

Performans ve Dil Kabiliyeti

Yanıt Doğruluğu: Sistem, kullanıcı inputu method ve lokasyon detayı içerdiğinde istenilen cevapları verebiliyor. Bu, sistemin belirli koşullarda yüksek doğrulukla çalıştığını gösteriyor.

Dil Kapsamı: Sistem yalnızca İngilizce sorgularla çalışıyor, bu nedenle çok dilli destek sağlamıyor. Bu, dil kabiliyetinin sınırlı olduğu anlamına geliyor.

Yanıt Üretimi: Vector eşleştirmesi yapıldıktan sonra ilgili loglar Gpt-4o-mini ile harmanlanarak kullanıcıya yanıtlar sunuluyor. Bu yöntem, yanıtların log verisi ile sınırlı kaldığını, dolayısıyla yaratıcılığın sınırlı olduğunu gösteriyor.

Yanıtın Log Verisi ile Sınırlılığı: Gpt-4o-mini'nin çıktılarında herhangi bir halüsinasyon (yanıltıcı bilgi üretimi) olmuyor, ancak yanıtlar log datası çevresinde kalıyor. Yaratıcılık ve esneklik, bu sınırlama nedeniyle düşük kalıyor.

Veri Yetersizliği ve Etkileri

Vector Veritabanı: Vector veritabanında yeterli verinin olmaması durumunda, en yakın vector istenilen logdan farklılık gösterebiliyor. Bu, bazı durumlarda hatalı veya yetersiz yanıtların verilmesine neden olabilir.

Veri Hacmi: Sistem, yeterli detaya ve hacme sahip veri ile daha verimli çalışıyor. Yani, veri miktarı ve kalitesi arttıkça sistemin performansı da artıyor.

Genel Değerlendirme

Sistem, özellikle yeterli ve ilgili verilerle beslenirse, kullanıcı sorgularına oldukça doğru ve ilgili yanıtlar verebiliyor. Dil kabiliyeti ve veri hacmi açısından bazı sınırlamalar mevcut olsa da, doğru koşullarda yüksek performans sağlıyor. Sistem geliştirilmeye devam edildikçe, bu sınırlamaların üstesinden gelmek ve daha geniş bir dil desteği sunmak mümkün olabilir.

Veri Zenginleştirme ve Log Kaynaklarının Genişletilmesi

- Detaylı Log Verileri:** Standart log kayıtlarından ziyade, sunucudaki tüm logları daha detaylı metadata ile zenginleştirmek, sistemin daha güçlü ve anlamlı embeddings oluşturmasını sağlar. IP adresi, zaman damgası, HTTP yöntemleri coğrafi konum gibi mevcut bilgilerin ötesinde, kullanıcı agentı, oturum bilgileri gibi ek metadatalar eklenebilir. Bu zenginleştirme, sorgularda daha hassas eşleşmeler sağlayacak ve sonuçların doğruluğunu artıracaktır.

Gelişmiş Preprocessing ve Vektör Eşleştirme

- Gelişmiş Embedding Modeli:** Zenginleştirilmiş log verileri, daha gelişmiş bir embedding modeliyle vektör haline getirilebilir. OpenAI embedding modelleri ile test edilip şuan kullanılandan farklı bir model denenmeli.

RAG Modelinin İyileştirilmesi

- **Daha Hassas Vektör Eşleştirme:** Embedding teknikleri daha detaylı araştırılıp daha uygun bir hassas vektör eşleştirme sağlanabilir. Bu, kullanıcı sorgularının doğru log kayıtlarıyla eşleşmesini ve daha doğru cevapların verilmesini sağlar.
- **Bağlamsal Anlam Zenginleştirme:** Log kayıtlarından sadece içerik değil, bağlam da çıkarılabilir. Kullanıcıların niyetini ve duygularını anlayarak daha bağlamsal yanıtlar üretilebilir. Bunun için kullanıcı inputu direkt vektöre çevrilmeden bir dil modeli ile sorguya daha uygun hale getirilebilir.

Gpt-4o-mini Yanıtlarının Kalibre Edilmesi

- **Yanıt Formatının Detaylandırılması:** Gpt-4o-mini'nin yanıt formatını detaylandırarak, daha yapılandırılmış ve anlaşılır cevaplar oluşturabilir. Yanıtların formatı, belirli bir standarda oturtulup kalibre edilerek hata payı minimize edilebilir.
- **Kural Tabanlı ve Özelleştirilmiş Cevaplar:** Özellikle sık sorulan sorular için önceden tanımlanmış kurallar ekleyebilir, bu sayede bu tür sorulara daha tutarlı ve doğru cevaplar verilebilir.

Performans Takibi ve Geri Bildirim

- **Gerçek Zamanlı İzleme:** Sistem performansını gerçek zamanlı izleyerek, belirli sorguların neden hatalı sonuçlar verdiğini hızlıca tespit edip düzeltebilirsiniz. Bu, sistemin sürekli olarak optimize edilmesine katkı sağlar.
- **Kullanıcı Geri Bildirimi:** Kullanıcı geri bildirimlerini toplamak ve bu geri bildirimler doğrultusunda sistemi iyileştirmek, yanıt kalitesini artırmanın etkili bir yoludur.

Bu öneriler, sistemin daha detaylı, hassas ve kullanıcıya yönelik cevaplar vermesini sağlayacaktır. Log verilerini zenginleştirmek ve daha güçlü bir embedding modeli kullanmak, sistemin genel performansında kayda değer bir iyileşme sağlayabilir. Ayrıca, Gpt-4o-mini yanıtlarının formatını detaylandırmak ve kalibrasyon yapmak, hata payını minimize eder ve kullanıcı deneyimini geliştirir.