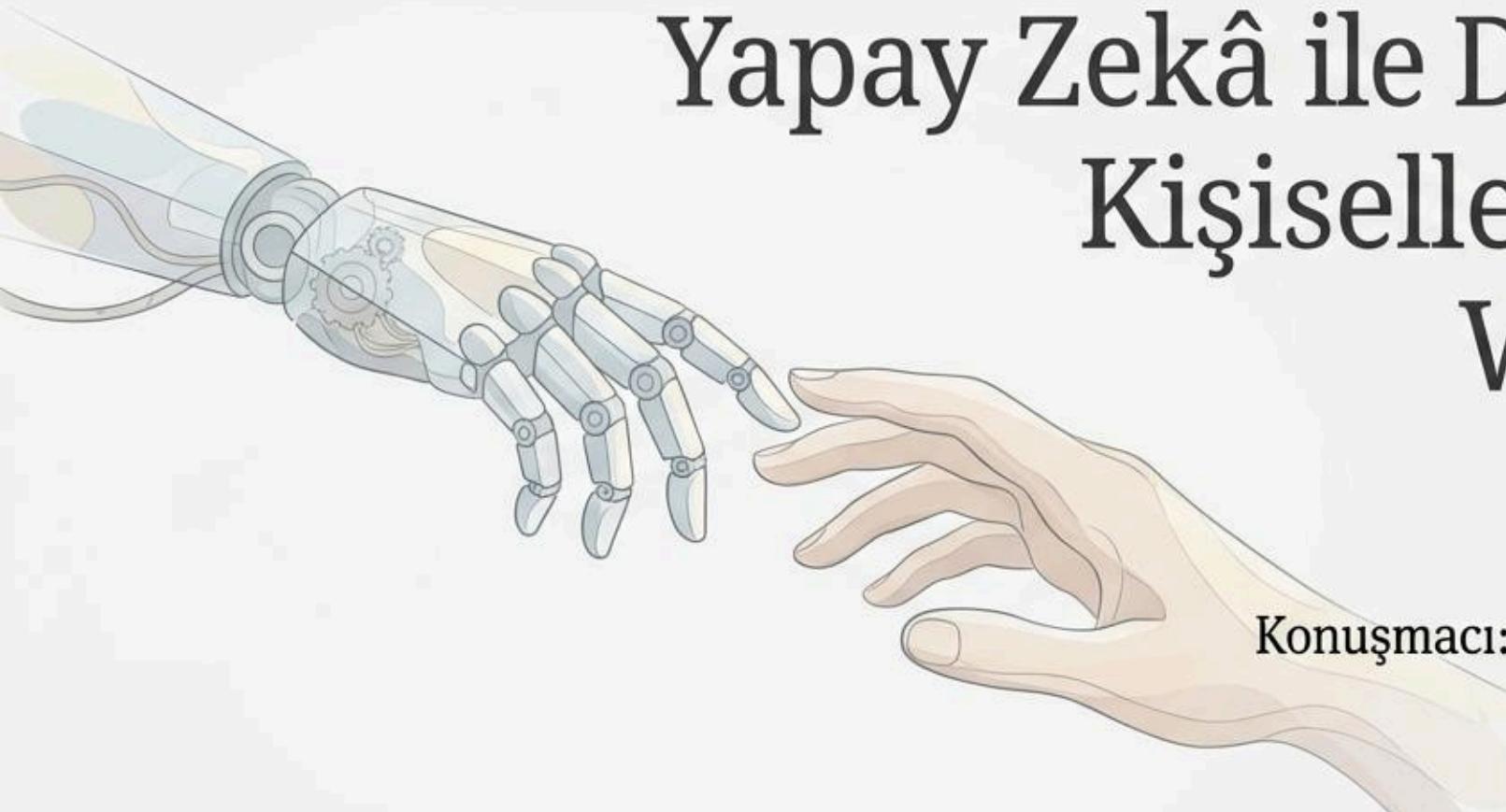




TMMOB
Elektrik Mühendisleri Odası
İstanbul Şubesi

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



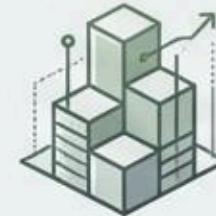
TMMOB

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?



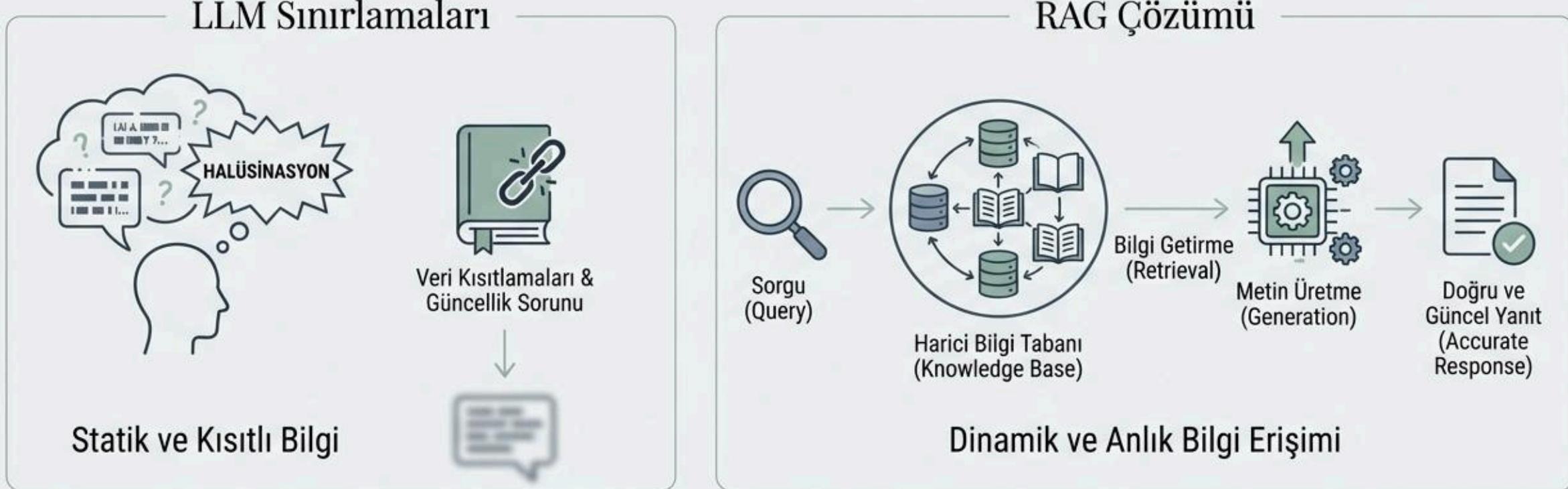
- Entegrasyon: Mevcut web sitelerine LLM entegrasyonu nasıl sağlanır?



