

LLM'ler Seviye Atlıyor - Daha İyi, Daha Hızlı, Daha Ucuz: Haziran 2024

"Ekonomik Araştırma için Üretken Yapay Zeka" Bölüm 3 Güncellemesi: *Journal of Economic Literature* 61(4)'te yayınlanan "Use Cases and Implications for Economists" başlıklı makalenin 3. Bölümünün güncellenmesi†

ANTON KORINEK

Ekonomik Araştırmalarda Üretken Yapay Zeka Uygulamaları

Üretken yapay zeka (YZ) ortamı, bu makalenin orijinal versiyonunun *Journal of Economic Literature*'da yayınlanmasından bu yana hızla gelişmiştir (Korinek 2023b). Bu bölüm, ana yenilikleri özetlemekte ve orijinal makale ile aynı altı alanda sınıflandırılan ekonomik araştırmalardaki en son büyük dil modellerinin (LLM'ler) güncellenmiş kullanım örneklerini göstermektedir: fikir ve geri bildirim, yazma, arka plan araştırması, kodlama, veri analizi ve matematiksel türetmeler. Her bir alan için genel bir açıklama ve LLM yeteneklerinden nasıl yararlanılacağına dair birkaç özel kullanım örneği sunuyorum. LLM'lerin kullanışlılığının dengeli bir versiyonunu sunmak için yazım sırasında LLM'lerin hem yeteneklerini hem de başarısızlıklarını gösteriyorum.

Ucuz: Haziran 2024 'Ekonomik Araştırma için Üretken Yapay Zeka' Bölüm 3 Güncellemesi: *Journal of Economic Literature* 61(4)'te yayınlanan 'Use Cases and Implications for Economists' başlıklı makalenin 3. Bölümünün Haziran 2024 Güncellemesi." Gelecekteki güncellemeler için <https://www.aeaweb.org/articles?id=10.1257/jel.20231736#additionalMaterials> adresini kontrol edin.

† Lütfen şu şekilde alıntılayın Korinek, Anton, 2024. "LLM'ler Seviye Atlıyor - Daha İyi, Daha Hızlı, Daha

Yeni Geliřmelerin Özeti - LLM'lerdeki son geliřmeler, daha iyi performans, büyüyen bağlam pencereleri, böylece LLM'lerin aynı anda daha fazla veriyi işleyebilmesi, daha iyi hatırlama, daha hızlı işleme ve düşen maliyetler ile karakterize edildi. Bu makalenin ilk versiyonunun Aralık 2023'te yayınlanmasından bu yana, önde gelen üç yapay zeka laboratuvarının her biri, diğeri özelliklerin yanı sıra görme yetenekleri ve gerçek zamanlı ses işleme özelliklerini içeren LLM tekliflerinde önemli güncellemeler yayınladı. Mayıs 2024'te yazarken, OpenAI'nin GPT-4o'su halka açık en yetenekli LLM'dir ve benim

en çok kullandığım LLM'dir. OpenAI ayrıca ChatGPT uygulaması içinde GPT-4'ün kullanışlı özelleştirilmiş sürümlerini sunan bir GPT mağazası da sunuyor. Ayrıca, yazma yetenekleri için Anthropic'in Mart 2024'te piyasaya sürülen Claude 3'ünü düzenli olarak **k u l l a n ı y o r u m** . Google DeepMind, Ocak 2024'te şu anda en geniş bağlam penceresini sunan Gemini Pro 1.5'i piyasaya sürdü - halka açık sürüm aynı anda 1 milyon jetonu (yaklaşık 1500 sayfa metin) işleyebiliyor. Ocak 2024'te piyasaya sürülen kardeş modeli Gemini Ultra 1.0, Google DeepMind'ın en yetenekli modelidir ancak kamuya açık değildir.

TABLO 1 YAYGIN Olarak KULLANILAN LLM'lere GENEL BAKIŞ, 15 MAYIS 2024 İTİBARIYLA Özellikler

| Yapay Zeka Laboratuvarı | Model | En İyi Versiyon | Serbest bırakıldı | Jetonlar | Veri kesme | URL |
|-------------------------|----------|-----------------|-------------------|----------|--------------|---------------------|
| OpenAI | GPT-4o | GPT-4o | Mayıs 2024 | 128k | Aralık 2023 | chatgpt.com* |
| Antropik | Claude 3 | Opus | Mart 2024 | 200k | Ağustos 2023 | claude.ai |
| Google | İkizler | Ultra | Şubat 2024 | 32k | Temmuz 2023 | gemini.google.com* |
| Mistral | Mixtral | 8x22b | Nisan 2024 | 65k | Eylül 2021 | OS, chat.mistral.ai |
| Meta | LlaMA 3 | Llama-3 | Nisan 2024 | 8k | Aralık 2023 | OS, meta.ai |

Not: *, internette de gezinebilen sohbet robotlarını; OS ise açık kaynaklı modelleri göstermektedir.

Her üç laboratuvarın sınır modellerinin performansı benzerdir, ancak OpenAI hafif bir liderliğe sahiptir.

Açıklanan öncü modeller tescilli ve yalnızca internet üzerinden kullanılabilirken, giderek artan sayıda daha küçük (daha az yetenekli olsa da) modeller açık kaynak olarak yayınlanmış ve kullanıcıların bunları güvenlik veya gizlilik açısından hassas uygulamalar için kendi sunucularında kullanmalarına izin vermiştir. Bu modellerden bazıları dizüstü bilgisayarlarda ve hatta cep telefonlarında çalıştırılacak kadar küçüktür.

Bölüm 3.1'de bu yeni gelişmeleri açıklamanın yanı sıra, aşağıdaki alt bölümlerde LLM'ler için birkaç yeni örnek ve kullanım durumunu da ele alıyorum:

- Otomatik istem oluşturma
- Sesli sohbetler aracılığıyla araştırma hakkında geri bildirim
- Elle yazılmış denklemleri LaTeX'e dönüştürme
- Sunum slaytları oluşturma
- YouTube videolarını özetleme
- Kodlama projesi için bir planın ana hatlarının çizilmesi
- Veri kaynaklarını bulma

Ekonomi alanında, AEA dergilerinin çoğu yakında yazarlardan araştırmalarında LLM'leri kullanıp kullanmadıklarını ve nasıl kullandıklarını beyan etmelerini isteyecektir. Yazarların sundukları içerikten yalnızca kendilerinin sorumlu

olduğunu söylemeye gerek yok.

Veri Gizliliği.-Araştırmacılar için önemli bir konu da veri gizliliğinin nasıl sağlanacağıdır.

LLM'lere girdikleri verilerin özgünlüğü. OpenAI, ChatGPT uygulamasında bir "geçici sohbet" seçeneğinin yanı sıra kullanıcı ayarlarında bir gizlilik seçeneği ("Modeli herkes için geliştir" seçeneğini kapatın) sunarak kullanıcıların girdilerinin gelecekteki LLM'leri eğitmek için kullanılmasını devre dışı bırakmalarını sağlar. OpenAI, uygulama programlama arayüzü (API) aracılığıyla girilen kullanıcı verilerini eğitim amacıyla kullanmamaktadır. Anthropic, açık bir tercih dışında veya nadir durumlarda güvenlik incelemesi için işaretlenmediği sürece kullanıcı verilerini gelecekteki eğitimler için kullanmaz. Google, kullanıcılara Gemini uygulamalarına gizli bilgiler girmemelerini tavsiye eder, çünkü girilen veriler gelecekte eğitim amacıyla kullanılabilir.

3.1 Son Teknoloji LLM'lere Genel Bakış

Tablo 1, Mayıs 2024 itibarıyla yaygın olarak kullanılan LLM'lere genel bir bakış sunmakta ve bunların yayınlanma tarihi, bağlam penceresi ve eğitim verisi sınırlaması gibi bazı temel özelliklerini ve sınırlamalarını içermektedir. Bağlam penceresi, modelin işleyebileceği maksimum belirteç sayısını yansıtır ve bu sınır tipik olarak kullanıcı istemi ile oluşturulan tamamlamanın toplamına uygulanır. Ortalama olarak, dört belirteç İngilizce dilinde üç kelimelik metne karşılık gelir. Eğitim verisi sınırı, LLM'nin ön eğitimden geçirildiği verilerin ne kadar yeni olduğunu yansıtır. Modeller, internete erişim kapasitesine sahip olmadıkları sürece, genellikle bu tarihten sonra meydana gelen gerçekler hakkında bilgi sahibi değildir.

