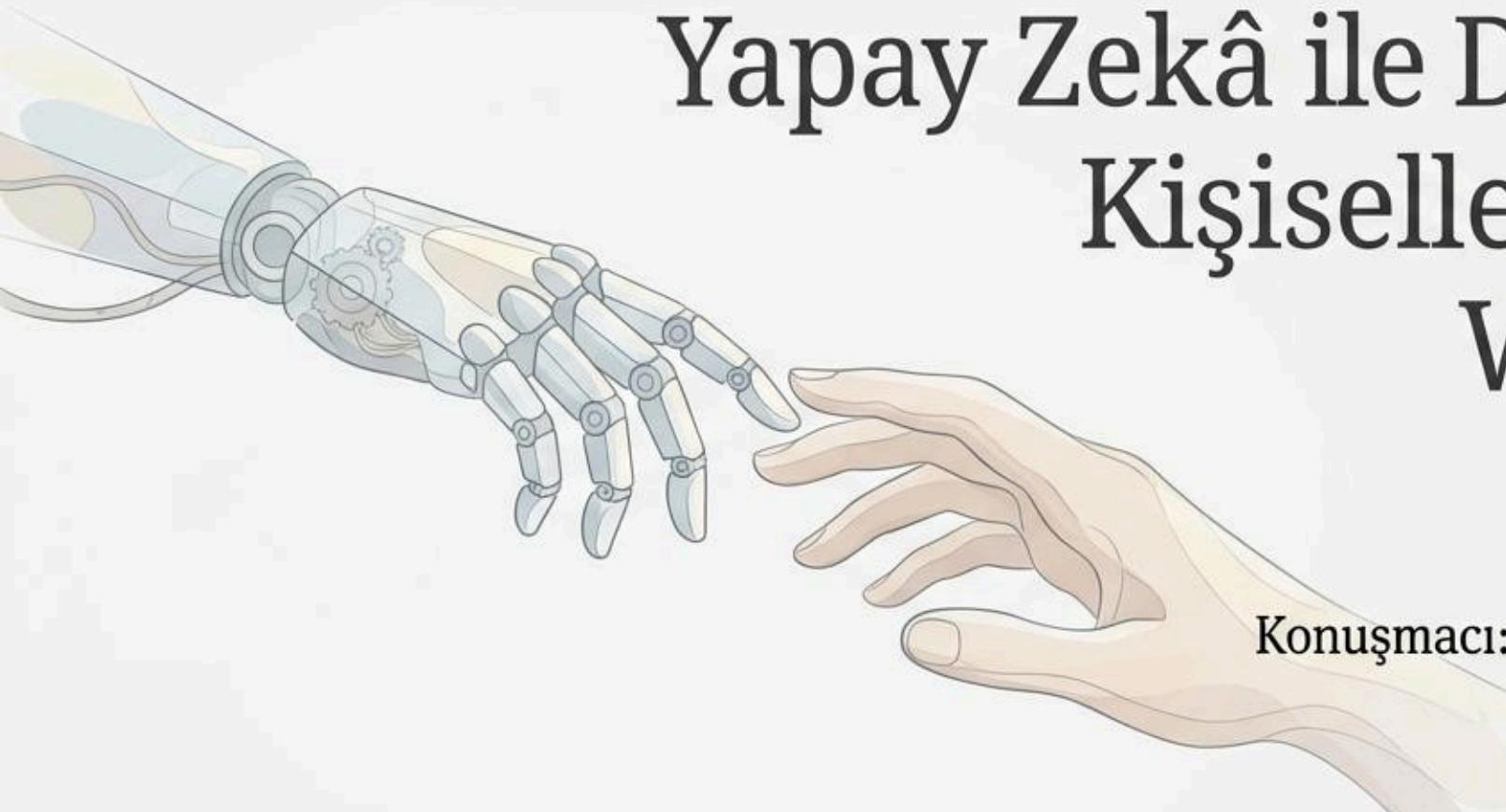




TMMOB
Elektrik Mühendisleri Odası
İstanbul Şubesi

Yapay Zekâ ile Dijital Dönüşüm: Kişiselleştirilmiş Akıllı Web Asistanları



Konuşmacı: Akın ARSLAN / IoT Proje Mühendisi

Moderatör: Müslüm Can ÖZATA

Düzenleyen: TMMOB Elektrik Mühendisleri Odası İstanbul Şubesi



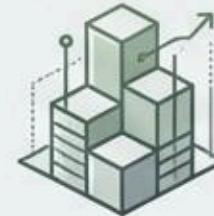
TMMOB

Elektrik Mühendisleri Odası
İstanbul Şubesi

Bugün Neleri Yanıtlayacağız?



