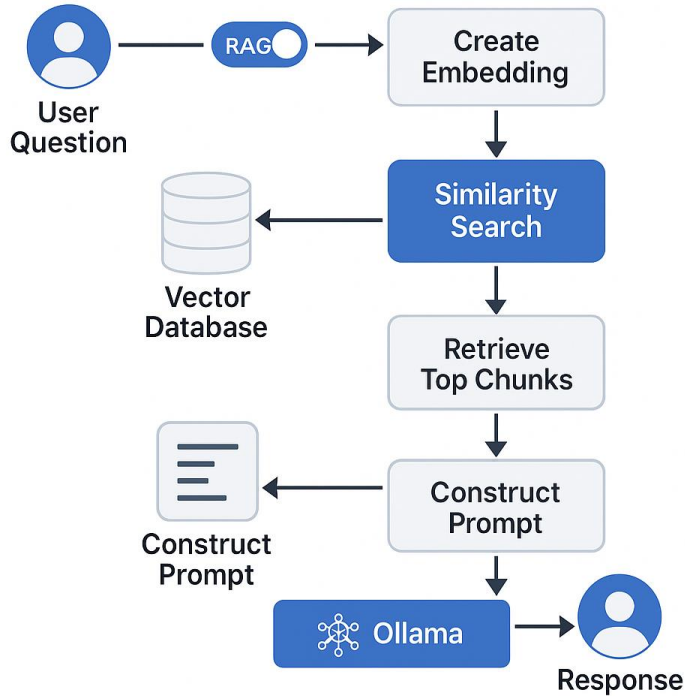
System Architecture Overview



Ana Bileşenler:

Kullanıcı → Web UI →
Flask Backend
Ollama ve RAG
(PostgreSQL +
pgvector) ile bilgi
sorgulama
SQLite ile sohbet
geçmişi ve Docker
konteyner yönetimi

RAG Process Flow

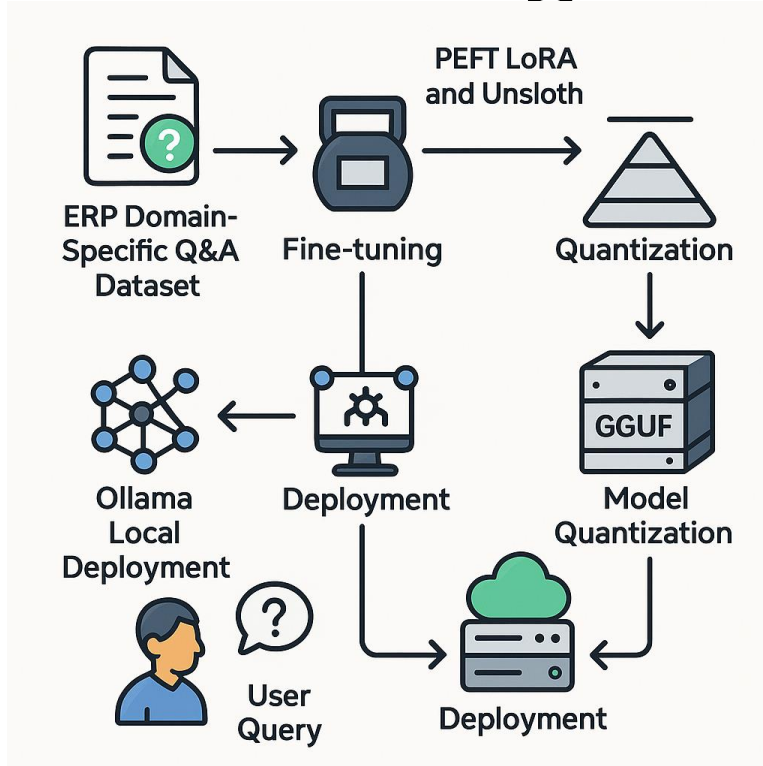


RAG Adım Adım:

Kullanıcı sorusu
embedding'e
dönüştürülür

Vektör veritabanında
similarity search ile
ilgili chunk'lar alınır
Prompt oluşturulur ve
Ollama'ya gönderilir
Model cevabı döner
ve kullanıcıya sunulur

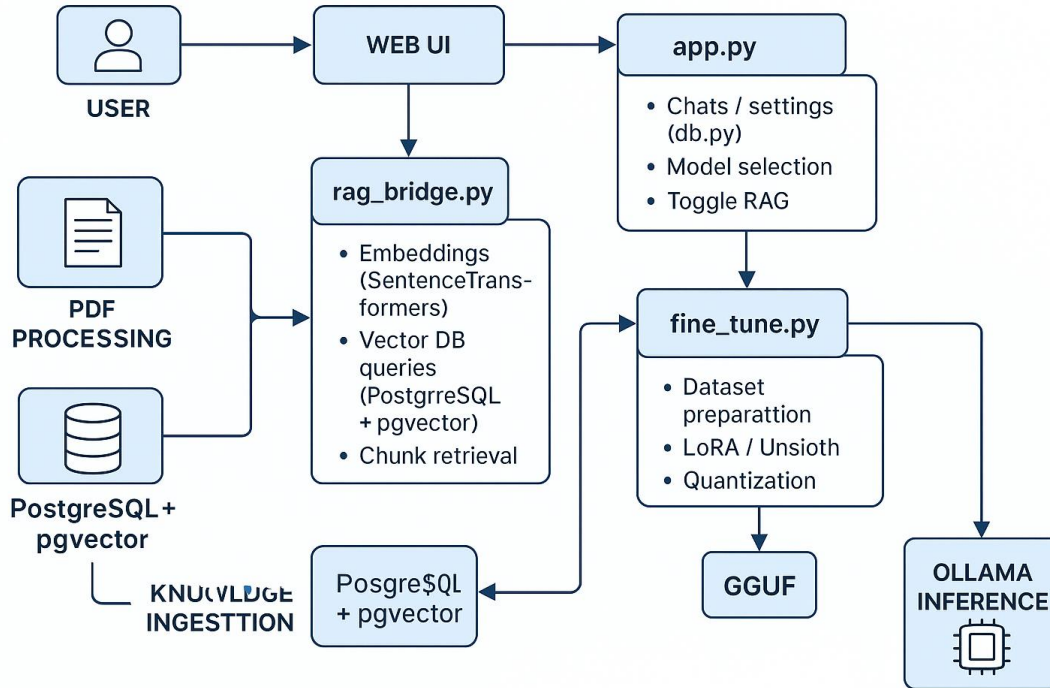
Fine-tuning & Deployment Pipeline



Pipeline Özeti:

ERP'ye özel Q&A veri seti hazırlanır
PEFT LoRA ve Unsloth ile fine-tuning yapılır
Quantization (Q4) ve GGUF dönüşümü ile model optimize edilir
Ollama üzerinden yerel dağıtım ile kullanıcı sorgularına cevap verilir

Python Modülleri ve Akış Diyagramı



Modül Sorumlulukları:

app.py: Kullanıcıdan gelen istekleri yönetir, model/RAG toggle, chat geçmişi için db.py ile etkileşim.

rag_bridge.py: RAG sorgusunu orkestre eder; benzer chunk'ları alır ve prompt oluşturur.

rag_utils.py: Embedding üretimi, vector DB sorguları, chunk retrieval.

fine_tune.py: ERP domain-specific model fine-tuning, LoRA/Unsloth ve quantization işlemleri.