- 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.

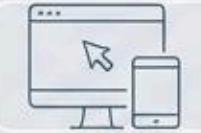


TMMOB

Elektrik Mühendisleri Odası
İstanbul Şubesi

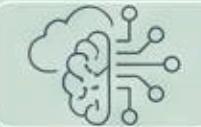
Web-AI Stack Bileşenleri

Sunum Katmanı



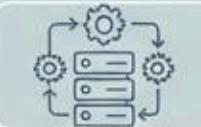
- **Sunum Katmanı:** Web Frontend (Kullanıcı Arayüzü).

Zeka Katmanı



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

Servis Katmanı



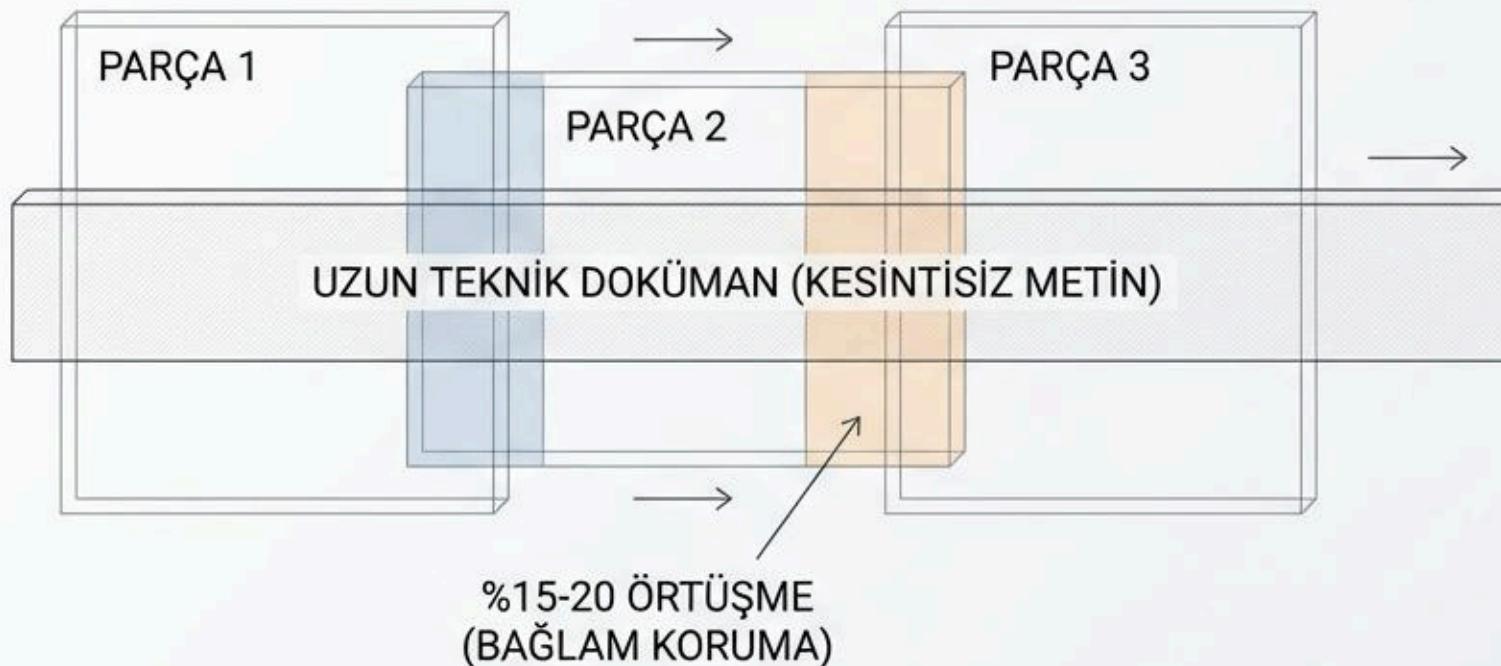
- **Servis Katmanı:** Backend Orchestrator (Orkestrasyon Birimi).