- Etkileşim: Web sitemiz kullanıcılarla gerçekten "konuşabilir" mi?



- Geliştirme: Kişiye özel bir web asistanı mimarisi nasıl inşa edilir?



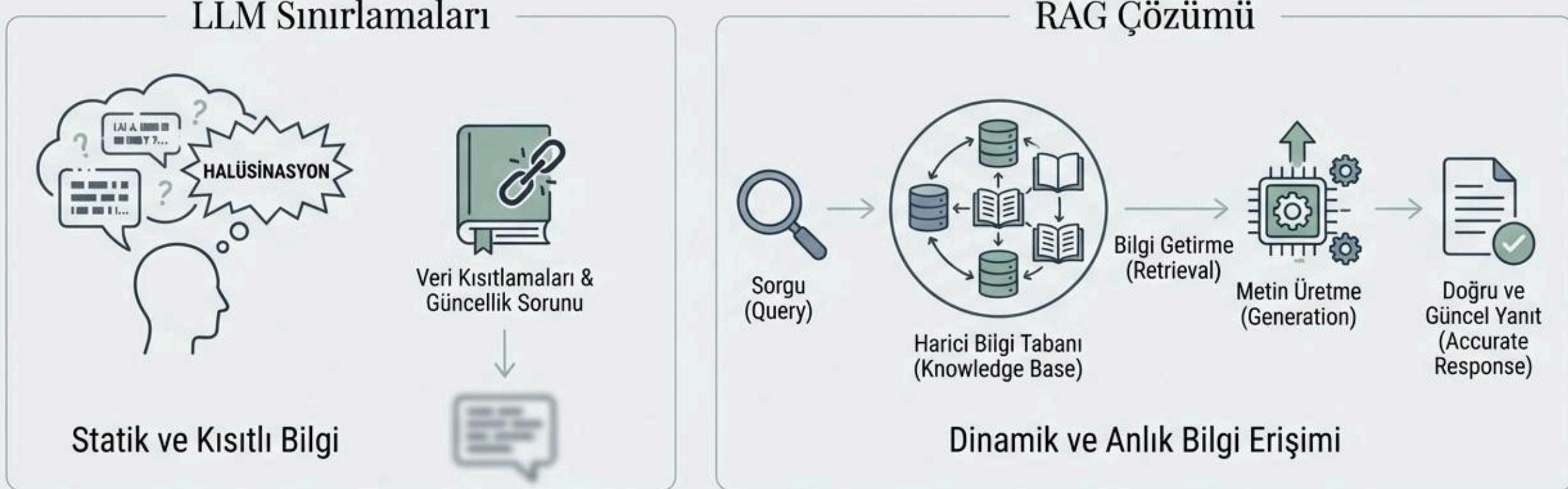
- Entegrasyon: Mevcut web sitelerine LLM entegrasyonu nasıl sağlanır?



- Güvenlik: Kurumsal verinin gizliliği ve güvenliği nasıl korunur?



Neden RAG? (Retrieval-Augmented Generation)



- Teorik Çerçeve: LLM'lerin eğitim verisi kısıtlamaları ve halüsinsiyon riski.
- Akademik Yaklaşım: Modeli yeniden eğitmek (Fine-tuning) yerine, harici bilgi tabanından (Knowledge Base) anlık veri çekme.
- Metodoloji: "Bilgi Getirme" (Retrieval) ve "Metin Üretme" (Generation) aşamalarının ayrıştırılması.

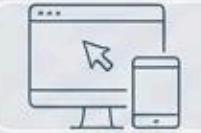


TMMOB

Elektrik Mühendisleri Odası
İstanbul Şubesi

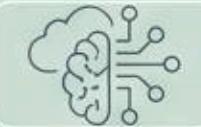
Web-AI Stack Bileşenleri

Sunum Katmanı



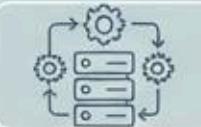
- **Sunum Katmanı:** Web Frontend (Kullanıcı Arayüzü).

Zeka Katmanı



- **Zeka Katmanı:** OpenRouter üzerinden erişilen LLM API'leri.