3.1.1 Erişim Modları

Listelenen LLM'lere dört farklı şekilde erişilebilir:

- **Web tabanlı Sohbet Robotları:** Tablo 1'deki modellerin tümüne son sütunda listelenen URL'ler altında sohbet robotları olarak erişilebilir. Bu sohbet robotu arayüzleri, kullanıcıların modelleri asistan veya eğitmen olarak yönlendirmesine olanak tanır. Aşağıda belgelenen kullanım örneklerinin çoğu bu etkileşim modunu göstermektedir. Çoğu sohbet robotunun yanıtlarında belirli bir miktarda random varyasyonu ortaya çıkarak çeşitli programlandığını, bu nedenle sohbet arayüzlerindeki sonuçların tam olarak tekrarlanabilir olmadığını unutmayın. Bu sohbet robotlarının ücretsiz sürümleri ya kullanım kısıtlamalarıyla birlikte gelir ya da yalnızca daha az güçlü model sürümlerine erişim sağlar. Benim deneyimlerime göre bu durum, hayal kırıklığını önlemek ve günlük işlerde LLM kullanımından anlamlı üretkenlik kazanımları elde etmek için daha az kısıtlamaya sahip ilk üç modelden birine veya birkaçına (ChatGPT 4o, Claude 3 Opus veya Gemini Advanced) erişmek için gereken aylık 20 dolarlık abonelik ücretini ödemeye değer kılıyor.
- **Gerçek Zamanlı Sesli Asistanlar:** OpenAI en son modeli GPT-4o ile konuşma dilinde doğal etkileşime olanak tanıyan etkileşimli bir sesli asistan sunuyor ve kamera ve/veya ekran görüntüleri aracılığıyla, ChatGPT telefon uygulaması aracılığıyla ve iOS. Bazı kullanıcılar, ChatGPT uygulamasını belirli iş görevleri boyunca ve hatta gün boyunca telefonlarında ses modunda açık tuttıklarını, böylece herhangi bir noktada parmaklarını kaldırmadan dijital asistanlarından

kolayca yararlanabildiklerini bildiriyor. Google, benzer işlevselliğe sahip Project Astra (gelişmiş gören ve konuşan duyarlı ajan) asistanının bir demosunu yayınladı ve bu demo 2024'ün ilerleyen aylarında halka açık olarak yayınlanacak.

• **Web Tabanlı Deneme Platformları:**

Tüm büyük LLM sağlayıcıları, sohbet robotlarından **d a h a** fazla işlevsellik ve esneklik sunan ancak pro- gramlama bilgisi gerektirmeyen web tabanlı arayüzler de sunmaktadır. OpenAI Playground (<https://playground.openai.com>),

Anthropic Console (<https://console.anthropic.com>)

gibi bu platformlar,

ve Google AI Studio (<https://aistudio.google.com>), kullanıcıların farklı model ayarlarını denemelerine olanak tanır.

perature ve top-p sampling, and provide more control over the input and output formats compared to chatbots. Bu tür deney platformları, LLM'lerin kapasitelerini keşfetmek, istemleri test etmek ve belirli görevler için modellere ince ayar yapmak için özellikle yararlıdır.

- **API'ler:** Maksimum düzeyde özelleştirme ve entegrasyon için, listelenen modellere programcıların LLM'leri doğrudan kendi yazılım uygulamalarına entegre etmelerine olanak tanıyan API'ler aracılığıyla da erişilebilir. Bu, tekrarlayan görevleri otomatikleştirmek veya doğal dil işleme tekniklerini kullanarak büyük veri kümelerini analiz etmek gibi çok çeşitli daha gelişmiş ve özelleştirilmiş kullanım durumlarını mümkün kılar. API'ler yukarıdaki iki seçeneğe kıyasla daha fazla esneklik ve kontrol sağlar ve kullanım başına ödeme esasına göre kullanılabilir, ancak aynı zamanda etkili bir şekilde kullanmak için daha yüksek düzeyde teknik uzmanlık gerektirir. API'ler aracılığıyla LLM'lere erişmek genellikle model sağlayıcıdan bir API anahtarı için kaydolmayı, tercih edilen programlama dilinde bir istemci kütüphanesi kurmayı ve API uç noktalarıyla etkileşim için kod yazmayı içerir. Bu süreç bir chatbot kullanmaktan daha karmaşık olsa da, gerekli programlama becerilerine sahip olanlar için LLM'lerin tüm

potansiyelini ortaya çıkarır. Bu makalenin çoğaltma paketi, LLM'leri otomatik olarak sorgulamak için API'lerin nasıl kullanılacağını göstermektedir.

Tablo 1'de listelenen modeller, her biri performans, hız ve maliyet arasında kendi ödünleşimlerine sahip farklı boyutlardaki ailelerde mevcuttur. Daha büyük modeller daha "akıllıdır" ve genellikle daha iyi performans ve daha fazla yetenek sunar, ancak aynı zamanda daha fazla hesaplama kaynağı gerektirir ve isteklerin işlenmesi daha uzun sürer, bu da onları daha pahalı hale getirir. Öte yandan daha küçük modeller daha hızlı ve daha uygun maliyetlidir, ancak çıktılarında aynı düzeyde kalite sağlayamayabilirler. Bu, kullanıcıların uygulamaları için uygun model boyutunu seçerken kendi özel ihtiyaçlarını ve bütçelerini göz önünde bulundurmalarını sağlar.

3.1.2 Önde Gelen Tescilli Modeller

Tablodaki ilk üç model profesyoneldir, yani bunları geliştiren laboratuvarlara aittir ve kaynak kodları, mimarileri ve model ağırlıkları kamu kullanımı, modifikasyon veya dağıtım için serbestçe kullanılamaz. Bunlara yalnızca laboratuvarların sunucularında, öngördükleri koşullara ve kontrollere tabi olarak erişilebilir.

Önde gelen modellerin tümü, dil işlemeyi görsel bilgi işleme becerisiyle birleştirir, bu da onların görme-dil modelleri (VLM'ler) olduğu anlamına gelir. Bu sayede kullanıcılar sohbet arayüzüne ya da API'lere görüntü yükleyebilmekte ve ilerleyen bölümlerde bazı kullanım örneklerinde inceleyeceğimiz gibi bu görüntülere dayalı yanıtlar isteyebilmektedir. Bununla birlikte, bunlara ortak terim olan LLM'leri kullanarak atıfta bulunmaya devam edeceğiz.

OpenAI'nin GPT-4 modeli ilk olarak Mart 2023'te kendi sınıfında lider model olarak piyasaya sürülmüştür ve açık ara en popüler LLM'dir. En son sürüm Mayıs 2024'te piyasaya sürülen GPT-4o'dur; burada "o" son eki "omni" anlamına gelir ve modelin gerçek zamanlı bir dijital asistan olarak hareket etmek için metin, görüntü ve sesi aynı anda işleyebileceğini yansıtır. GPT-4o ayrıca

modelin yeteneklerini geliştiren çeşitli uzantılara da erişime sahiptir, bunlar arasında veri

analizi, web'de gezinme yeteneği ve aşağıda açıklanan GPT Mağazası. GPT-4o ücretsiz olarak kullanılabilir ancak ChatGPT'de kısmen talebe bağlı olan kullanım sınırlarına tabidir. Ödeme yapan kullanıcılar daha güvenilir erişim elde edebilirler. GPT-3.5 Turbo, GPT-4'ün daha küçük bir kardeşidir ve ChatGPT'nin ücretsiz sürümü aracılığıyla da erişilebilir. Alternatif olarak, OpenAI'nin tüm modellerine web tabanlı OpenAI Playground üzerinden ve API aracılığıyla erişilebilir.

ChatGPT 4o'nun ekonomistler için özellikle yararlı olan bir özelliği, doğrudan ChatGPT arayüzü içinde korumalı bir ortamda Python kodu yazabilen ve çalıştırabilen ve sonuçları görüntüleyebilen ve bunlar üzerinde inşa edip yineleyebilen bir araç olan Gelişmiş Veri Analizi'dir. Gelişmiş Veri Analizi, kullanıcıların dosya yüklemelerine ve dosya dönüşümlerinden regresyon gibi karmaşık analizlere kadar değişen veri işleme görevlerini gerçekleştirmelerine olanak tanır. Aşağıda, örneğin Solow modeli simülasyonumuzda bu kabiliyetlerden birkaçını ele alacağız.

OpenAI'nin GPT Mağazası, ChatGPT arayüzüne entegre edilmiş dijital bir pazar yeridir ve kullanıcılara üçüncü taraf içerik oluşturucular tarafından geliştirilen çok çeşitli uzantı ve uygulamalara göz atma, yükleme ve kullanma olanağı sunar. Bu "özel GPT'ler", yazı yazmaktan ekonomik analiz araçlarına kadar özel bilgiler, etkileşimli araçlar ve bireysel kullanıcı ihtiyaçlarına göre uyarlanmış özelleştirilmiş deneyimler ekleyerek GPT-4o'nun işlevselliğini geliştirir. (Mart 2024'e kadar mevcut olan ChatGPT eklentilerinin yerini almışlardır). Özel bir GPT'yi kullanmak için ChatGPT arayüzünün sol tarafındaki menü çubuğunda "GPT'leri Keşfet" seçeneğine tıklayıp arama yaparak bulunabilir. Ekonomistlerin faydalı bulabileceği birkaç özel GPT aşağıdakileri içerir:

- **Veri Analisti**, OpenAI tarafından oluşturulmuş, kullanıcıların **veri** yüklemesine olanak tanıyan özel bir GPT'dir.

Elektronik tablolar ve içerdiği verileri analiz etme veya görselleştirme.

- **Wolfram**, Mathematica'nın yapımcısı Wolfram Alpha'dan hesaplama, matematik, seçilmiş bilgi ve gerçek zamanlı verilere erişim sağlar.
- **Consensus**, 200 milyon akade- mik makaleyi tarayarak, temel makalelere atıfta bulunan bilimsel temelli cevaplar sunan yapay zeka tabanlı bir araştırma asistanı sunmaktadır.