Veri Katmanı



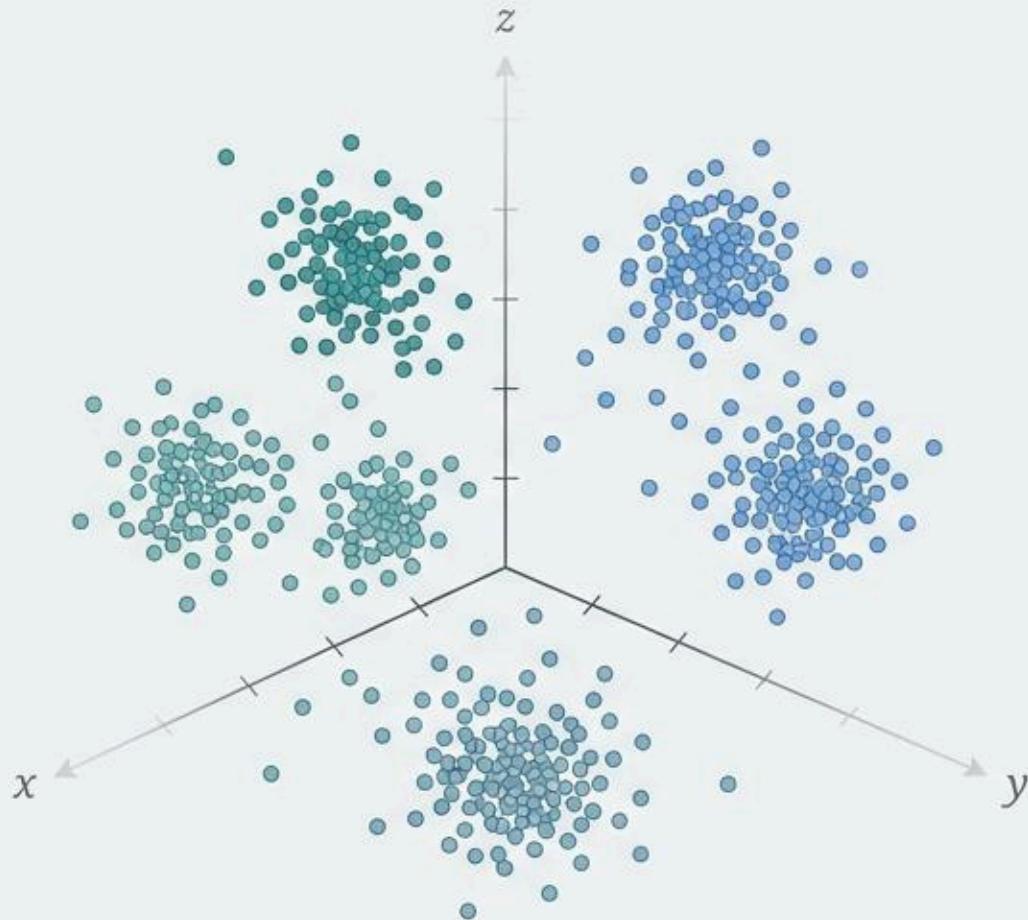
- **Veri Katmanı:** Vektör Veritabanı (Semantic Memory).

Metin Parçalama (Chunking) Stratejileri



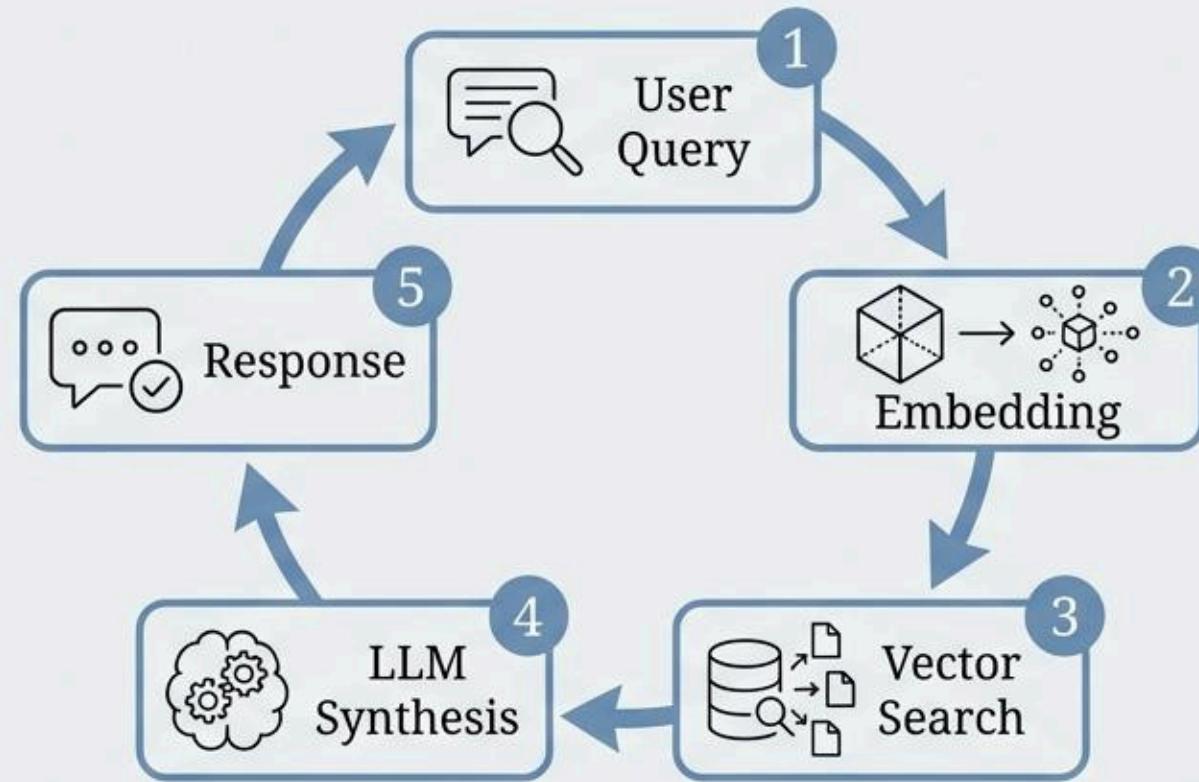
- **Data Chunking:** Uzun teknik dokümanların anlamsal bütünlüğü bozmadan küçük parçalara bölünmesi.
- **Overlap Mantığı:** Bağlamın (Context) korunması için parçalar arasında %15-20 oranında örtüşme bırakılması.