Servis Katmanı



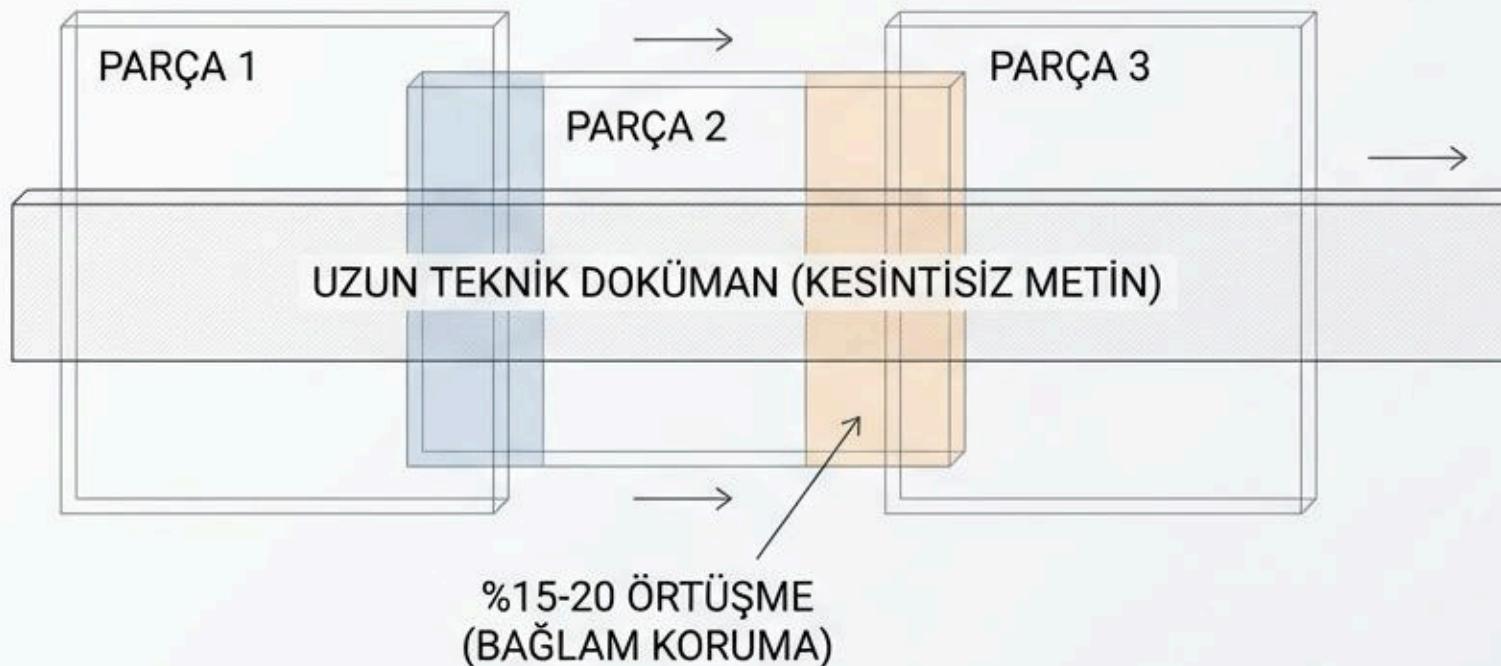
- **Servis Katmanı:** Backend Orchestrator (Orkestrasyon Birimi).

Veri Katmanı



- **Veri Katmanı:** Vektör Veritabanı (Semantic Memory).

