



Dil Modelleri Vites Yükseltiyor

Son yılların en popüler konularından olan dil modelleri gelişmeye devam ediyor. Google, Gemini Pro 1.5'in girdi işleme kapasitesini iki kat artırarak 2 milyon tokene çıkardığını duyurdu. Yani kabaca 1,5 milyon sözcükten oluşan bir metni tek seferde işleyebiliyor. OpenAI ise GPT-4 Turbo'dan 2 kat daha hızlı ve %50 daha ucuz bilgi üreten ve çoklu girdiyi destekleyecek şekilde geliştirilmiş olan GPT-4o'yu piyasaya sürdü. Bu gelişmeleri 18 aylık baş döndürücü bir trendin son halkası olarak değerlendirmek mümkün. ChatGPT'nin Kasım 2022'de piyasaya sürülmesinden bu yana, GPT-4, Gemini 1.5 Pro, Claude 3 Opus ve Llama 3-70B'nin piyasaya sürülmesini içeren önemli kilometre taşlarıyla birlikte, birçok model sağlayıcı yeteneklerini iki önemli şekilde geliştirdi: LLM (geniş dil modeli)'lerin karmaşık kavramlar üzerinde düşünmesine ve karmaşık talimatları izlemesine olanak tanıyan akıl yürütme ve daha uzun veri girişine imkân sağlayan kapasite artırımları.

GPT-4 ve diğer gelişmiş modellerin akıl yürütme kabiliyetleri karmaşık cümleleri anlamalarını sağlıyor. Birçoğumuz bu modelleri bir iki cümlelik sorular sormak için kullansak da uzun talimatları yerine getirme kapasiteleri son derece iyi hale geldi. Artık verdikleri cevaptaki bir bölümü seçip bu bölüm üzerine daha ayrıntılı konuşabilme gibi özellikler sunuluyor. Ayrıca, veri giriş kapasitelerinin artmış olması bu araçların kullanımını çok farklı bir noktaya taşıdı. Tek seferlik girdi kapasitesi olarak bakıldığında

GPT-4o 128.000, Claude 3 Opus 200.000 ve Gemini 1.5 Pro 2 milyon tokeni destekliyor. Böylece birkaç kitaba ait PDF dosyalarını sisteme yükleyip bunlarla ilgili karmaşık sorular sorabiliyorsunuz.

Diğer taraftan OpenAI'nin en son modeli GPT-4o, ses, video, imaj gibi farklı veri tipleriyle çalışmada çıtayı yükseltmiş görünüyor. GPT-4o'nun sonundaki "o" omni yani çok kanallı sözcüğünün kısaltması. Daha önce kullanıcı tarafından sesli bir girdi verildiğinde bunu metne dönüştüren bir ara sistem kullandıktan sonra işleme başlayan GPT-4 artık doğrudan sesi anlayabiliyor. Model geliştirilirken farklı ortam türlerinden verilerle eğitilmiş. Böyle olunca farklı tonlamaları, sesteki sevinme veya üzülmeye gibi farklılıkları algılayabiliyor. Cevap verirken de benzer şekilde vurgular yapabiliyor. Aynı şekilde videoyu da hızlı şekilde işleyerek yorumlayabiliyor. Bu özellikler aşama aşama kullanıcıların kullanımına açılrsa da örnekler hayli etkileyici. Mesela bir örnekte eğlenceli ve olağanüstü gerçekçi tonda bir kadın sesiyle konuşan GPT-4o bir hikâye anlatırken, robotikten aşırı dramatik tonlara kadar farklı tonlar benimsiyor, İngilizce ve İtalyanca arasında akıcı bir şekilde çeviri yapıyor ve bir akıllı telefon kamerasıyla yakalanan yüz ifadelerini yorumluyor. Üstelik model gelişmiş olmasına rağmen kullanım maliyeti GPT-4'ün yarısına düşürülmüş, cevaplama hızı ise iki kat artırılmış. Sesli komutlara yanıt süresi 0,3 saniyeye düşürülmüş. GPT-4 için bu süre 5,4 saniye civarıydı.

Kaynak: openai.com