LLM Yetenekleri ve Sınırlamaları

LLM'lerin gerçekten güçlü olduğu alanlar, hallucination'ın neden olduğu, knowledge cutoff'un etkisi, akıl yürütme sınırları, ve eğitim verisindeki bias'ın nasıl yeniden üretildiği.

Orta 20 dk

Son güncelleme: 6 Ekim 2026
Çevrimiçi oku: https://www.learnforgex.com/tr/topics/llm-capabilities-and-limitations

LLM Yetenekleri ve Sınırlamaları

Bu, "Büyük Dil Modelleri" kategorisindeki dördüncü ve son ders, ve son üç dersin sessizce üzerine inşa ettiği ders bu. "Large Language Model'ler Nasıl Çalışır?" mekanizmayı açıkladı, "Token'lar ve Context Window'lar" onun sert sınırlarını kapsadı, ve "Prompting and Prompt Engineering" onu bilinçli olarak nasıl kullanacağınızı kapsadı -- bu ders, bir şey için bir LLM'e güvenip güvenmeyeceğinize ve nasıl güveneceğinize karar verirken gerçekten önemli olan soruyu soruyor: gerçekten neyi iyi yapabilir, ve nerede güvenilir biçimde yetersiz kalır? Bu konuda dürüst olmak, sonraki kategoriler için gerekli bir zemindir -- "Tool'lar ve MCP" ve "AI Agent'lar," ikisi de büyük ölçüde burada kapsanan sınırlamaların etrafından dolaşmak için var.

LLM'ler Gerçekten Neyde İyidir?

Sınırlamaları kapsamadan önce, gerçek güçlü yönler hakkında somut olmaya değer, çünkü bir LLM'in ne zaman kullanılacağına dair iyi kararlar vermek için ikisi de önemlidir. Modern LLM'ler şunlarda güçlüdür: metni dönüştürmek (özetlemek, çevirmek, farklı bir tonda yeniden yazmak), kod taslağı hazırlamak ve açıklamak, eğitim verisinde bulunan genel bilgiye dayanan soruları yanıtlamak, iyi belirtilmiş talimatları takip etmek ("Etkili Prompt'lar Yazmak"ı hatırlayın), ve birkaç örnekten bir örüntü almak ("Zero-Shot, Few-Shot ve Prompt'larda Örnekler"i hatırlayın). Bu güçlü yönlerin hepsi "Large Language Model'ler Nasıl Çalışır?"ta anlatılan aynı köke dayanır: devasa miktarda metin üzerinde pretraining, dili işlemede ve üretmede gerçekten geniş, esnek bir yetenek üretir.

Hallucination: Kendinden Emin, Akıcı, Yanlış

"Generative AI," **hallucination**'ı "bir generative AI sisteminin akıcı, kendinden emin ama olgusal olarak yanlış çıktı ürettiği durum" olarak tanıtmıştı -- bu ders, LLM'lerin nasıl çalıştığı

hakkında artık bildiğiniz her şeyi göz önünde bulundurarak *neden* olduğunu açıklıyor. "Large Language Model'ler Nasıl Çalışır?"tan, pretraining'in tüm hedefinin *doğrulanmış* değil *makul* bir sıradaki token'ı tahmin etmek olduğunu hatırlayın -- üretilen metni ortaya çıkarmadan önce gerçek bilgiyle karşılaştıran yerleşik bir mekanizma yoktur. Bir model bir olguya sahip olmadığında (ya pretraining'de yoktu, ya da mevcut context'in dışındadır), varsayılan olarak "bilmiyorum" demenin güvenilir bir yolu yoktur -- basitçe doğru olmayan, belirli, kendinden emin, iyi biçimlendirilmiş bir olgu gibi okunabilecek, istatistiksel olarak en makul devamı üretir. "Generative AI"ın ilk belirttiği gibi, akıcılık ve doğruluk tamamen ayrı özelliklerdir, ve hallucination, ikincisi olmadan birincisi mevcut olduğunda olan şeydir.

> Hallucination nadir ya da egzotik değildir -- bu modellerin metni nasıl ürettiğinin doğrudan, yapısal bir sonucudur, her LLM'de bir dereceye kadar mevcuttur. Akıcı, kendinden emin görünen bir cevabı asla otomatik olarak doğru sayma, özellikle belirli olgular, tarihler, alıntılar ya da sayılar için -- gerçekten önemli olan her şeyi doğrulayın.

