

ReAct Tabanlı RAG Sistemi

1. Projenin Amacı

Bu çalışmanın amacı, Türk mesafeli satış sözleşmeleri ve cayma hakkı kapsamında hazırlanmış hukuki dokümanlar üzerinde çalışan, **ReAct (Reasoning + Acting)** yaklaşımını kullanan bir **RAG (Retrieval-Augmented Generation)** sistemi geliştirmek ve bu sistemin performansını **benchmark testleri** ile değerlendirmektir.

Sistem, kullanıcı sorularına yalnızca yüklenen PDF belgelerinden bilgi çekerek yanıt üretmekte, her cevap için **kaynak PDF adı ve sayfa numarasını** zorunlu kılmaktadır.

2. Kullanılan Veri Seti

Projede iki adet PDF doküman kullanılmıştır:

- **on-bilgilendirme-formu.pdf**
- **cayma_hakki_akademik.pdf**

Bu belgeler:

- PyMuPDF kullanılarak sayfa bazlı metin çıkarımı yapılmış,
- Metinler anlam bütünlüğü korunacak şekilde **chunk'lara bölünmüştür**,
- Basit kelime-frekans temelli bir vektör uzayı ile RAG sistemine aktarılmıştır.

Sistem, yalnızca bu iki PDF dışında hiçbir harici bilgi kaynağı kullanmamaktadır.

3. Sistem Mimarisi (ReAct + RAG)

Sistem aşağıdaki bileşenlerden oluşmaktadır:

- **Retriever:** PDF chunk'ları arasında cosine similarity ile en alakalı parçaların seçilmesi

- **Tool Kullanımı:**
 - ecommerce_policy_search_tool: PDF içi arama
 - calculator: Sayısal hesaplama (gerektiğinde)
- **ReAct Döngüsü:**
 - Thought → Action → Observation → Answer

Model, yanıt üretmeden önce mutlaka ilgili PDF içeriğini tool ile çağrırmak zorundadır.

4. Benchmark Tasarımı

Toplam **50 adet benchmark sorusu** hazırlanmıştır:

4.1 Dataset Soruları (40 adet)

- Cevabı **doğrudan PDF içinde bulunan**
- Hukuki metinden birebir çıkarılabilir sorular

4.2 Reasoning Soruları (10 adet)

- Birden fazla bilgiyi birleştirme
- Süre hesaplama
- İstisna ve koşul yorumlama gerektiren sorular

4.3 Otomatik Skorlama

Her soru için:

- **1.0** → Doğru içerik + kaynak belirtilmiş
- **0.5** → Kısmi doğru + kaynak
- **0.0** → Yanlış veya kaynaksız cevap

Skorlar otomatik olarak hesaplanmış ve CSV dosyalarına kaydedilmiştir.

5. Benchmark Sonuçları (Llama 3.1 8B – Ücretsiz Model)

Benchmark testleri Llama 3.1 8B (Groq – Free) modeli üzerinde başarıyla çalıştırılmıştır.

5.1 Sonuç Tablosu

Model	Dataset (40)	Reasoning (10)	Toplam (50)
Llama 3.1 8B	29 / 40 (%72.5)	4 / 10 (%40.0)	33 / 50 (%66.0)

5.2 Yorum

- Dataset sorularında yüksek başarı, sistemin **doğru doküman retrieval** yaptığını göstermektedir.
- Reasoning sorularında skorun düşmesi, küçük model boyutunun çok adımlı muhakemede sınırlı kaldığını göstermektedir.
- Buna rağmen model, hukuki metinlere dayalı cevap üretiminde tutarlı ve denetlenebilir sonuçlar sunmuştur.

6. Ücretli Model (Gemini) Karşılaştırması – Durum Açıklaması

Benchmark çalışması kapsamında Gemini (ücretli/pro) modeliyle karşılaştırma planlanmıştır.

Ancak yapılan API denemelerinde:

- Bazı Gemini modelleri için **404 NOT_FOUND** (model erişilemez)
- Erişilebilir görünen modeller için **429 RESOURCE_EXHAUSTED (quota exceeded)**

hataları alınmıştır.

Bu nedenle:

- Gemini için **tekrarlanabilir ve loglanabilir** sayısal skor üretilememiştir.
- Hata çıktıları ve deneme logları Colab notebook içerisinde mevcuttur.

Bu çalışmada yalnızca **otomatik ve doğrulanabilir sonuçlar** rapora dahil edilmiştir.

7. Sonuç

Bu projede:

- ReAct + RAG mimarisi başarıyla uygulanmış,
- Hukuki PDF'ler üzerinde kaynaklı cevap üretimi sağlanmış,
- 50 soruluk benchmark seti ile sistem performansı ölçülmüş,
- Sonuçlar CSV formatında kayıt altına alınmıştır.

Model karşılaştırması planlanmış ancak API kısıtları nedeniyle yalnızca Llama 3.1 8B için nicel sonuçlar raporlanabilmiştir.

8. GitHub ve Reproduksiyon

- Tüm kodlar, PDF'ler, benchmark sonuçları ve notebook dosyası GitHub deposunda paylaşılmıştır.
- Çalışma Colab ortamında tekrar çalıştırılabilir niteliktedir.

Ayşe Gençalioğlu

2022556463