Özel bir GPT'yi ilk kez kullandıktan sonra, herhangi bir ChatGPT görüşmesinde @ işaretini ekleyerek, örneğin "@Wolfram" yazarak bu GPT'ye başvurabilirsiniz. Kullanıcılar ayrıca "GPT'leri Keşfet" ve "Oluştur" düğmesine tıklayarak ve takip eden süreci adım adım izleyerek kendi özel GPT'lerini oluşturabilirler. Ortaya çıkan özel GPT'ler özel olarak kullanılabilir ya da herkese açık olarak paylaşılabilir.

Microsoft Copilot (eski adıyla Bing Chat)

Microsoft'un GPT-3.5/4 OpenAI modelleri ailesini temel alan sohbet motorudur. Microsoft, Copilot'u tüm ofis ürünleri. Copilot ayrıca web'de gezinebiliyor ve kullanıcılara internette mevcut olan en son bilgilere dayanan sonuçlar sunuyor. Kullanıcıların Hassas, Dengeli veya Yaratıcı olmak üzere 3 mod arasından seçim yapmasına olanak tanır.

Anthropic'in Claude 3'ü Mart 2024'te piyasaya sürüldüğünde halka açık en iyi LLM idi ancak şimdi GPT-4o tarafından gölgede bırakıldı. Claude, LLM'yi bir dizi üst düzey etik ilkeyi takip edecek şekilde eğitmek için anayasal yapay zeka adı verilen bir süreç kullanarak kendisini yardımsever, onurlu ve zararsız olarak markalamaktadır (Bai ve ark. 2022). Claude'un üç versiyonu bulunmaktadır: Claude 3 *Opus* en güçlü modeldir ve yazma stilini beğendiğim için düzenli olarak yazı yazmak için kullanıyorum; Claude 3 *Sonnet* hız ve beceriler arasında bir uzlaşma sunar; *Claude 3 Haiku* en küçük modeldir

ve en hızlı modeldir, bu da mevcut en sofistike modeli gerektirmeyen toplu araştırma uygulamaları için kullanışlıdır. Claude 3, 200 bin jetonluk bir bağlam penceresine sahiptir, bu da tek seferde yaklaşık 150.000 kelimeyi işleyebilmesini sağlar; örneğin, kullanıcı Claude sohbet arayüzüne birkaç akademik makale yükleyebilir ve aynı anda işleyebilir. Modeller ayrıca web tabanlı Anthropic Console ve kullanıcılara dil modelinin harici araçlardan yararlanmasına izin verme olanağı sunan API aracılığıyla da kullanılabilir.

Google DeepMind'in Gemini serisi LLM'leri ilk olarak Aralık 2023'te piyasaya sürüldü ve yanıtlarına gerçek zamanlı bilgileri dahil etmek için internete de erişebilen ve kullanıcıların sonuçları çapraz kontrol etmesine ve kaynaklarına giden bağlantıları takip etmesine olanak tanıyan isimsiz bir sohbet robotu aracılığıyla erişilebilir. Gemini model ailesinin dört boyutu bulunmaktadır. *Gemini Ultra 1.0* en güçlü modeldir ancak halka açık değildir. *Gemini Pro 1.5* Şubat 2024'te piyasaya sürülmüştür ve halka açık sürümünde 1 m'lik bir belirteç bağlam penceresine sahip olma özelliğini taşımaktadır - bu, halka açık tüm LLM'ler arasında en uzun olanıdır. Bu, yeni kullanım alanları sunmaktadır; örneğin, araştırmacıların çalışmalarının önemli bir kısmını tek seferde yüklemelerine ve sorguları buna göre işlemelerine olanak sağlamaktadır. *Gemini Flash 1.5* Mayıs 2024'te piyasaya sürüldü ve daha düşük maliyetle daha yüksek hızlar sunuyor ancak performansı biraz daha düşük. Hem *Pro 1.5* hem de *Flash 1.5*, web tabanlı Google AI stüdyo platformunun yanı sıra API aracılığıyla da kullanılabilir. Gemini Nano 1.0 cep telefonlarında çalışabilecek kadar küçüktür.

3.1.3 Önde Gelen Açık Kaynak Modelleri

Tablo 1'de listelenen son iki LLM açık kaynak kodludur, yani kullanımı, değiştirilmesi ve dağıtımı serbesttir.¹ Bu

¹ Halka açık olarak yayınlanan birçok LLM'nin "açık kaynak" teriminin geleneksel katı tanımını tam olarak karşılamadığını unutmayın; bu da kaynak kodunun, eğitim verilerinin ve sonuçta ortaya çıkan parametre ağırlıklarının

ekonomik arařtırmalar için çeřitli faydalar sunmaktadır. İlk olarak, açık kaynaklı modellerin řeffaflığı, arařtırmacıların altta yatan mimariyi incelemelerine olanak tanıyarak modelin yapısını daha iyi anlamalarını ve üretilen çıktıları değerlendirmelerini sağlar. İkinci olarak, açık kaynaklı projeler işbirliğine dayalı inovasyona olanak tanır ve ekonomik arařtırmalar gibi özel ihtiyaçlara göre uyarlanmış LLM'lerin geliştirilmesini hızlandırabilir.

Arařtırmacılar uzmanlıklarıyla katkıda bulunabilir, iyileřtirmeler önerebilir ve modelleri kendi özel ihtiyaçlarına göre uyarlayabilirler. Üçüncü olarak, arařtırmacıların düşük maliyetli bilgi işlem kaynaklarına erişimi varsa, mali maliyetlere katlanmadan çalışmaları için açık kaynaklı modellerden yararlanabilirler. Son olarak, açık kaynaklı modeller daha fazla tekrarlanabilirliğe olanak tanır, bu da diğ er arařtırmacıların rapor edilen sonuçları doğrulamasına ve üzerine inşa etmesine olanak tanıdığı için ekonomik arařtırmalarda bilimsel bütünlüğün sağlanmasına yardımcı olur. Bu faydalar, açık kaynaklı dil modellerini, çalışmalarında doğal dil işleminin gücünden yararlanmak isteyen ekonomi arařtırmacıları için cazip bir seçenek haline getirmektedir.

Ekonomik bir perspektiften bakıldığında Açık kaynak modelleri, LLM'ler tarafından yaratılan ekonomik sosyal artığı serbestçe dağıttıkları ve inovasyonu teşvik ettikleri için oldukça faydalıdır. Öte yandan, açık kaynaklı LLM'ler daha yetenekli hale geldikçe, artan güvenlik riskleri de ortaya çıkmaktadır (Anderljung ve ark. 2023). Örneğ in, LLaMA, arařtırmacıların yukarıda listelenen tüm LLM'lerin güvenlik kısıtlamalarını aş an düşmanca saldırılar oluřturmasına zaten izin vermiştir (Zou ve ark. 2023). Seger ve diğ erleri (2023), LLM'lerin açık kaynak kullanımının artılarını ve eksilerini ve bu şekilde arzu edilebilecek ara çözümleri tartışmaktadır

modeller daha yetenekli hale gelmekte ve artan güvenlik riskleri oluřturmak t a d ı r .

Mistral'in Mixtral 8x22b modeli, Nisan 2024'te piyasaya sürülen önde gelen açık kaynaklı modellerden biridir. Mistral, eski Google DeepMind ve Meta çalışanları tarafından kurulan Paris merkezli bir girişimdir. Modelin adı, şirketin adının yanı sıra karmařık görevleri verimli bir şekilde çözmek için uzman olarak adlandırılan sekiz özel alt modeli birleřtiren modelin temelindeki uzmanlar karışımı (MoE) mimarisi üzerine bir kelime oyunudur.² Mistral ayrıca bir dizi tescilli model ve daha küçük açık kaynaklı modellerin yanı sıra bir sohbet arayüzü, web tabanlı bir platform ve modellerine API erişimi sunmaktadır.

Meta'nın LLaMA 3 serisi, ilk ikisi Nisan 2024'te piyasaya sürülen 8B, 70B ve 400B parametrelili bir model ailesini kapsamaktadır. En büyük modelin kısa süre içinde piyasaya sürülmesi bekleniyor. Meta, modeli eğitmek için kullanılan verileri saklı tutarken, temel kodu ve eğitilen modellerin ağırlıklarını serbestçe dağıtmıştır. 70B parametre sürümü GPT-3.5'ten daha iyi performans gösterir ve Microsoft Azure, AWS ve Hugging Face gibi önde gelen bulut bilişim platformlarında kullanılabilir.

Açık kaynaklı modellerin ortaya çıkması, arařtırmacıların LLM'leri kendi bilgisayarlarında çalıştırmalarına da olanak sağlamış ve ařağıdaki gibi avantajlar sunmuřtur

² Üretilen her belirteç için tüm parametrelerin kullanıldığı geleneksel yoğun LLM'lerin (örneğin, GPT-3) aksine, MoE modelleri her girdi için farklı alt parametre setlerini dinamik olarak seçerek daha verimli

Hızlı, Daha Ucuz

kısıtlama olmaksızın serbestçe paylaşılabilir. Örneğin, tabloda listelenen modeller eğitim kaynak kodlarını ve verilerini paylaşmamaktadır. Dahası, bunlardan bazıları, örneğin Meta'nın LLaMA 3'ü, modellerinin büyük ölçekli ticari kullanımına sınırlamalar getirmektedir. Yapay zeka laboratuvarı TogetherAI, LLaMA veri setinin açık kaynaklı bir sürümünü çoğaltmak ve dağıtmak için RedPajama adlı bir projeye sahiptir.