Knowledge Cutoff ve Eksik Güncel Bilgi

"Large Language Model'ler Nasıl Çalışır?"tan, bir modelin **knowledge cutoff**'unun, pretraining verisinin durduğu nokta olduğunu hatırlayın -- o tarihten sonra gerçekten yeni olan hiçbir şey hakkında, nasıl sorulursa sorulsun, hiçbir bilgisi yoktur. Bu, hallucination'dan (bir modelin kendinden emin biçimde bilgi uydurması) farklı bir sınırlamadır, ama ikisi genellikle birlikte ortaya çıkar: cutoff'undan sonraki bir şey hakkında sorulan bir model ya bilmediğini söyleyebilir, ya da -- daha kötüsü -- boşluğu tanımak yerine makul görünen ama uydurma bir cevap hallucinate edebilir. Bu belirli sınırlama, tam olarak bu kursun ilerleyen "Tool'lar ve MCP" kategorisini motive eden şeydir: bir modele, yalnızca pretraining sırasında gömülü olana güvenmek yerine, araçlar aracılığıyla güncel, harici bilgiye erişim sağlamak.

Akıl Yürütme Sınırları

LLM'ler adım adım ve mantıklı görünen akıl yürütme üretebilir ("Prompting and Prompt Engineering"teki chain-of-thought prompting'i hatırlayın), ve bu genellikle çok adımlı problemlerde doğruluğu gerçekten iyileştirir. Ama gerçekte neyin olduğu konusunda kesin olmaya değer: model, bir hesap makinesinin ya da bir derleyicinin yaptığı gibi garantili, biçimsel olarak doğrulanmış bir mantıksal süreç yürütmüyor -- pretraining sırasında öğrenilen örüntülere dayanarak, birer birer, akıl yürütmeye benzeyen metin üretiyor. Kesin aritmetik, tam çok adımlı mantık, ya da aynı anda birçok kısıtlamayı dikkatli biçimde takip etmeyi gerektiren görevlerde, LLM'ler gerçek hatalar yapabilir ve yapar, bazen tamamen kendinden emin ve tutarlı görünen bir açıklama üretirken. Bu, chain-of-thought tekniklerinin işe yaramaz olduğu anlamına gelmez -- ölçülebilir biçimde yardımcı olurlar -- yanlış olmanın gerçekten önemli olduğu her şeyde, çıktılarının varsayılan olarak güvenilmesi değil kontrol edilmesi gerektiği anlamına gelir.

Eğitim Verisindeki Bias

"Large Language Model'ler Nasıl Çalışır?"taki "Pretraining: Dünyanın Metninden Öğrenmek" ten, bir modelin örüntülerini, var olan devasa bir insan yazımı metin örnekleminden öğrendiğini hatırlayın. O metin, kaynaklarında bulunan gerçek önyargıları, klişeleri ve dengesizlikleri yansıtır -- onunla eğitilmiş bir model, kimse öyle amaçlamasa bile, aynı örüntüleri kendi çıktısında yeniden üretebilir. Bu, hallucination'dan ya da akıl yürütme sınırlarından ayrı, ilgisiz bir kusur değildir -- aynı altta yatan gerçeğin (bir model, eğitim verisindeki istatistiksel örüntüleri yansıtır) farklı bir biçimde ortaya çıkmasıdır. Bunun farkında olmak, bir modelin çıktısının gerçek insanları haksız biçimde etkileyebileceği herhangi bir uygulama için (örneğin, insanları elemek, sıralamak ya da değerlendirmek) özellikle önemlidir.

Bu Sınırlamalar Neden Var?

Yukarıdaki dört sınırlamayı, ilgisiz bir hata listesi olarak ele almak yerine, hepsini ortak tek bir nedene bağlayarak geri adım atmaya değer. Bu dersteki her şey, "Large Language Model'ler Nasıl Çalışır?"taki aynı mekanizmaya dayanır: bir LLM, pretraining sırasında öğrenilen istatistiksel örüntülere dayanarak metnin makul devamlarını tahmin eder -- yerleşik bir olgu doğrulayıcısı, biçimsel bir mantık motoru, güncel olaylara canlı bir bağlantı, ya da açık bir bias düzeltme süreci yoktur. "LLM'ler Gerçekten Neyde İyidir?"te kapsanan her güçlü yön ve bu derste kapsanan her sınırlama, tam olarak aynı kaynaktan gelir. Bu yeniden çerçeveleme pratik olarak önemlidir: bu kursun ilerleyen kategorilerinin (Tool'lar ve MCP, AI Agent'lar) altta yatan modeli "düzeltmeye" çalışmamasının nedeni budur -- bunun yerine, güncel bilgi sağlayan, çıktıları doğrulayan ve modelin ne yapmasına izin verildiğini kısıtlayan sistemleri modelin *etrafına* inşa ederler, bu sınırlamaların kendiliğinden ortadan kalkacağını varsaymak yerine onlarla birlikte çalışırlar.