Vektörleştirme ve Semantik Uzay



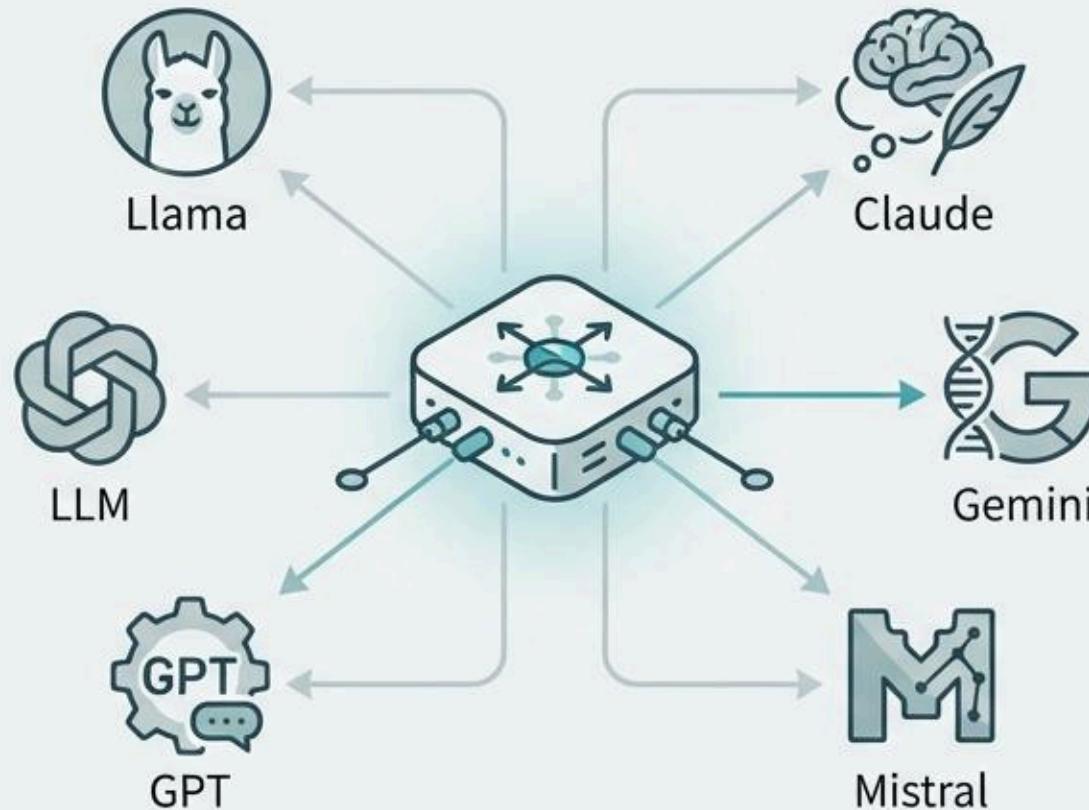
- Embedding Süreci: Metnin sayısal koordinatlara ($f(text) \rightarrow \mathbb{R}^d$) dönüştürülmesi.
- Similarity Search: Kelime eşleşmesi değil, “anlamsal yakınlık” (Cosine Similarity) araması.
- Indexleme: Milyonlarca satır veri içinde milisaniyeler düzeyinde erişim hızı.

Sorgu Döngüsü (The Retrieval Loop)



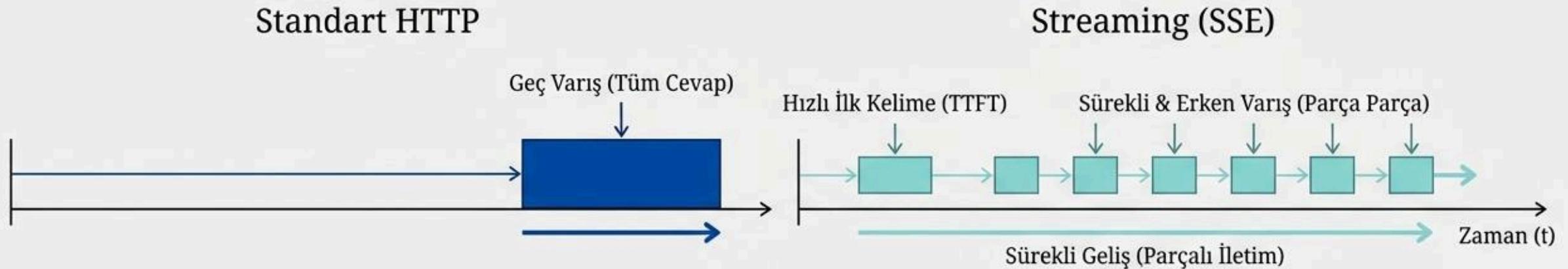
- **Soru:** Kullanıcının doğal dil sorgusu.
- **Arama:** Sorgunun vektöre çevrilip veritabanında en yakın dökümanların bulunması.
- **Entegrasyon:** Bulunan bilginin “**Sistem Promptu**” içeresine eklenmesi.
- **Yanıt:** OpenRouter üzerinden LLM’in cevabı üretmesi.

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.