Bu yaklaşımın uygulanması daha zordur ancak MoE modellerinin yoğun muadillerine kıyasla hesaplama açısından daha verimli olmalarının yanı sıra en son teknolojiye sahip performansa ulaşmalarını sağlar. Önde gelen üç profesyonel laboratuvar (Anthropic, Google DeepMind, OpenAI) önde gelen modellerinin tam mimarisi konusunda sessiz kalsa da, uzmanlar hepsinin MoE mimarilerini kullandığına inanıyor.

veri gizliliği, maliyet etkinliği, kişiselleştirme ve çevrimdışı erişilebilirlik. Hesaplama kaynağı gereksinimleri, çoğu masaüstü bilgisayarda yalnızca küçük modellerin makul bir hızda yürütülebileceği anlamına gelir. Ancak, hesaplama kapasitesindeki ilerlemelerin yanı sıra daha küçük modellerin daha büyük yeteneklere sahip olmasını sağlayan LLM'lerin hızlı verimlilik kazanımları, LLM'lerin yerel kullanımını hızla daha cazip hale getirmektedir. Şu anda LLM'lerin yerel olarak kullanılmasını kolaylaştıran iki çözüm bulunmaktadır:

- LM Studio, kullanıcıların VLM'ler de dahil olmak üzere bir dizi açık kaynaklı LLM'yi kişisel bilgisayarlarına veya sunucularına indirmelerine ve çalıştırmalarına olanak tanır.
- llamafire, LLM'leri tek bir dosyada indirmeyi ve çok çeşitli bilgisayar sistemlerinde çalıştırmayı mümkün kılar.

Açık kaynaklı LLM Gemma 2b, Google DeepMind tarafından cep telefonlarında kullanılmak üzere geliştirilmiştir. Microsoft'un phi-3-mini 3.8b modeli de yerel makinelerde kullanım için optimize edilmiştir.

LLM Etkileşimi ve Denemeleri için Merkezi Merkezler - Ara sıra kullanıcılara önde gelen tüm LLM'lere erişim sağlayan kullanıcı dostu bir sohbet arayüzü sunan faydalı bir web sitesi <https://poe.com>'dur. Benzer şekilde, kullanıcılara çok çeşitli modellere erişim sağlayan web tabanlı bir deney platformu sunan bir web sitesi de <https://nat.dev/> adresidir.

Her ne kadar herkesin kendi deneyimleri ve yaparak öğrenme yoluyla LLM'lerle etkileşim deneyimi kazanması faydalı olsa da, önde gelen model sağlayıcıların çoğu artık LLM'lerin nasıl yönlendirileceğine ilişkin stratejileri açıklayan kamuya açık dokümanlar da sunmaktadır. Aşağıda bu tür kılavuzların bir listesi yer almaktadır:

- <https://docs.anthropic.com/claude/docs/prompting'e> giriş ve <https://docs.anthropic.com/tr/prompt-library/library> adresinde ilgili bir prompt kütüphanesi
- <https://platform.openai.com/docs/guides/prompt-engineering>
- https://ai.google.dev/docs/prompt_intro
- <https://llama.meta.com/get-started/#prompting>

Anthropic ayrıca, kullanıcıların Anthropic Console Dashboard aracılığıyla LLM'leri sorgulamak için kullandıkları istemleri optimize etmelerine yardımcı olan bir "istem oluşturucu" sunar ("Bir İstem Oluştur" seçeneğine tıklayın). Bu araç, kullanıcının niyetini açıklayan yazılı dildeki temel talimatları, LLM'leri daha yüksek kalitede çıktı üretmeye teşvik eden daha karmaşık ve spesifik bir dizi talimata dönüştürür. Doğal olarak kullanıcı, amacını mümkün olduğunca iyi yakaladığından emin olmak için kullanımdan önce oluşturulan talimatları düzenleyebilir ve ayarlayabilir. Bunu aşağıdaki örnekle açıklayacağım. Bu bölümün geri kalanında, LLM'ler tarafından üretilen tüm örnekleri kutular içinde sunacağım, istem başlıkta ve LLM'nin ürettiği yanıt kutunun gövdesinde yer alacaktır.

Sohbet 1 (Anthropic Prompt Generator): *Bir ekonomi makalesi yazdım ve Brookings için bir blog taslağı oluşturmak istiyorum.*

Göreviniz, Brookings Enstitüsü için bir ekonomi araştırma makalesine dayanan bir blog yazısı hazırlamaya yardımcı olmaktır. Makalenin tam metni burada:

```
<kağıt>
{{PAPER}}
```

```
</paper>
Bu blog yazısının hedef kitlesi:
```

```
<izleyici>
{{AUDIENCE}}
```

```
</audience>
```

İlk olarak, makalenin tamamını dikkatlice okuyun.

```
<scratchpad>
```

bölümünde kilit noktalar, ana bulgular ve çıkarımlar hakkında notlar alın.

Ardından, blog yazısı için kısa bir taslak yazın

```
<outline>
```

bölümü. Hedef kitleyi içeride tutun

Neye odaklanacağımıza ve bilgiyi nasıl sunacağımıza karar verirken aklınızda bulunsun.

Ardından, taslağınızı bir kılavuz olarak kullanarak blog yazısının tam taslağını bir <blog_post_draft> bölümünde yazın. Şunları yapmayı unutmayın:

-Hedef kitle tarafından kolayca anlaşılabilir olacak açık, ilgi çekici bir dil kullanın

-Gönderiyi daha ilişkilendirilebilir ve ilginç kılmak için uygun yerlerde ilgili örnekler, anekdotlar veya benzetmeler ekleyin

- Orijinal kağıttan kelimesi kelimesine kopyalamaktan kaçının. Anahtar kavramları kendi kelimelerinizle ifade edin.

-Teknik ayrıntılara çok fazla girmek yerine araştırmanın sonuçlarına ve "ne olmuş yani "sine odaklanmak

Taslağı bitirdikten sonra, makaledeki en önemli ve ilginç noktaları hedef Brookings kitlesinde yankı uyandıracak şekilde ele aldığınızdan emin olmak için gözden geçirin. Taslağınızı göndermeden önce son düzeltmeleri yapın.

Oluşturulan bilgi istemi, mümkün olan en iyi sonuçları elde etmek için birkaç son teknoloji bilgi istemi tekniğini kullanan dikkate değer bir ayrıntı düzeyi sağlar: modelden bir "karalama defteri" bölümünde not alarak başlamasını ister, bu da modeli metnin ana derslerini sentezlemeye teşvik eder. Ardından modelden bir taslak hazırlamasını ister, bu da modeli bir plan oluşturmaya teşvik eder. Ancak bu adımlardan sonra bilgi istemi modelden metnin taslağını yazmasını ister. Komut istemi ayrıca kullanıcıların söz konusu makaleyi eklemesi ve dinleyiciyi tanımlaması için iki konum bırakmaktadır. Ben bunları basitçe "Makalenin tam metni ektedir" yazarak ve hedef kitle olarak "politika yapımcılar"ı ekleyerek değiştirdim. Her ne kadar üretilen blog yazısı Brookings gibi bir düşünce kuruluşunun derinliği ve kalitesinde olmasa da, bu istem önde gelen LLM'lerin üçünde de böyle bir projede önemli ölçüde zaman kazandıracak faydalı sonuçlar üretti.³

³ Anthropic Prompt Generator'dan birkaç farklı mecra için blog yazılarına yönelik ipuçları oluşturmasını istedim. Nedense, Brookings'i hedefleyen bir blog istemek, başka herhangi bir blog istemektan daha kaliteli bir ipucu ortaya çıkardı.

Bsharat ve diğerleri (2024), önde gelen LLM'leri yönlendirmek için 26 farklı yol gösterici ilkeyi sistematik olarak incelemiş ve bunların etkinliğinin sistematik bir değerlendirmesine dayanarak yönlendirme tavsiyeleri sunmuştur. İlginç bir şekilde, LLM'lere (hayali olarak) bir ipucu sunmak veya bir ceza tehdidinde bulunmak gibi stratejilerin performansı artırabileceğini bulmuşlardır.

3.1.4 Tekrar Üretilirlik

Bu makalenin geri kalanındaki uygulamaların çoğu, makalenin Aralık 2023 JEL sürümünün yazıldığı tarihte kamuya açık olan önde gelen LLM'yi, OpenAI'nin GPT-4, gpt4-0613 sürümünü kullanmaktadır. Bu makaleyle ilişkili çevrimiçi materyallerde (makalenin ön sayfasındaki dipnota bakınız), OpenAI'nin API'sini çağırarak sonuçları yeniden üretmek için Python kodu sağlıyorum. Kod, "sıcaklık" parametresini sıfıra ayarlayarak LLM yanıtlarını deterministiğe yakın hale getiriyor. Programcı olmayanlar sonuçları <https://platform.openai.com/playground> web tabanlı deneyim platformunda yeniden lisanslayabilirler. Hem OpenAI API hem de Playground platformu GPT-4'e erişmek için ücretli bir alt abonelik gerektirmektedir.⁴

Yeniden üretimi sınırlayan iki faktör vardır Sonuçlarının doğruluğu. İlk olarak, OpenAI "sıcaklığı 0'a ayarlamanın çıktıları çoğunlukla deterministik hale getireceğini, ancak az miktarda değişkenliğin kalacağını" belirtmektedir. Bu tekrarlanabilirlik sınırlarını özellikle şu yanıtları içeren örnekler için gözlemledim

çıkış denedim. Benim en iyi açıklamam, modelin Brookings'in yüksek kaliteli analizleriyle tanıdığını içselleştirmiş olmasıdır. (Yasal Uyarı: Yazar Brookings'te yerleşik olmayan bir araştırmacıdır).