Metin Parçalama (Chunking) Stratejileri



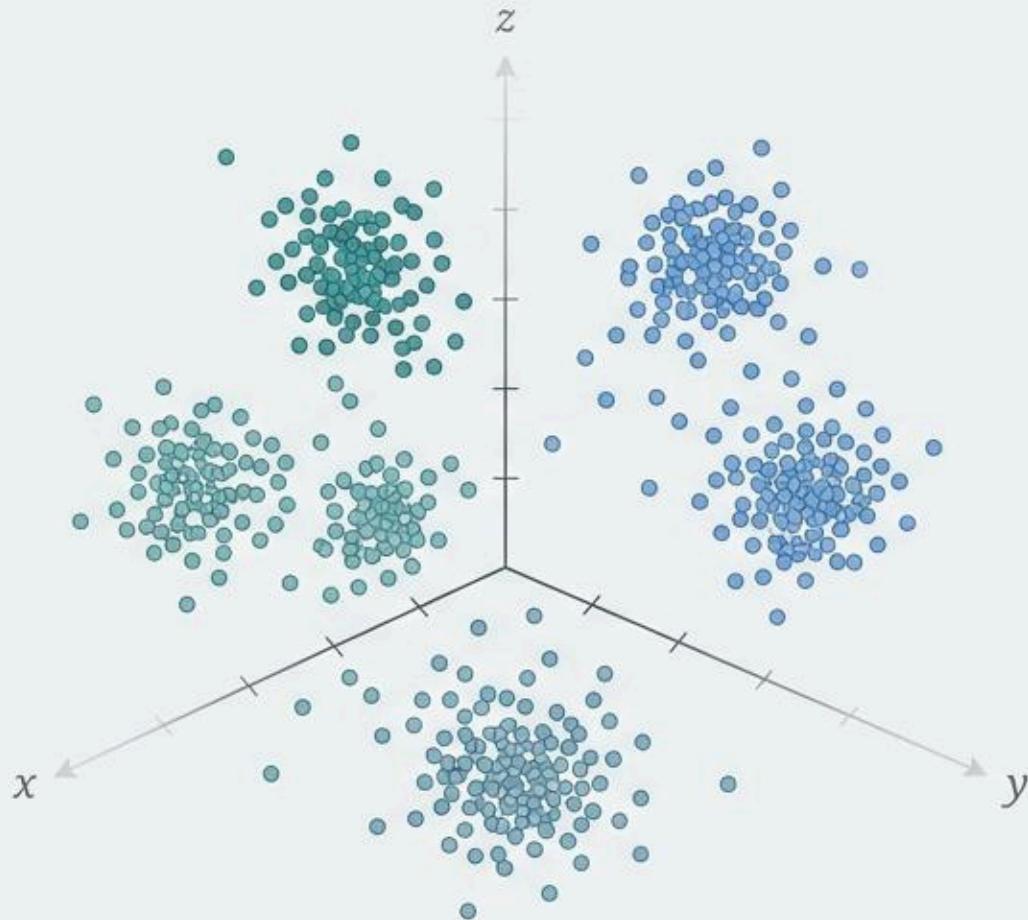
- **Data Chunking:** Uzun teknik dokümanların anlamsal bütünlüğü bozmadan küçük parçalara bölünmesi.
- **Overlap Mantığı:** Bağlamın (Context) korunması için parçalar arasında %15-20 oranında örtüşme bırakılması.



TMMOB

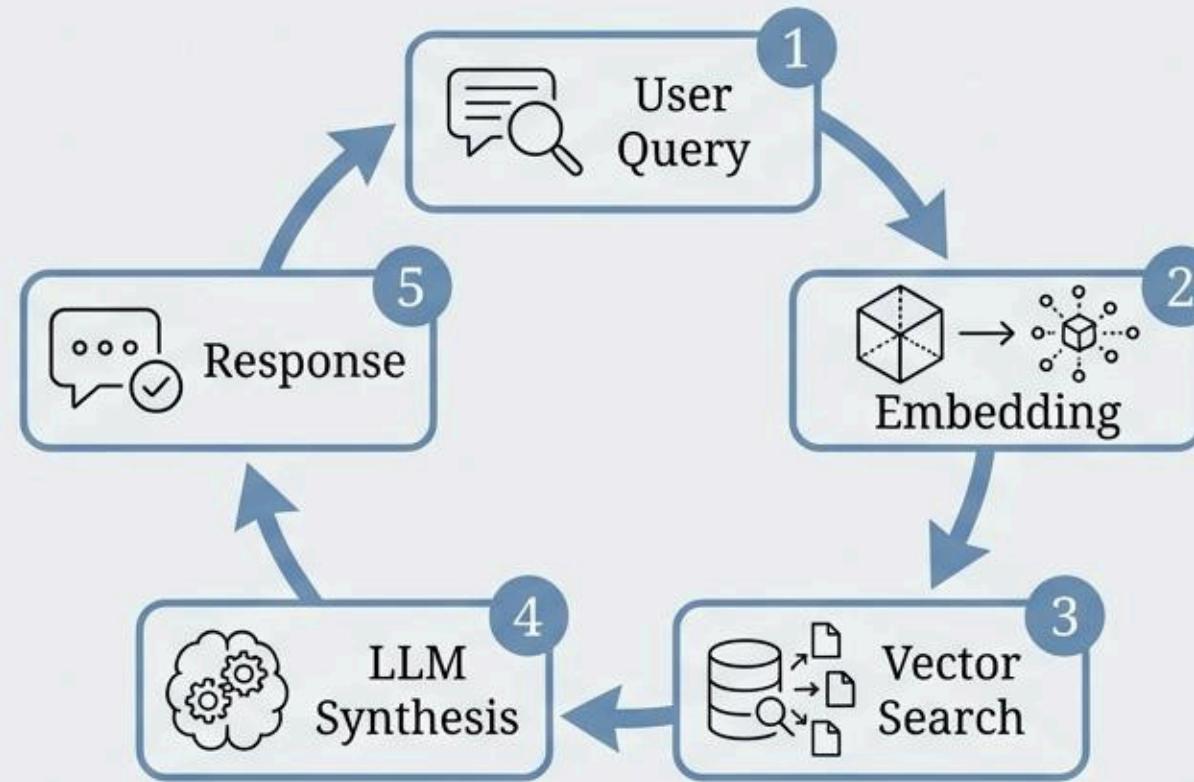
Elektrik Mühendisleri Odası
İstanbul Şubesi