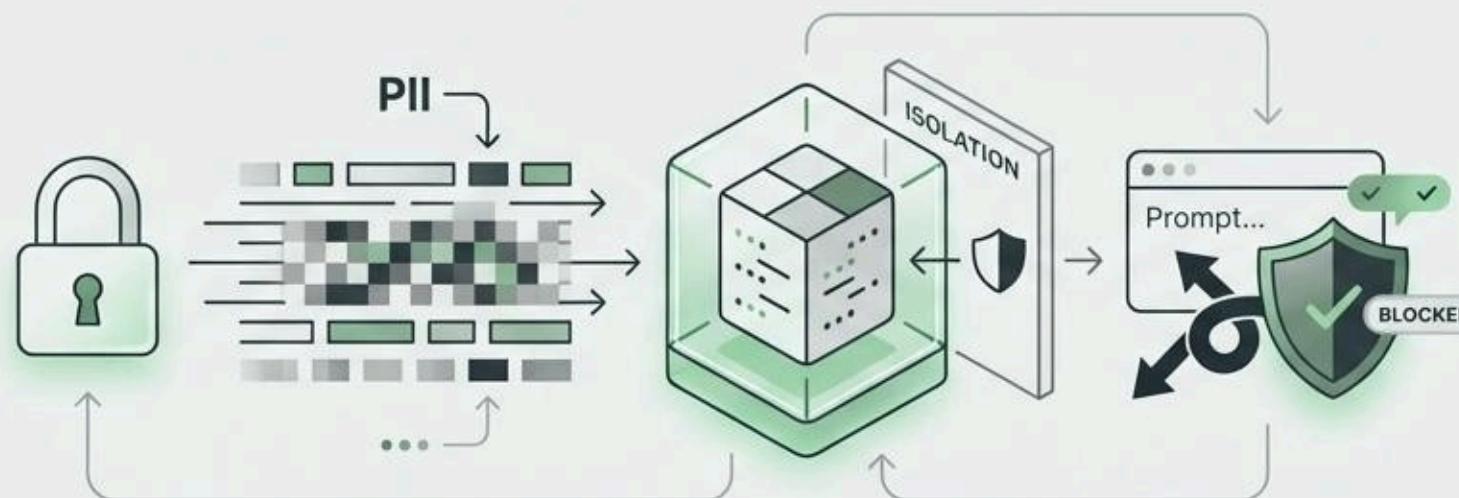
Streaming (SSE) ve Kullanıcı Deneyimi



- **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.

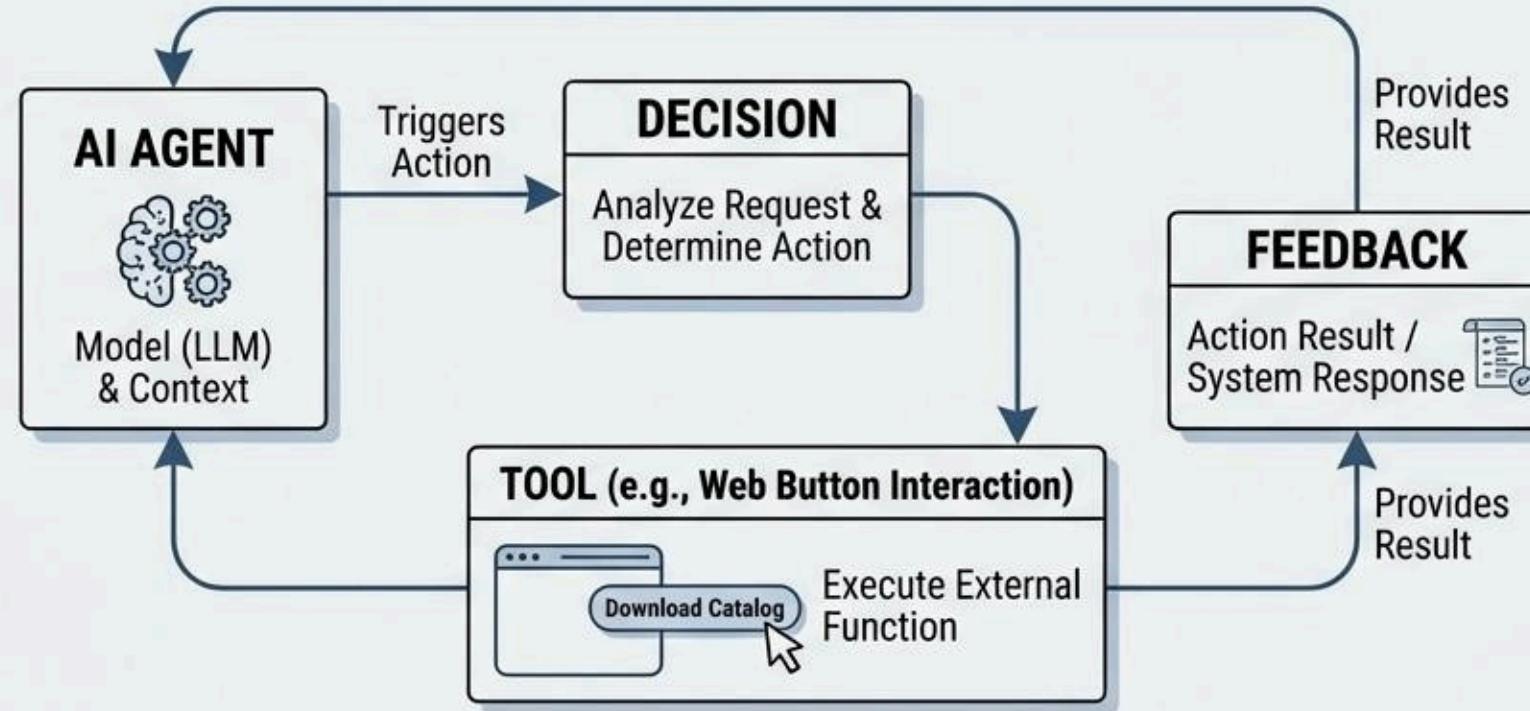


Veri Güvenliği ve Gizlilik



- **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.

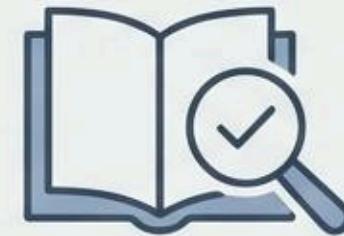
RAG Başarı Metrikleri



- 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.



Soru - Cevap ve Demo

The screenshot shows a web browser window with a light blue header. Inside, there's a search bar with a magnifying glass icon and the word 'Search'. Below it is a 'Content retrieval' section with a 'All document' button and a 'Formats' dropdown set to 'Report Needs'. There are two document thumbnails: 'Document 1' and 'Document 2', both containing placeholder text ('Lorem ipsum dolor sit amet...'). To the right is a 'Chatbot' window with a green profile picture. The chatbot's messages are: 'Hello, lists about information and interests what's our computer?', 'What do you do cool', and 'Map our custom interests'. A user message 'When are you going?' is also visible. At the bottom of the browser window is a text input field with the placeholder 'Type your message...'.

Sorularınız?