⁴ Aşağıda GPT3.5/GPT-4 olarak etiketlenen tüm örneklerin çalıştırılması için 5k'dan biraz fazla girdi ve 5k çıktı gerektirmiştir her biri token. Ekim 2023'te toplam maliyet 50 sentin biraz

altındaydı, ancak bu maliyet o zamandan beri önemli ölçüde azaldı. GPT-4o kullanıldığında, aynı token sayısı yaklaşık 10 sente mal olmaktadır. OpenAI'nin modelleri için güncel fiyatlandırma bilgileri <https://openai.com/pricing> adresinde mevcuttur.

birden fazla cümleyi kapsamaktadır.⁵ İkinci olarak, OpenAI "daha güvenli ve daha yetenekli modeller piyasaya sürdükçe, eski modelleri düzenli olarak kullanımdan kaldırıyoruz" demektedir. Ayrıca, "yeni bir sürüm piyasaya sürüldükten sonra, eski sürümler genellikle 3 ay sonra kullanımdan kaldırılacaktır." Eğer gpt4-0613 modeli kullanımdan kaldırılırsa, sonuçların artık tekrarlanabilir olmayabilir.⁶

En uygun kullanıcı arayüzü <https://chat.openai.com/> adresinde bulunan ChatGPT'dir ve sıfırdan büyük bir "sıcaklık" parametresi kullanır, bu da modelin yanıtlarına daha fazla değişkenlik katar. GPT-4'e bu arayüz üzerinden erişmek için ChatGPT Plus'a ücretli abonelik gerekmektedir. Bu, kullanıcıların aşağıdaki GPT-4'ü kullanan tüm örneklerin ruhunu denemelerine olanak tanır, ancak ekstra değişkenlik, kesin sonuçların bir komut istemi her çalıştırıldığında farklı olacağı anlamına gelir. Aynı durum ChatGPT'nin bir parçası olan ChatGPT Gelişmiş Veri Analizi için de geçerlidir, ve Claude 2/3'e dosya yükleme olanağı sunuyor. Bu nedenle benim çoğaltma kodum ikinci modellerin sonuçlarını içermemektedir.

Şimdi eko- nomik araştırmalarda üretken yapay zekanın kullanımına dönüyorum.

matter, resulting in different word choices. Ve bir kelime değiştiğinde, onu takip eden her şey farklılaşır.

⁶ Ayrıca, OpenAI'nin model kullanımdan kaldırma politikasının yanı sıra mevcut modellerin ne kadar süreyle kullanılabilir kalacağının garanti edildiğine ilişkin mevcut zaman çizelgeleri hakkında <https://platform.openai.com/docs/deprecations> adresine bakın.

⁵ Sıfır sıcaklıkta bile model çıktısının tutarsızlığı hakkında daha fazla bilgi için <https://platform.openai.com/docs/guides/gpt/why-are-model-outputs-inconsistent> ve LLM çıkarımını verimli bir şekilde gerçekleştirmenin doğasında var olan belirsizliğe ilişkin bir tartışma için <https://community.openai.com/t/a-question-on-determinism/8185> adreslerine bakın. Özetle, yüz milyarlarca parametreye sahip LLM'lerin verimli bir şekilde yürütülmesi, hesaplamaların paralelleştirilmesini gerektirir. Ancak, bilgisayarların ayrık doğası göz önüne alındığında, $(a \cdot b) \cdot c$ gibi hesaplamalar bazı- kez $a \cdot (b \cdot c)$ 'den biraz daha farklı bir sonuç verir. Bir LLM hangi kelimenin en yüksek olasılığa sahip olduğunu hesapladığında bility to be next, the minor differences in the parallelization of the exact same calculations sometimes come to

3.2 Fikir Üretme ve Geri Bildirim

Araştırma, fikir üretme, geliştirme ve seçme süreci ile başlar. LLM'leri araştırmaya iki nedenden dolayı fikir ve geri bildirim içeren kullanım senaryoları ile başlıyorum. Birincisi, fikirlerle başlamak araştırmanın doğal sırasını takip eder. İkincisi, fikir üretme ve geri bildirim, LLM'leri araştırmadaki derin öğrenmenin önceki uygulamalarından keskin bir şekilde ayıran yeni bir dizi yeteneği sergiliyor - uzun zamandır insanlara ayrılmış bir yaratıcılık biçimini ortaya çıkarıyorlar. Fikir üretme ve geri bildirim, mevcut en gelişmiş LLM'leri kullanmanın karşılığını veren alanlardır. Girotra, Terwiesch ve Ulrich (2010) tarafından geliştirilen bir fikir üretme modeli, yaratıcı performansın (i) fikirlerin miktarına, (ii) fikirlerin ortalama kalitesine ve (iii) ne kadar istisnai fikir üretildiğini belirleyen varyansa bağlı olduğunu gözlemlemektedir. Girotra ve diğerleri (2023), GPT-4'ün yenilikçi yeni ürün fikirleri geliştirme yarışmasında, ABD'nin önde gelen bir işletme okulundaki MBA öğrencilerinden her üç boyutta da daha iyi performans gösterdiğini bulmuştur. Sonuç olarak, fikir üretmedeki darboğazın giderek fikir üretmekten fikir değerlendirmeye doğru kaydığını ileri sürmektedirler.

İleride göreceğimiz gibi, her ne kadar İleri teknoloji LLM'lerin fikir üretme ve geri bildirim alanlarındaki mevcut yetenekleri etkileyicidir, ancak aynı zamanda sınırlamaları da vardır. Ayrıca daha geniş potansiyel tuzaklar da vardır. Fikir üretme ve geri bildirim için LLM'leri kullanan herhangi bir araştırmacı, elbette, belirli bir kullanım durumunda hangi noktaları kullanıp hangi noktaları reddettikleri konusunda dikkatli olmalıdır - tıpkı meslektaşlarımızla fikirlerimizi tartışırken yaptığımız gibi. Zaman içinde ortaya çıkan ince dezavantajlar da olabilir. LLM tarafından üretilen fikirlere duyulan güven, bireysel araştırmacıların otomasyona daha fazla güvenmesine ve

kendi eleştirel düşüncelerini daha az uygulamalarına neden olabilir. Dahası, giderek daha fazla sayıda ekonomist fikir üretmek ve geri bildirim almak için aynı bir ya da iki ileri teknoloji LLM'ye güvenirse, bu durum

ekonomistlerin üzerinde çalıştığı fikirler giderek daha homojen hale gelecek ve daha az sayıda gerçekten yeni fikir içerecektir. Bu homojenleşme riski Bommasani ve diğerlerinde de (2021) tartışılmaktadır. Dahası, beyin fırtınası veya geri bildirim için LLM'leri kullanırken, eğitim verilerinin ne zaman kesildiğini akılda tutmak önemlidir (bkz. Tablo 1).

3.2.1 *Beyin Fırtınası*

Son teknoloji LLM'ler, tanımlanmış bir temayla ilgili fikir ve örneklerin beyin fırtınası (ya da belki de daha uygun bir ifadeyle sinir ağı fırtınası) için oldukça kullanışlıdır. Tüm insan bilgisinin bir kesitini temsil eden büyük miktarda veri üzerinde eğitilmiş olan son teknoloji LLM'lerin eğitim verilerinden geliştirdikleri dünya temsilinin genişliği, ekonomi hakkında oldukça fazla bilgi içerir. Bununla birlikte, derinlik söz konusu olduğunda insan uzmanların hala bir üstünlüğü vardır, bu nedenle LLM'ler kişinin uzman olmadığı alanlarda beyin fırtınası yapmak için en uygun yöntemdir.

