

Endüstriyel RAG: doğru yanıt nasıl ölçeriz?

PLC kılavuzlarından gelen bir yanıtın işe yararlılığını; kaynak, bağlam ve yanıt vermeme davranışıyla birlikte değerlendirin.

Kısaca

- Arama başarısı yanıt doğruluğunun yerine geçmez, bu yüzden üretilen yanıtta her teknik iddia bir kaynak bölümüne bağlanmalıdır.
- Kaynak bulunamadığında genel model bilgisinden kesin bir parametre üretmek başarı sayılmaz, doğru davranış boşluğu açıkça söylemektir.
- Kabul kaydı için test kümesi değişiklikten önce dondurulmalı ve tanım, karşılaştırma, arıza ve kapsam dışı soruları içermelidir.

01 / Önce soruyu sınırlandırın

Bir bakım mühendisi “Sürücü neden duruyor?” diye sorduğunda ürün serisi, hata kodu ve çalışma koşulu eksiktir. Yanıtın ilk işi bu bilgileri istemektir. Belgede geçen benzer bir hata kodunu başka cihaz ailesine taşımak, metin olarak ikna edici fakat teknik olarak yanlış bir sonuç üretir. NQ ile çalışırken cihaz, sürüm, belirti ve daha önce yapılan kontrolü soruya ekleyin.

02 / Arama ile yanıtı ayrı değerlendirin

Bir test tablosunda soru, beklenen belge, ilgili bölüm ve kabul edilebilir yanıtı tutun. Önce ilk beş sonuç içinde doğru belgenin bulunup bulunmadığını ölçün. Sonra üretilen yanıtta her teknik iddiayı ilgili bölüme bağlayın. Arama başarısı, yanıt doğruluğunun yerine geçmez. Aynı cihazın eski kılavuzu, yüksek benzerliğe rağmen geçersiz bir referans olabilir.

03 / Cevapsız soruları teste katın

Koleksiyonda bulunmayan bir ürün, çelişen iki sürüm ve eksik hata kodu içeren sorular ekleyin. Beklenen davranış, boşluğu açıkça söylemek ve gerekli bilgiyi istemektir. Kaynak bulunamadığında genel model bilgisinden kesin parametre üretmek başarı sayılmaz. Arama puanını doğruluk yüzdesi olarak sunmayın; bu değer sadece sonuçların göreceli sırasını anlatır.

04 / Pilot kabul kaydı

Örnek pilot için her soru türünden en az birkaç kayıt seçin; tanım, karşılaştırma, arıza ve kapsam dışı soru ayrı gruplar olsun. Test kümesini değişiklikten önce sabitleyin. Sürüm, veri koleksiyonu, yanıt süresi ve kaynak bağlantılarını kaydedin. İki mühendisin farklı değerlendirdiği yanıtları uzlaşma toplantısına ayırın. Bu önerilen iş akışı, sahada doğrulanmış bir doğruluk iddiası değildir.

Uygulayın

NQ içinde aynı soruyu önce cihaz modeli olmadan, sonra model ve sürüm bilgisiyle sorun. Yanıtın hangi bölümünün gerçekten kaynağa dayandığını karşılaştırın.

Kaynaklar ve çevrimiçi rehber

Güncel çevrimiçi rehberi açın

NIST AI Risk Management Framework

Önerilen pilot adımları saha doğruluk iddiası içermez. Sentetik örnekler gerçek tesis ölçümü değildir.