Vektörleştirme ve Semantik Uzay



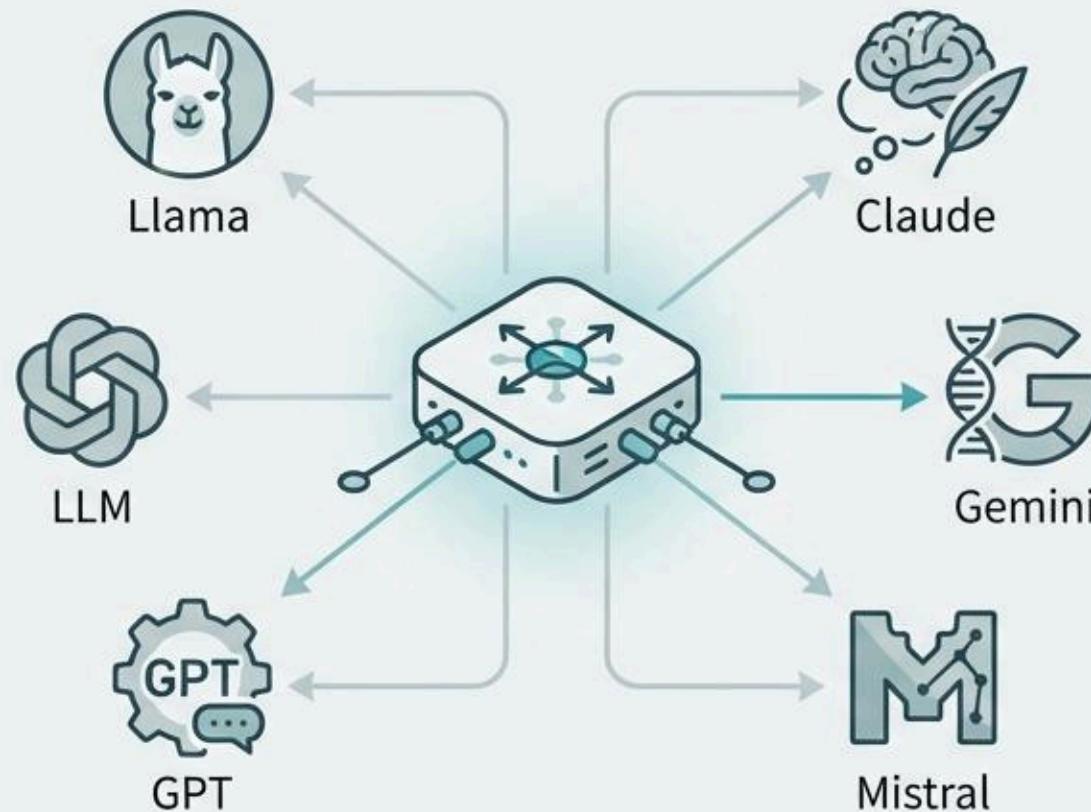
- Embedding Süreci: Metnin sayısal koordinatlara ($f(text) \rightarrow \mathbb{R}^d$) dönüştürülmesi.
- Similarity Search: Kelime eşleşmesi değil, “anlamsal yakınlık” (Cosine Similarity) araması.
- Indexleme: Milyonlarca satır veri içinde milisaniyeler düzeyinde erişim hızı.

Sorgu Döngüsü (The Retrieval Loop)