Aşağıdaki komut istemi basit bir GPT-4 kullanan örnek. Yanıtın her bir nokta için on kelimeyle sınırlandırılması için bir talimat eklediğime dikkat edin; aksi takdirde LLM her bir nokta için bütün bir paragraf üretir ki bu genel olarak faydalı olabilir ancak buradaki açıklayıcı amaçlarımız için çok uzun olacaktır:

Sohbet 2 (GPT-4): *Yapay zekanın eşitsizliği artırabileceği 20 kanal hakkında beyin fırtınası yapabilir misiniz? Yanıtınızı her bir nokta için 10 kelime ile sınırlayın.*

1. Düşük vasıflı sektörlerde otomasyon nedeniyle iş kaybı.
2. Teknoloji meraklısı kuruluşlarda yapay zeka odaklı servet yoğunlaşması.
3. Dijital okuryazarlık açığı ekonomik eşitsizliğe yol açıyor.
4. Yapay zeka teknolojisine evrensel erişim eksikliği.
5. İşe alım ve seçim süreçlerinde yapay zeka kaynaklı önyargı.
6. Yapay zeka uzmanlıkları nedeniyle iş

7. Hassas nüfusları etkileyen veri gizliliği sorunları.
8. Ağırlıklı olarak varlıklı tüketicileri hedefleyen yapay zeka odaklı hizmetler.
9. Algoritmalar önyargı yoluyla sosyal eşitsizliği daha da kötüleştiriyor.
10. Kapsayıcı yapay zeka ürün geliştirme eksikliği.
11. Yapay zeka ile geliştirilmiş ürünler/hizmetler nedeniyle daha yüksek fiyatlar.
12. Teknoloji merkezli şehirlerde yapay zeka destekli soylulaştırma.
13. Yapay zeka uygulamaları ile desteklenen rekabete aykırı uygulamalar.
14. Yapay zeka tarafından yerinden edilen işler için işçi haklarının olmaması.
15. Yapay zeka öğrenme araçları nedeniyle eğitimde dengesizlik.
16. Alt sosyoekonomik gruplar hariç olmak üzere sağlık hizmetlerinde yapay zeka.
17. Yapay zekanın siyasi kampanyalardaki orantısız etkisi.
18. Yapay zeka lehine insan becerilerinin değersizleştirilmesi.
19. Önyargılı YAPAY ZEKA sistemler ayrımcılığın sürdürülmesi.
20. Yapay zeka, veri analizi yoluyla toplumsal hiyerarşileri güçlendiriyor.

konusu üzerinde beyin fırtınası yapıp her birini bir cümleyle açıklayabilir misiniz?

Model tarafından listelenen kanallar konuyla ilgili, büyük ölçüde makul ve aklıma gelenlerin geniş bir yelpazesini ve biraz daha fazlasını kapsıyor. Bununla birlikte, bazı madde işaretleri biraz tekrarlayıcı.

Altı çizilmesi gereken önemli bir husus, beyin fırtınası gibi faaliyetler sırasında LLM'ler tarafından üretilen yanıtların dikkat çekici hızı ve hacmidir; bu da kendine özgü bir yararlılık biçimi oluşturur. Bu gibi örneklerde 20 öneriden sadece bir tanesinin faydalı olduğu kanıtlanırsa bile, araştırmamızı önemli ölçüde daha üretken hale getirebilir.

Yararlı bulduğum diğer beyin fırtınası ipuçları şunlardır:

- Yapay zeka ve eşitsizlik üzerine çalışan bir ekonomistim. Bir beyin fırtınası yapabilir misin-satırında [konuyu girin]?
- Yapay zeka üzerine çalışan bir ekonomistim ve Eşitsizlik. Beş potansiyel makale

- Ben [konuyu giriniz] üzerine akademik bir makale üzerinde çalışan bir ekonomistim. Benim için bir araştırma planı beyin fırtınası yapabilir misiniz?

3. Sınırlı Veri: AGI'nin henüz ilk aşamalarında olduğu göz önüne alındığında, ekonomik analiz için sınırlı veri mevcuttur. Bu veri eksikliği, ekonomik

3.2.2 Geri bildirim

LLM'ler ayrıca fikirleri değerlendirebilir, örneğin farklı hipotezlerin veya araştırma planlarının artılarını ve eksilerini vurgulayabilir. Aşağıdaki örnekte, LLM'den belirli bir araştırma alanında çalışmanın artılarını ve eksilerini listelemesi istenmektedir. Bu, LLM'lerin farklı araştırma yönleri hakkında faydalı girdiler sağlayabileceğini göstermektedir.

Sohbet 3 (GPT-4): *AGI ekonomisi üzerinde çalışmanın en önemli üç avantajı ve dezavantajı nedir?*

AGI Ekonomisi Üzerine Çalışmanın Artıları:

1. Yüksek Etki: AGI'nin (Yapay Genel Zeka) gelişimi, sağlıktan finansa ve ulaşılmaya kadar ekonominin birçok sektöründe devrim yaratma potansiyeline sahiptir. AGI ekonomisi üzerinde çalışmak, bu dönüşümün ön saflarında yer almanıza ve potansiyel olarak yönünü şekillendirmenize olanak tanır.
2. Disiplinlerarası Yaklaşım: AGI ekonomisi, diğer disiplinlerin yanı sıra ekonomi, bilgisayar bilimi ve felsefenin bir kombinasyonunu içerir. Bu disiplinler arası yaklaşım entelektüel açıdan teşvik edici ve ödüllendirici olabilir.
3. Yenilikçilik Potansiyeli: AGI'nin nispeten yeni ve hızla gelişen bir alan olduğu göz önüne alındığında, yenilikçi düşünme ve yeni yaklaşımlar için çok fazla alan vardır. Bu, problem çözmekten ve yaratıcı düşünmekten hoşlananlar için heyecan verici ve tatmin edici olabilir.

AGI Ekonomisi Üzerine Çalışmanın Eksileri:

1. Belirsizlik: AGI'nin geleceği oldukça belirsizdir, bu da doğru tahminler yapmayı veya sağlam ekonomik modeller geliştirmeyi zorlaştırmaktadır. Bu belirsizlik işi zorlaştırabilir ve bazen de sinir bozucu olabilir.
2. Etik ve Toplumsal Etkiler: AGI'nin geliştirilmesi ve yaygınlaştırılması, işlerin yer değiştirmesinden mahremiyet kaygılarına ve potansiyel kötüye kullanıma kadar bir dizi etik ve toplumsal sorunu gündeme getirmektedir. AGI ekonomisi üzerinde çalışmak, bu karmaşık ve çoğu zaman tartışmalı konularla boğuşmak anlamına gelir.

hipotezleri test etmek veya kesin sonuçlara varmak zordur.

güçlendirin. Görünüm bölümü düşündürücüdür ancak spekülatif kalmaktadır.

Geribildirim almak için yararlı bir ipucu örneği de şudur:

- Ben [konuyu girin] üzerine akademik bir makale üzerinde çalışan bir ekonomistim. Bu konuyu araştırırken karşılaştığınız başlıca zorluklar nelerdir? konu? Onlara en iyi nasıl hitap edebilirim?

Yineleme: Beyin fırtınası ve değerlendirme arasında yineleme yapmak özellikle yararlıdır. Bir araştırmacının fikirler üretmesi, en umut verici olanları seçmesi ve bunları geliştirmesine benzer şekilde, LLM'lerden beyin fırtınası yapmaları, en umut verici olarak değerlendirdikleri fikirleri seçmeleri ve bunlar üzerinde daha fazla beyin fırtınası yapmaları istenebilir.

Claude 2'nin uzun bağlam penceresi, tüm araştırma makalelerini LLM'ye yüklemeyi ve geri bildirim istemeyi mümkün kılar. Bu makalenin Şubat 2023 NBER çalışma raporu versiyonunu (Korinek 2023a) Claude 2'ye yükledim ve aşağıdakileri sordum:

Sohbet 4 (Claude 2): *Bu makaleyi geliştirmenin yollarını önerebilir misiniz?* [Ek: w30957.pdf]

Büyük dil modelleri ve bilişsel otomasyon hakkındaki makaleyi geliştirmek için bazı öneriler:

- Farklı kullanım durumlarını açıklarken daha spesifik ayrıntılar ve örnekler verin. Bazıları için açıklamalar biraz belirsiz. Daha somut örneklerin eklenmesi bu bölümleri güçlendirecektir.
- LLM'lere aşırı güvenmenin sınırlamaları ve potansiyel olumsuzluklarına ilişkin tartışmayı genişletin. Çalışma çoğunlukla faydalara odaklanmaktadır, ancak risklerin daha dengeli bir şekilde tartışılması faydalı olacaktır.
- Uzun vadeli sonuçlara ilişkin analizi

Tahminleri desteklemek için daha titiz analizler eklenmesi daha ikna edici olabilir.

- Kullanım senaryolarını birden fazla LLM üzerinde test etmeyi ve yeteneklerini karşılaştırmayı düşünün. Örneklerin hepsi tek bir LLM kullanıyor gibi görünüyor, ancak Farklı olanları test etmek, güçlü/zayıf yönlerdeki farklılıkları ortaya çıkarabilir.
- LLM çıktılarının kullanılmadan önce nasıl doğru bir şekilde değerlendirileceği konusunda rehberlik sağlayın. LLM'ler hata yapabildiğinden, sonuçların iki kez kontrol edilmesine ilişkin bazı ipuçları faydalı olacaktır.
- (ek öneriler yer nedeniyle atlanmıştır)

Genel olarak, daha fazla ayrıntı, analiz, örnek ve yapı sağlamak, makaleyi umut verici bir başlangıçtan daha parlak ve ikna edici bir nihai ürüne götürebilir. Bu önerilerden herhangi birini detaylandırmamı isterseniz bana haber verin!

Listelenen tüm madde işaretleri iyi önerilerdir ve revizyon sürecinde bunların birçoğunu makaleye dahil ettim.