Best Practices

- Akıcı çıktıya varsayılan olarak güvenmek yerine belirli olguları, sayıları, tarihleri ve alıntıları doğrulayın ("Hallucination: Kendinden Emin, Akıcı, Yanlış"a bakın) -- özellikle yanlış olmanın gerçek bir maliyeti olan her bağlamda.
- Gerçekten zamana duyarlı bir şey için, doğrulanmamış bir cevaba güvenmeden önce bilginin modelin knowledge cutoff'undan sonrasına düşebileceğini kontrol edin ("Knowledge Cutoff ve Eksik Güncel Bilgi"e bakın).
- Adım adım akıl yürütme çıktısını bir garanti değil, yararlı bir yardımcı olarak ele alın ("Akıl Yürütme Sınırları"na bakın) -- bir hatanın önemli olacağı her şeyde gerçek mantığı kontrol edin.
- Gerçek insanlar hakkında kararlar içeren herhangi bir bağlamda bir modeli devreye almadan önce çıktısından kimin etkilenebileceğini düşünün ("Eğitim Verisindeki Bias"a bakın).

Yaygın Hatalar

- **Kendinden emin görünen bir cevabı doğruluğun kanıtı saymak.** "Hallucination: Kendinden Emin, Akıcı, Yanlış"ın açıkladığı gibi, akıcılık ve doğruluk bir LLM'in çıktısının birbiriyle ilgisiz özellikleridir.
- **Bir modelin bilgisi eksik olduğunda "bilmiyorum" diyeceğini varsaymak.** "Hallucination: Kendinden Emin, Akıcı, Yanlış" ve "Knowledge Cutoff ve Eksik Güncel Bilgi" de kapsandığı gibi, bir modelin kendi boşluğunu güvenilir biçimde tanıyıp işaretlemesinden çok, makul görünen bir tahmin üretmesi çok daha olasıdır.
- **Bu dersteki her sınırlamayı, ayrı ayrı etrafından dolaşılacak ilgisiz, bağımsız bir hata olarak ele almak.** "Bu Sınırlamalar Neden Var?"ın gösterdiği gibi, hallucination, knowledge cutoff, akıl yürütme hataları ve bias hepsi aynı altta yatan mekanizmaya dayanır -- bu mekanizmayı anlamak, ayrı başarısızlık modlarının bir listesini ezberlemekten daha yararlıdır.

Özet, Cheat Sheet ve Terimler Sözlüğü

Özet

- LLM'ler metni dönüştürmede, kod taslağı hazırlayıp açıklamada, genel bilgi sorularını yanıtlamada, ve iyi belirtilmiş talimatları takip etmede gerçekten güçlüdür -- bunlar doğrudan büyük ölçekli pretraining'den gelen yeteneklerdir.
- **Hallucination,** pretraining'in doğrulanmış değil makul devamlar için optimize edilmesinden kaynaklanır -- akıcılık doğruluğu ima etmez.
- Bir modelin **knowledge cutoff**'u, eğitim verisi toplandıktan sonra gerçekten yeni olan hiçbir şey hakkında bilgisi olmadığı anlamına gelir.
- LLM "akıl yürütmesi," biçimsel olarak doğrulanmış bir süreç değil, adım adım mantığa benzeyen üretilmiş metindir -- gerçek hatalar yapabilir ve yapar.
- Modeller gerçek dünya metninden öğrendiği için, o eğitim verisinde bulunan **bias'ları** yeniden üretebilirler.
- Bu sınırlamaların tümü ortak tek bir nedene dayanır: bir LLM, yerleşik bir olgu doğrulayıcısı, mantık motoru, canlı veri bağlantısı ya da bias düzeltmesi olmadan, öğrenilmiş örüntülere dayanarak makul metin tahmin eder.

Cheat Sheet

- Güçlü yön = dil işleme, taslak hazırlama, açık talimatları takip etme.
- Hallucination = akıcı + yanlış. Her zaman belirli olguları doğrulayın.
- Knowledge cutoff = pretraining verisi toplandıktan sonrası hakkında bilgi yok.
- Akıl yürütme = adım adım görünür, biçimsel olarak garanti edilmez -- kontrol edin.
- Bias = eğitim metnindeki örüntüleri (istenmeyenler dahil) yansıtır.
- Dördünün de kök nedeni: makul-metin tahmini, yerleşik doğrulayıcı yok.

Terimler Sözlüğü

- **Hallucination:** bir LLM'in akıcı, kendinden emin ama olgusal olarak yanlış çıktı ürettiği durum.
- **Knowledge cutoff (bilgi kesim tarihi):** pretraining verisi o tarihten önce toplandığı için, bir modelin ondan sonrası hakkında hiçbir bilgisi olmadığı zaman noktası.
- **Akıl yürütme (LLM'lerde):** diğer her çıktının üretildiği aynı şekilde üretilen, adım adım mantıksal akıl yürütmeye benzeyen metin -- biçimsel olarak doğrulanmış bir mantıksal süreç değil.
- **Bias (LLM'lerde):** eğitim verisinde bulunduğu için bir modelin çıktısında yeniden üretilen, klişeler ya da dengesizlikler dahil sistematik örüntüler.