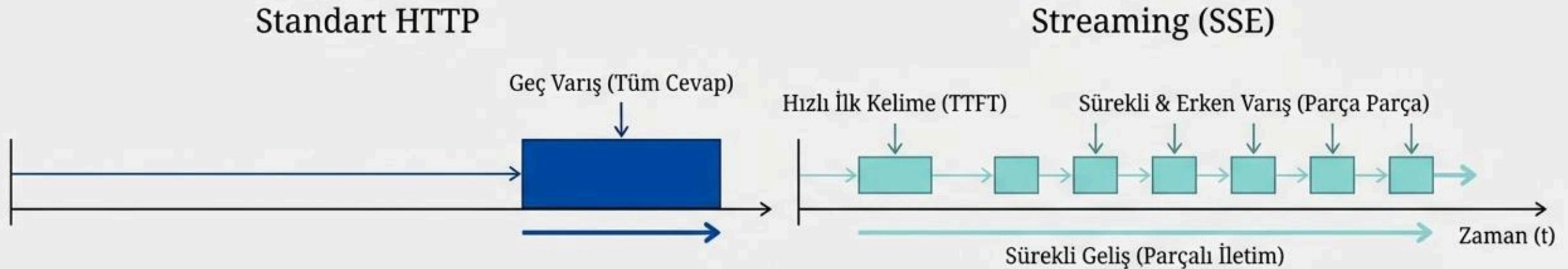
- **Soru:** Kullanıcının doğal dil sorgusu.
- **Arama:** Sorgunun vektöre çevrilip veritabanında en yakın dökümanların bulunması.
- **Entegrasyon:** Bulunan bilginin “**Sistem Promptu**” içeresine eklenmesi.
- **Yanıt:** OpenRouter üzerinden LLM’in cevabı üretmesi.

OpenRouter: Çoklu Model Orkestrasyonu



- **Soyutlama:** Tek bir API üzerinden farklı modelleri kullanabilme esnekliği.
- **Maliyet ve Performans:** Basit sorgularda hafif (small), karmaşık analizlerde güçlü (large) model seçimi.
- **Sistem Güvenilirliği:** Model bazlı hata durumlarında otomatik yedek modele (fallback) geçiş.

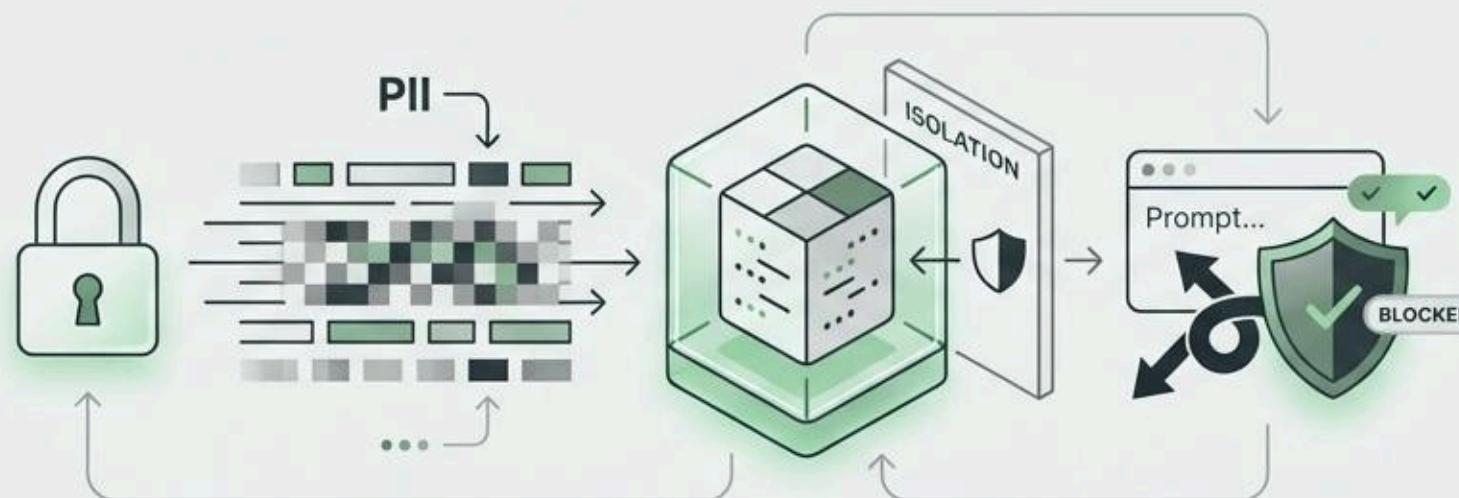
Streaming (SSE) ve Kullanıcı Deneyimi



- **Teknik:** Server-Sent Events (SSE) ile cevabın parça parça (token-by-token) iletilmesi.
- **Performans:** İlk kelimeye kadar geçen sürenin (TTFT) minimize edilmesi.
- **UX:** Kullanıcının bekleme süresi algısının iyileştirilmesi.