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

Geleceğin Enerjisi, Geleceğin Altyapısı

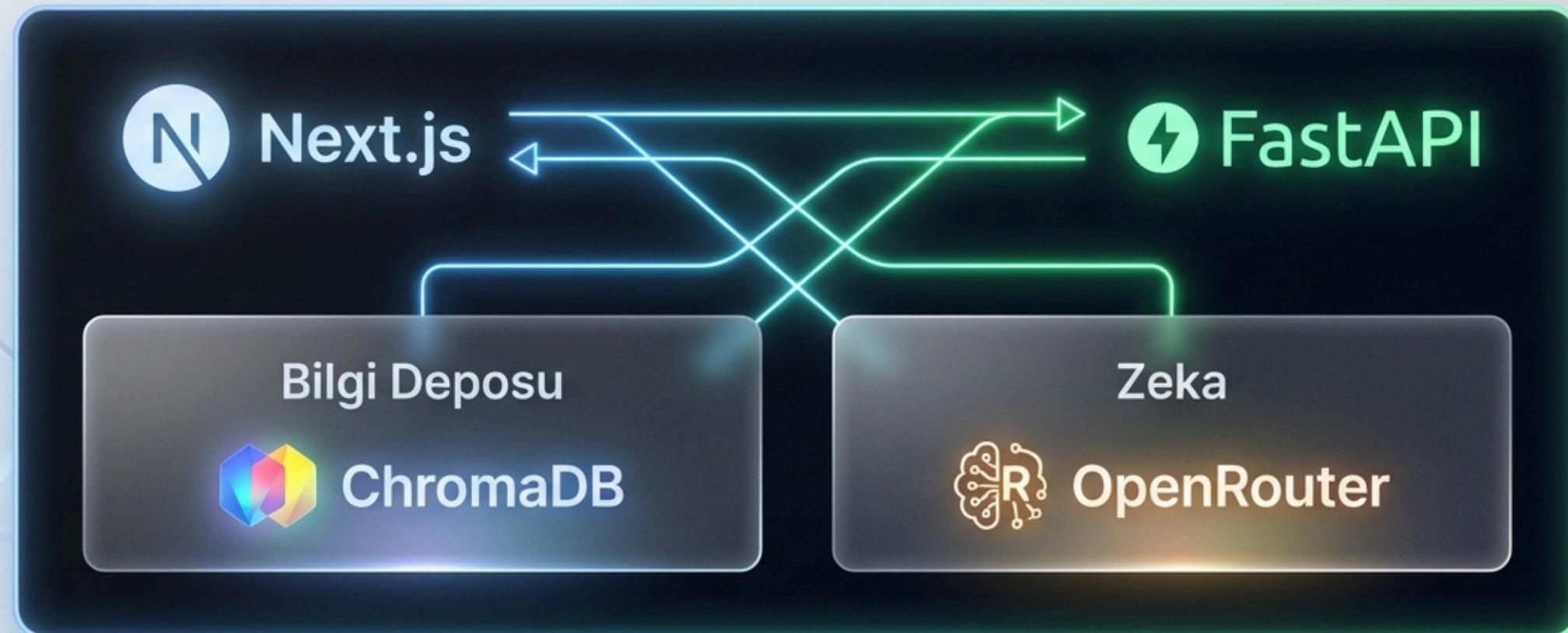
Frontend: Next.js 15+, Tailwind CSS 4.0, Framer Motion (Glassmorphism Arayüz).

Backend: FastAPI (Python) - Yüksek hız ve asenkron yetenekler.

Yapay Zeka Mekanığı: Orkestrasyon: LangChain (Döküman analizi ve RAG zinciri).

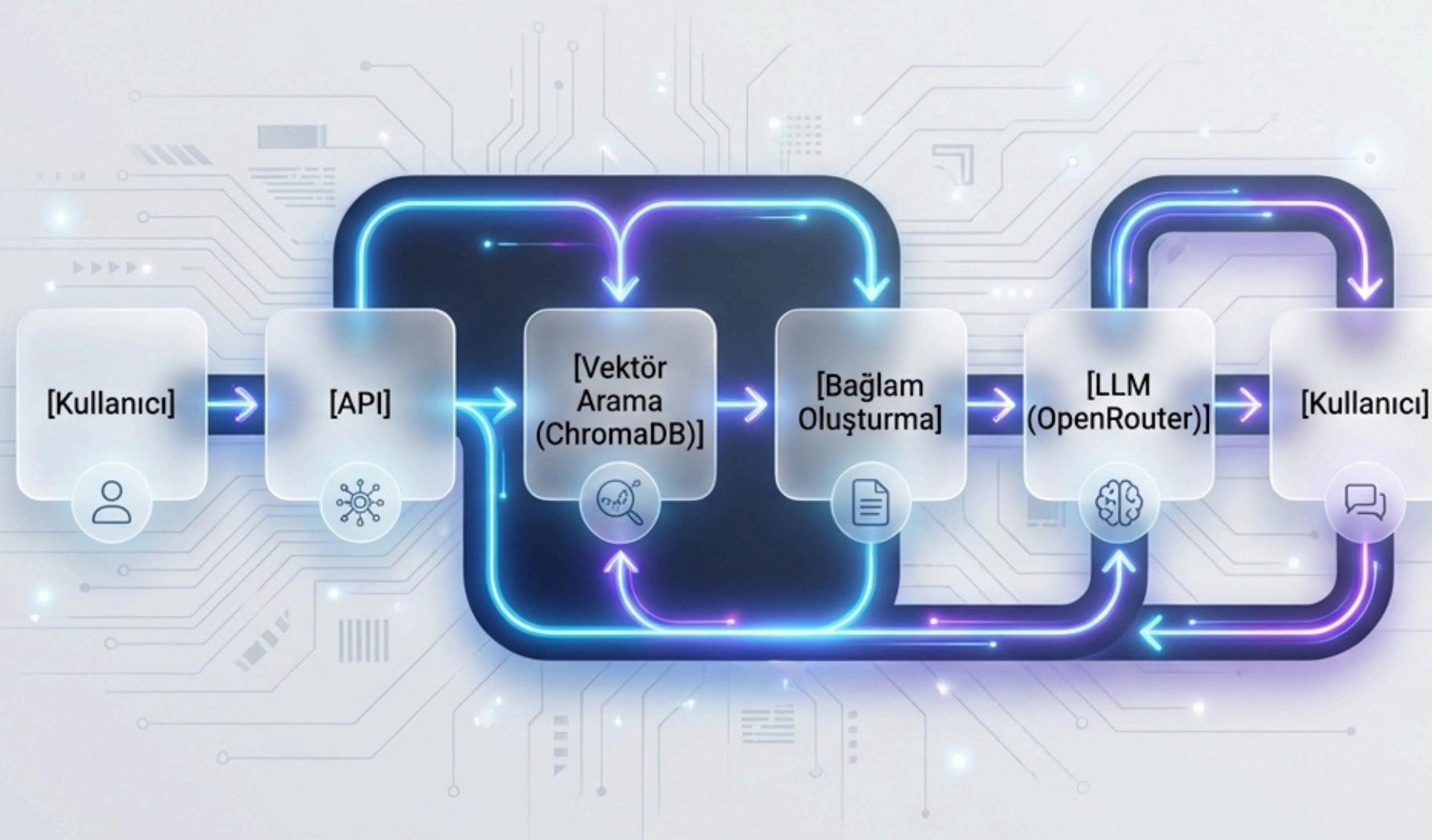
Beyin: OpenRouter (Qwen-3-Max, Gemini-2.0).