Claude 2 tüm makalenin içeriğini hafızasında tutabildiğinden, talep edilmesi halinde makalenin herhangi bir bölümü hakkında yorum sunabilir. Aşağıda yararlı istemlere ilişkin ek örnekler yer almaktadır:

- Bu makalenin temel güçlü ve zayıf yönleri nelerdir?
 - Makalede yeterince vurgulanmayan ana yeni fikirler nelerdir?
 - Bu makalede herhangi bir önyargı örneği tespit edebilir misiniz?
 - Bölüm [numarayı giriniz]'i nasıl geliştirebiliriz?
 - Bunun için bir hakem raporu hazırlayabilir misiniz?
- Ekonomik Edebiyat Dergisi* için bir makale mi?

Son örnekte ortaya çıkan yetenek, iyi ya da kötü, editörlük ve hakemlikte devrim yaratacak gibi görünüyor. Bir örnek vermek gerekirse, Claude 2'den aynı çalışma makalesi için bir hakem raporu hazırlamasını istedim (Korinek 2023a):

Sohbet 5 (Claude 2): Journal of Economic Literature için bu makale için bir hakem raporu hazırlayabilir misiniz? [Ek: w30957.pdf]

"Large Language Models and Cognitive Automation for Economic Research" başlıklı makalenin taslak hakem raporunu burada bulabilirsiniz:

Journal of Economic Literature editörlerine:

Anton Korinek tarafından kaleme alınan "Large Language Models and Cognitive Automation for Economic Research" başlıklı makaleye ilişkin incelememi tamamladım.

Bu makale, ChatGPT gibi büyük dil modellerinin (LLM'ler) ekonomistler için çeşitli araştırma görevlerini otomatikleştirmek ve geliştirmek için nasıl kullanılabileceğine dair zamanında ve yararlı bir bakış açısı sunmaktadır. Makale iyi yazılmış ve açık bir şekilde sunulmuştur. Küçük revizyonlardan sonra yayınlanmasını tavsiye ederim.

Özet:

Makale, LLM'lerin potansiyel uygulamalarını altı alana -fikir, yazma, arka plan araştırması, kodlama, veri analizi ve matematiksel türetmeler- ayırmaktadır. Her bir alan için yazar, hızlı mühendislik konusunda genel rehberliğin yanı sıra LLM'lerin yeteneklerini ve sınırlamalarını göstermek için özel örnekler sunmaktadır. GPT-3 modelini kullanan deneylere dayanarak, yazar her bir uygulamaya deneyselden (gözetim gerektiren tutarsız sonuçlar) oldukça yararlıya (muhtemelen araştırmacıya zaman kazandıracak) kadar değişen bir kullanılabilirlik derecesi atamaktadır.

Çalışma, LLM'lerin kısa ve uzun vadede ekonomik araştırma üretkenliğini nasıl etkileyebileceğini tahmin ederek sonuçlanmaktadır. Kısa vadede yazar, LLM'lerin mikro görevlerin otomatik olarak yerine getirilmesinde faydalı asistanlar olarak görev yapacağını ve araştırmacıların üretkenliğini artıracaklarını savunmaktadır. Uzun vadede ise yazar, LLM'lerin üstün ekonomik modeller üreterek potansiyel olarak insan ekonomistlerin yerini alabileceğini düşünüyor. Makale, ekonomistlerin bilişsel otomasyonun etkilerine ilişkin keşfetmeleri gereken yeni araştırma soruları önererek sona eriyor.

Yorumlar:

Çalışma, ekonomi uzmanlarının çalışmalarında LLM'lerden yararlanmaya başlamaları için erişilebilir bir giriş ve çerçeve sağlama amacıyla başarılıdır. Potansiyel uygulamalar kataloğu kapsamlı ve örnekler aydınlatıcıdır. Hızlı mühendislik, LLM'leri etkin bir şekilde kullanmak

için son derece önemlidir ve yazar,
tekrarlanabilirlik için yararlı ipuçları içermektedir.

Çalışmayı daha da güçlendirmek için sadece küçük önerilerim var:

Amerikan dergisinde yayınlanmasının uygun olacağını düşünmüyorum.

- Bu alandaki gelişmelerin hızı göz önüne alındığında, deneylerin zamanlaması ve kullanılan belirli LLM'nin yeteneklerinden bahsetmek, sonuçların genelleştirilebilirliğini çerçevelemeye yardımcı olacaktır.
- Kullanışlılık derecelendirmeleri önemlidir. LLM'leri kullanmış olan araştırmacılarla anket yapmak daha sağlam değerlendirmeler ortaya çıkarabilir.
- Tutsuz ve son derece faydalı kabiliyetleri belirleyen faktörlere ilişkin ilave rehberlik, beklentileri daha iyi belirleyebilir.
- LLM'lerin insanları tamamladığı işbirlikçi uygulamaları keşfetmek, uzun vadeli görünümü zenginleştirecektir.

Genel olarak, bu makalenin hızla gelişen bir alanda değerli bir katkı sağladığına inanıyorum. Üretkenliklerini artırmak için LLM'lerden yararlanmak isteyen araştırmacılar için yararlı bir rehber görevi görecektir. Küçük revizyonlardan sonra yayınlanmasını tavsiye ediyorum.

Herhangi bir açıklama veya ek geri bildirim sağlayabilirsem lütfen bana bildirin.

Saygılarımla,

[İsminiz burada]

LLM, makalenin gerçek (muhtemelen insan) hakemleri tarafından yapılan yorumlara benzeyen birkaç yorum sundu ve bu makalenin gözden geçirilmiş bir versiyonunun *Journal of Economic Literature* için uygun olacağını yazdı. Ancak ilginç bir şekilde, taslak hakem raporundaki yorumlar, önceki sohbetteki yorumlardan daha sığdır ve bu da doğrudan Claude 2'yi makaleyi geliştirmenin yolları için harekete geçirmiştir.

Claude 2'nin arkadaş canlısı ve iyimser olduğu iyi bilinmektedir. Bir önceki sohbetteki olumlu değerlendirmenin sadece olumlu bir önyargıyı yansıtıp yansıtmadığını kontrol etmek için Claude 2'ye makalenin *American Economic Review* için uygun olup olmayacağını da sordum:

Chat 5 (Claude 2) ctd: *Bu makale American Economic Review için uygun olur mu?*

İncelemelerime dayanarak, bu makalenin

Ekonomik İnceleme [AER] mevcut haliyle. İşte benim değerlendirmem:

American Economic Review, ekonomik analiz ve bilgi alanında önemli ilerlemeler sağlayan teorik ve ampirik araştırmaları yayınlayan en prestijli ve seçici ekonomi dergilerinden biridir.

Bu makale, ekonomistlerin araştırmalarında büyük dil modellerini nasıl uygulayabileceklerine dair faydalı ve güncel bir genel bakış sunsa da, yeni bir araştırma veya analiz sunmaktan ziyade daha açıklayıcı ve göstericidir. Bu nedenle, *AER* tarafından talep edilen önemli yeni ekonomik anlayışlara katkıda bulunma seviyesine yükselmeyebilir.

AER ile ilgili dikkate alınması gereken bazı faktörler uygunluk:

- Makale, potansiyel LLM uygulamalarını kataloglamakta ancak bunların ekonomik araştırma veya üretkenlik üzerindeki etkilerini titizlikle değerlendirmemektedir. Faydaları gösteren ve sayısallaştıran ampirik bir analiz makaleyi güçlendirecektir.
- Teorik veya kavramsal ekonomik analiz sınırlıdır. Ekonomik teori ve modelleme için çıkarımların genişletilmesi, *AER* için uygunluğu artırabilir.
- Matematiksel türetmeler bölümü oldukça sınırlı bir kapsama sahiptir. *AER* için daha teknik bir yaklaşım beklenebilir.
- LLM'lerin insan yeteneklerini aşmasına ilişkin uzun vadeli tahminler ilginç ancak spekülatiftir. Bunları resmi ekonomik büyüme modelleriyle ilişkilendirmek, onları daha titiz hale getirecektir.

Ek ampirik analiz, teknik derinlik ve resmi ekonomik modelleme ile bu araştırma hattı potansiyel olarak *AER* için uygun makaleler üretebilir. Ancak, uygulamalara genel bir bakış olarak mevcut haliyle, makale *AER*'nin yenilik ve titizlik standartlarını karşılamayabilir. *Journal of Economic Perspectives* gibi uygulamalara odaklanan dergiler daha uygun olabilir.

Umarım bu yorumlar faydalı olur. Makalenin *AER* için uygunluğuna ilişkin değerlendirmemi genişletmemi isterseniz lütfen bana bildirin. Ek geri bildirim sağlamakdan mutluluk duyarım.

önemli yeni ekonomik anlayışlara katkıda bulunma seviyesine yükselmeyebilir

Claude 2, "yeni bir araştırma veya analiz sunmaktan ziyade daha açıklayıcı ve gösterici olduğu" için makalenin uygun olmayacağını savunmuştur. Rapor şöyle devam etmiştir: "Bu nedenle, talep edilen

AER tarafından." Bu (doğru) değerlendirme, modelin halihazırda farklı dergilerdeki uygun araştırma makalelerinin neler olduğuna karar vermek için temel bir kapasiteye sahip olduğunu göstermektedir.

Claude 2, bunun gibi nitel bir makale hakkında gerçekçi bir geri bildirim sağlayabilirken, mevcut LLM nesli daha ana- litik veya nicel makaleleri anlayışlı bir şekilde değerlendirmekte zorlanmaktadır.