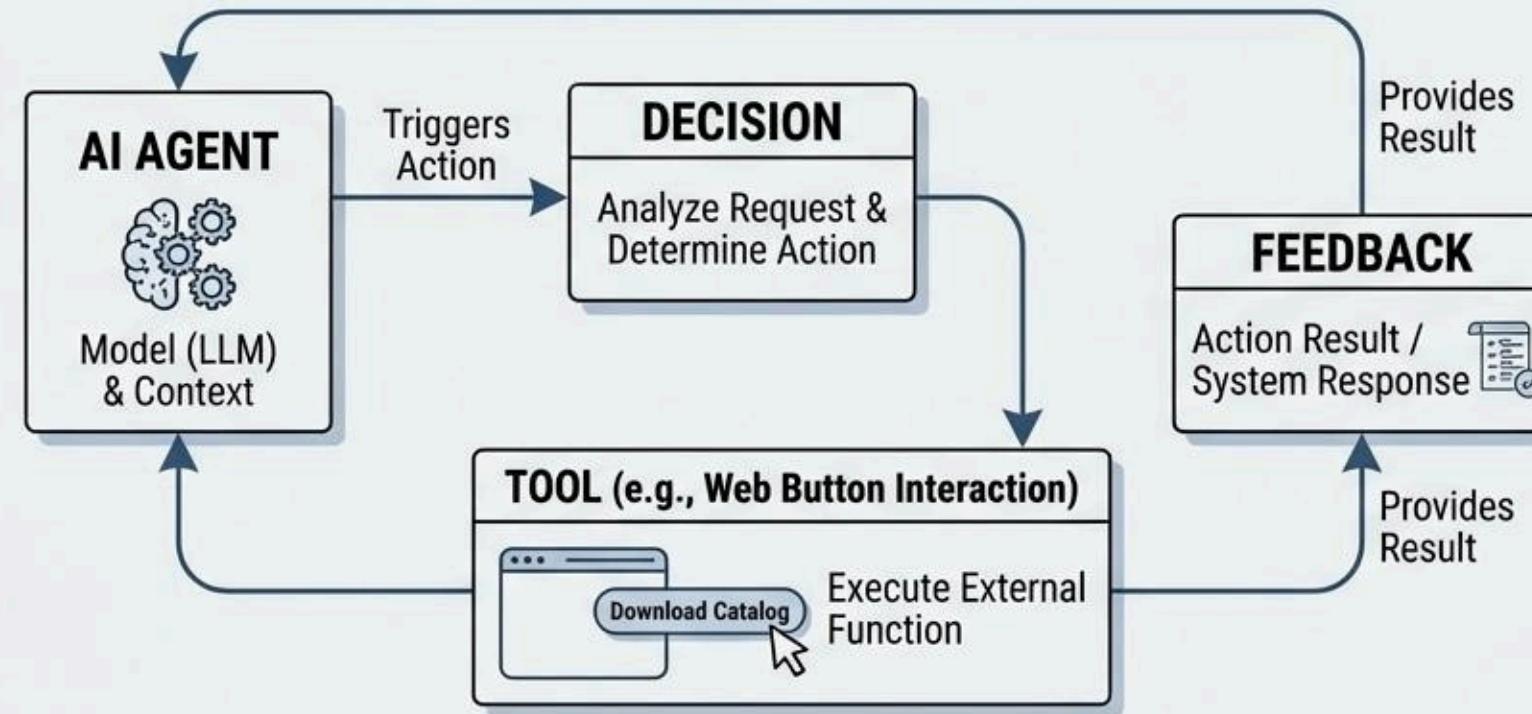


Veri Güvenliği ve Gizlilik



- **PII Masking:** Kişisel verilerin anonimleştirilmesi.
- **Prompt Injection:** Kötü niyetli kullanıcı girdilerine karşı güvenlik kalkanları.
- **Veri İzolasyonu:** Kurumsal API kullanımı ile verinin eğitimde kullanılmasının engellenmesi.

Agentic UI: Tool Calling Mekanizması



- Vizyon: Chatbot'un sadece konuşması değil, web sitesi üzerinde aksiyon alması.
- Mekanizma: Modelin metin yerine "Fonksiyon Çağrısı" (JSON) üretmesi.
- Örnek: "Kataloğu indir" komutuyla asistanın ilgili butonu tetiklemesi.

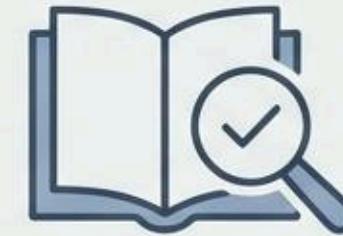
RAG Başarı Metrikleri



- Hassasiyet (Precision): Getirilen dökümanların soruya ilgisi.



- Sadakat (Faithfulness): Yanıtın sadece sağlanan dökümanlara dayanması.



- Akademik Doğrulama: Üretilen çıktıların teknik personel tarafından son kontrolü.

Özet ve Gelecek Projeksiyonu



- **Dönüşüm:** Statik arayüzlerden, bağlamı anlayan akıllı asistanlara geçiş.
- **Demokratikleşme:** Mühendislik dokümantasyonunun asistanlar aracılığıyla yaygınlaşması.