Hafıza: ChromaDB (Vektör Veritabanı).



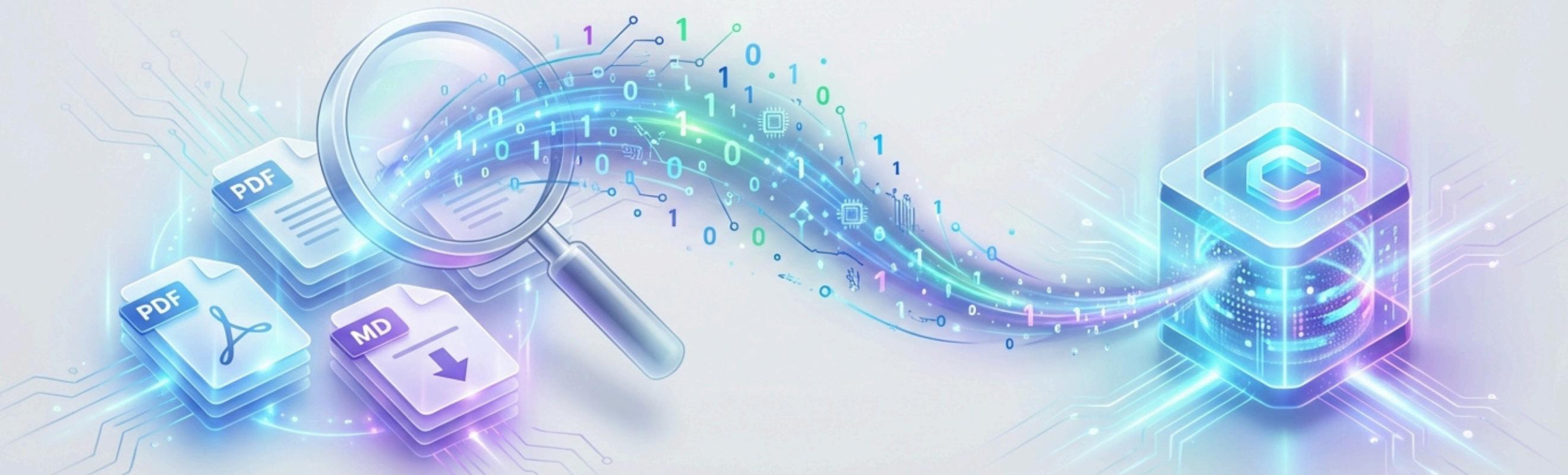


Mesajın Yolculuğu: Uçtan Uca Akış



- Giriş:** Kullanıcı Chatbot UI üzerinden "Trafo bakımı yapıyor musunuz?" sorusunu sorar.
- Sorgu İşleme:** Backend, soruyu alır ve ChromaDB içinde anlamsal (semantic) arama yapar.
- Bağlam (Context) Yakalama:** Teknik dökümanlardan en alakalı 500 karakterlik parçalar (chunk) bulunur.
- Prompt İnşası:** [Sistem Prompt'u] + [Döküman Kesitleri] + [Sohbet Geçmişi] birleştirilir.
- LLM Yanıtı:** OpenRouter üzerinden model (Qwen/Gemini) "Kısıtlı Bilgi" kuralıyla yanıt üretir.
- Çıkış:** Şık Markdown formatında cevap kullanıcıya ulaşır.