Referanslar ve İletişim



OpenRouter



Pinecone/Chroma



FastAPI

- **Github:** github.com/arslnakin/SmartWebAssistant-RAG
- **İletişim:** Akın ARSLAN / LinkedIn & E-posta bilgileri.

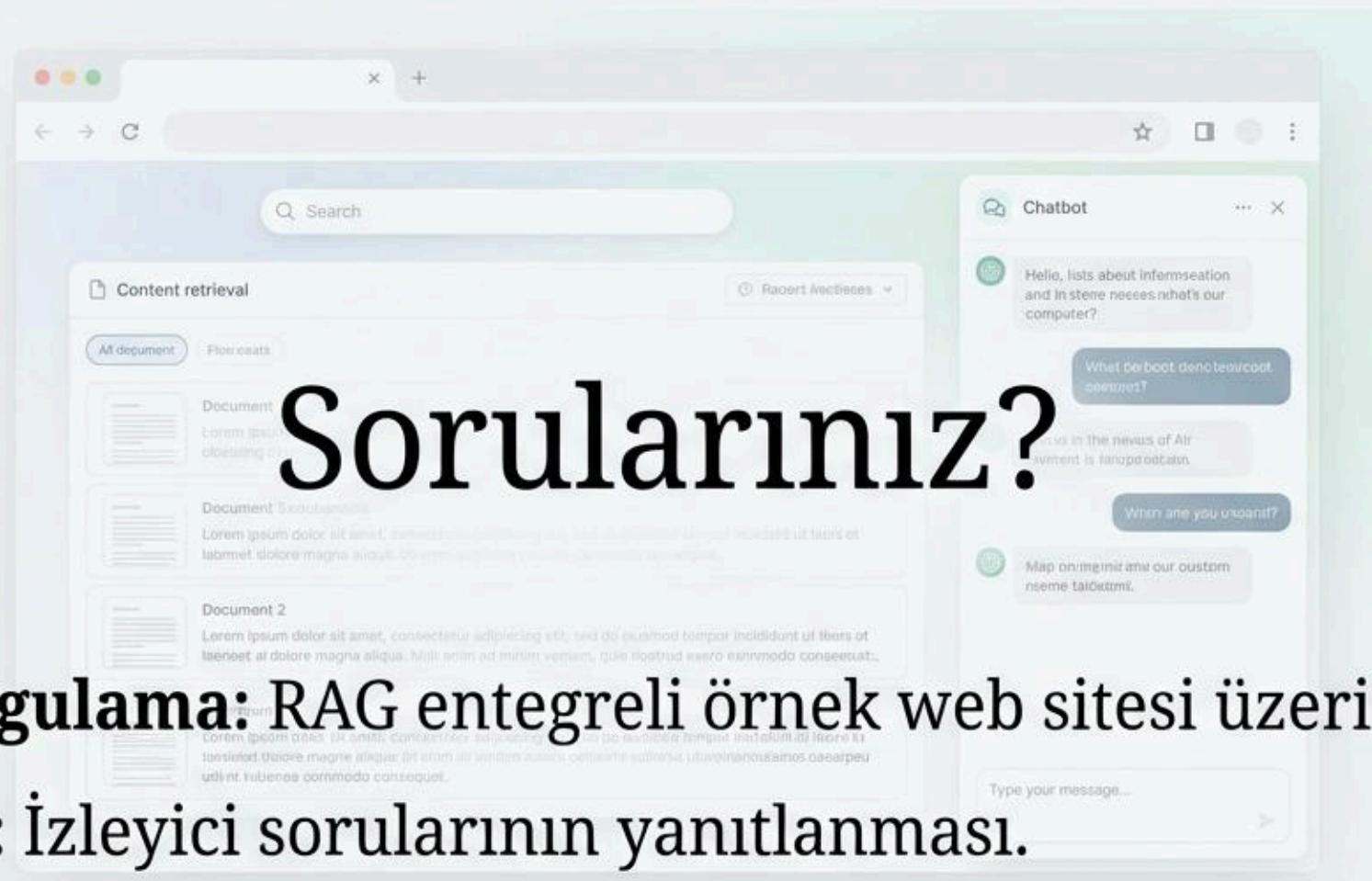


linkedin.com/in/akinaraslan



arslanakin97@gmail.com

Soru - Cevap ve Demo



- **Canlı Uygulama:** RAG entegreli örnek web sitesi üzerinde test.
- **Kapanış:** İzleyici sorularının yanıtlanması.