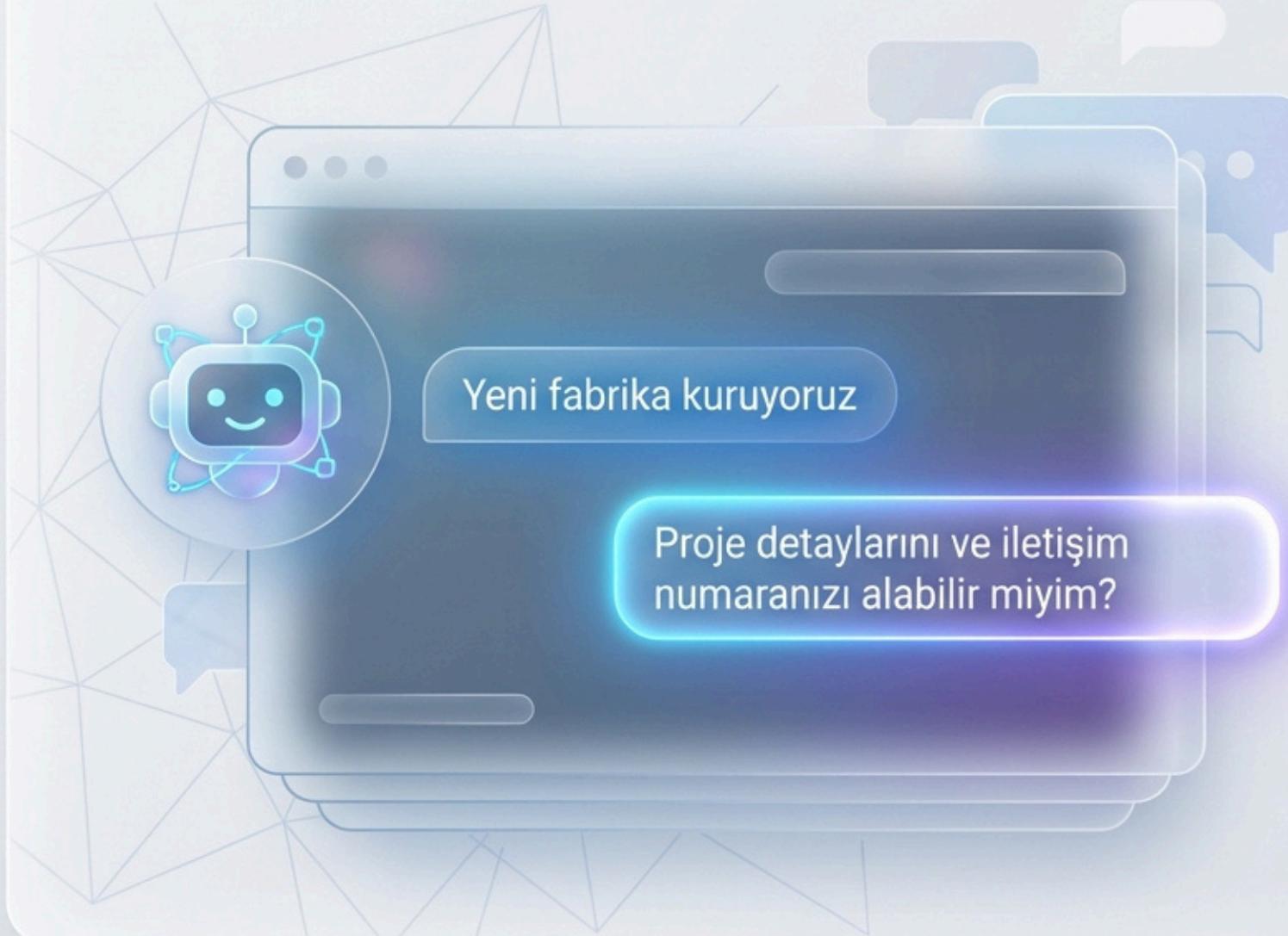
Halüsinasyon Yok, Gerçek Bilgi Var



- ▶ **Döküman Analizi:** Şirketin PDF ve Markdown dökümanları akıllı parçalara bölünür.
- ▶ **Anlamsal Embedding:** Cümlelerin sadece harfleri değil, "mühendislik anlamları" vektörlere dönüştürülür.
- ▶ **Similarity Search:** Kullanıcı "bakım" dediğinde, sistem dökümanlardaki "teknik periyodik hizmetler" bölümünü saniyeler içinde bulur.
- ▶ **Güvenlik Filtresi:** Döküman dışı bilgi verilmesi engellenerek "Kurumsal Güven" korunur.



Sadece Cevap Vermez, Satış Fırsatlarını Yakalar



Teknik Satış Asistanı: AI, pasif bir bot değil, aktif bir satış temsilcisi gibi davranır.

Lead Generation: Kullanıcı bir projeden bahsettiği an (Örn: "Yeni fabrika kuruyoruz"), asistan otomatik olarak iletişim bilgilerini talep eder.

Kalıcı Bağlam (Context): conversation_history sayesinde kullanıcı "Peki fiyatı ne?" dediğinde, asistan bir önceki cümlede "Trafo" konuşulduğunu bilir.

Profesyonel Fallback: Bilgi bulamadığında "Bilmiyorum" demez, "Sizi uzman mühendisimize bağlayalım" diyerek köprü kurar.



Kararlı ve Esnek Altyapı

Stateless API: Backend her isteği bağımsız ve hızlı işler, performans kaybı yaşanmaz.



Native Requests: Karmaşık kütüphaneler yerine doğrudan HTTP protokolü (Requests) ile OpenRouter'a bağlanarak maksimum uyumluluk sağlanır.



CORS & Veri Güvenliği: Sadece yetkili frontend uygulamalarının erişebileceği güvenlik katmanları.



Rate Limiting Kavramı: Model kotalarının ve API limitlerinin akıllı yönetimi.



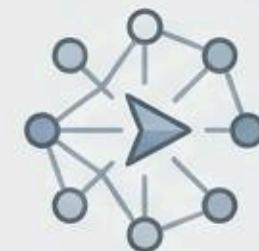
Yarının İhtiyaçlarına Hazır Portföy

- Dinamik Veri Ekleme: Yeni projeler veya teknik dosyalar eklendiğinde sistem saniyeler içinde kendini günceller (Auto-indexing).
- Çoklu Model Desteği: Tek bir ayar ile Qwen'den GPT-4'e veya Gemini'ye anlık geçiş imkanı.

- Analistik Entegrasyonu: Hangi konuların en çok sorulduğunu raporlayabilme potansiyeli.
- Omnichannel: Web sitesinden sonra WhatsApp veya Mobil uygulamaya kolayca entegre edilebilir altyapı.



Referanslar ve İletişim



OpenRouter

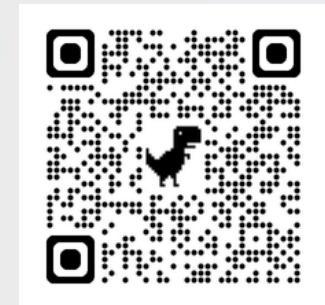


Pinecone/Chroma



FastAPI

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



linkedin.com/in/akinaraslan



arslanakin97@gmail.com