Sesli Sohbetler Aracılığıyla Geri Bildirim [Haziran 2024 Sürümünde Tümüyle Yeni] - Meslektaşlarımdan biri sesli sohbet etmenin son derece yararlı olduğunu bildirdi

ChatGPT veya Claude gibi önde gelen LLM sohbet robotları ile araştırma sorunları hakkında konuşmak, geri bildirim almak veya bir sorundan kurtulmak için. LLM'lerin eğitim verilerinde yakalanan geleneksel bilgeliği sentezleme, bir öğretmen olma veya sorulara daha geniş bir çerçeve sunma yetenekleri, onları bu amaç için çok uygun hale getiriyor. Kendi deneyimlerime dayanarak, LLM'lerle sesli sohbet yoluyla etkileşimde bulunmanın, metin yoluyla etkileşimde bulunmaktan farklı bir deneyim sağladığı konusunda hemfikirim - tıpkı bir meslektaşla bir sorun hakkında konuşmanın bazen yararlı olması gibi yararlı olabilecek bir deneyim.

3.2.3 Karşı Argüman Sağlama

Hangi noktayı tartışıyor olursak olalım, her zaman karşı argümanlar vardır. LLM'ler bir argümanın hangi tarafında olduklarını umursamazlar - belirli bir noktanın lehine argümanlar sunmada olduğu kadar karşı argümanlar sunmada da iyidirler. Ayrıca onlardan bir eleştiri istediğimizde egomuzu incitmekten de çekinmezler. Bu durum, insan beyninde yaygın olan teyit önyargısına karşı koymak için faydalı olabilir. Aşağıda bir örnek verilmiştir (yer darlığı nedeniyle kısa yanıtlar istenmektedir):

Bilişsel otomasyonun eşitsizliği artıracığını öngören model. En iyi beş karşı argüman nedir? Her biri bir satır.

1. Bilişsel otomasyon yeni sektörlerde iş yaratılmasına yol açarak diğer sektörlerdeki iş kayıplarını telafi edebilir.
2. Üretkenliği ve ekonomik büyümeyi artırarak genel refah artışına yol açabilir.
3. Otomasyon potansiyel olarak maliyetleri düşürebilir mal ve hizmetler, tüm tüketicilere fayda sağlar.
4. Beceri kazandırma ve yeniden beceri kazandırma için fırsatlar sağlayarak ücret artışına yol açabilir.
5. Bilişsel otomasyon insan hatalarını azaltarak verimliliği artırabilir ve potansiyel olarak daha adil sonuçlara yol açabilir.

Karşı argümanlar mantıklı, ancak şunu belirtmek isterim ki - çoğu meslekten olmayan kişinin aksine - ekonomistler 2. ve 3. noktaları aynı olarak görecektir, çünkü verimlilik daha ucuz mal ve hizmetlere göre ayarlanan gerçek bir kavramdır.

3.3 Yazma

LLM'lerin temel yetkinliği metin oluşturmaktır. Bu, madde işaretlerine dayalı cümleleri sentezlemekten metnin stilini değiştirmeye, metni düzenlemeye, stilleri değerlendirmeye ve başlıklar, başlıklar ve tweetler oluşturmaya kadar yazmayla ilgili birçok görev için oldukça yetenekli ve yararlı oldukları anlamına gelir. Aşağıda her birini sırayla ele alıyorum.

Temel yazma görevleri genellikle mevcut en ileri LLM'yi gerektirmez. Örneğin, temel metinleri sentezlemek için Tablo 1'de daha önce listelenen modellerden herhangi biri iyi bir iş çıkaracaktır. Ancak, LLM'den anlık talimatların ötesine geçmesi istendiğinde, her modelin yeteneklerine göre farklılıklar ortaya çıkmaya başlar. Bu yazının yazıldığı sırada, metin sentezleme araçları MS Office ve Google Workspace gibi ofis paketlerinin yanı sıra ilgili e-posta programları Outlook ve Gmail'e de dahil edilmeye başlanmıştır. Alternatif olarak, birçok yazılım için eklentiler sunan özel bir yazma asistanı

yaygın uygulamalar (web tarayıcıları ve ofis paketleri dahil) ve kullanıcı yazarken düzeltmeler ve stilistik iyileştirmeler öneren grammarly.com.

3.3.1 Metin Sentezleme

LLM'lerin yazma yardımı için en faydalı yeteneklerinden biri, kaba madde işaretlerini açık ve okunması kolay, iyi yapılandırılmış cümlelere dönüştürmek olabilir. Aşağıda bir örnek verilmiştir. Maddelerin aynı zamanda LLM'nin örnekler sunması ve çözümler önermesi için talepler içerdiğine dikkat edin.

Sohbet 7 (GPT-4): Akademik bir ekonomist tarzında aşağıdaki noktaları bütünüştiren bir paragraf yazın. Bir konu cümlesi ekleyin.

- * Büyük dil modelleri (LLM'ler) birçok bilişsel görevi otomatikleştirir
- * Akademik ekonomistlerin çalışmaları büyük ölçüde mantıksaldır [örnekler ekleyin]
- * Bu nedenle biz ekonomistler şu konularda endişelenmeliyiz
- * dil modellerinde son gelişmeler
- * Kısa vadede, bu modeller daha üretken olmamıza yardımcı olabilir
- * Uzun vadede, insan ekonomistlerin yerini almaları büyük bir olasılıktır
- * nasıl hazırlanacağımızı [çözümler sunacağımızı] düşünmeliyiz.

Çok sayıda bilişsel görevi otomatikleştiren büyük dil modellerinin (LLM'ler) ortaya çıkışı, veri analizi, ekonomik tahmin ve politika önerisi gibi görevleri içeren işimizin bilişsel doğası göz önüne alındığında, akademik ekonomi uzmanlarının farkında olması gereken önemli bir paradigma değişikliği sunmaktadır. Dil modellerindeki hızlı ilerlemeler, mesleğimizi potansiyel olarak sekteye uğratabilir ve bu gelişmekte olan teknolojiye proaktif bir yaklaşım gerektirir. Kısa vadede, bu modeller rutin görevleri otomatik hale getirerek üretkenliğimizi artırabilir ve böylece ekonomik araştırmanın daha karmaşık ve incelikli yönlerine odaklanmamızı sağlayabilir. Ancak, uzun vadede, bu modellerin karmaşık bilişsel görevleri yerine getirmede giderek daha becerikli hale gelmeleri nedeniyle insan ekonomistlerin yerini alma olasılığı oldukça yüksektir. Sonuç olarak, bu modellerin

Bu olasılığa hazırlıklı olmalıyız. Bu, bu modellerin yeteneklerinden etkili bir şekilde yararlanmak için bu modeller hakkında daha derin bir anlayış geliştirmeyi, beceri setlerimizi kolayca otomatikleştirilemeyen alanları içerecek şekilde çeşitlendirmeyi ve bu teknolojik gelişmelerle potansiyel olarak yerinden edilenler için adil bir geçiş sağlayan politikaları savunmayı içerebilir.

Bunu kullanmak, araştırmacıların enerjilerini sıradan yazma süreci yerine metinlerindeki fikirlere yoğunlaştırmalarını sağlar. Bazıları metnin biraz steril olduğunu ve insan yazısını karakterize eden kendine has özelliklerden ve sürpriz unsurlarından yoksun olabileceğini iddia etse de, metin genellikle yüksek kalitededir.⁷

LLM'ler farklı tarzlarda yazma taleplerine çok duyarlıdır. "Akademik bir ekonomist tarzda" ifadesini yazma talebiyle değiştirdiğimde hem etkileyici hem de komik sonuçlar elde ettim:

- günlük konuşma tarzında
- iktisatçı olmayanların da anlayabileceği bir üslupla
- tarzında [favori ekonomistinizi, politikacınızı vb. girin]⁸

LLM'ler ayrıca matematiksel ifadeleri gerekli formatta otomatik olarak biçimlendiren LaTeX formatında metin yazabilir ve madde işaretlerine dayalı LaTeX kodunda beamer slaytları hazırlayabilir.

Yazma elbette sadece araştırma için yararlı bir beceri değildir - birçok alanda bilişsel çalışanlar, örneğin ChatGPT'yi e-postalar, notlar, performans değerlendirmeleri veya birkaç maddeye dayalı özetler hazırlamak için kullandıklarını bildirmektedir. Aynı şekilde profesörler de

⁷ Bu, örneğin <https://gptzero.me/> adresinde bulunan GPTZero aracı gibi LLM ile yazılmış metin dedektörlerinin sıfırladığı bir özelliktir. Ancak, bu tür araçların her zaman güvenilir olmadığı ve kolayca atlatılabileceği unutulmamalıdır.

Hızlı, Daha Ucuz

⁸ "Bir gangster tarzında" diye rica ettiğimde, LLM "Burayı dinleyin, bakın, bu kodaman büyük dil modelleri (LLM'ler) bizim bölgemizde cirit atıyor..." diye başladı.

LLM'leri kullanarak tavsiye mektuplarının veya ders programlarının ilk taslağını oluşturmak.

açıklayabilir, böylece yazarlar yaptığı düzeltmelerden ders çıkarabilirler

3.3.2 Metin Düzenleme (Hatalar, Stil, Netlik, Basitlik... için)

Bir başka faydalı yetenek de düzenlemedir. LLM'ler metni dilbilgisi veya yazım yanlışları, stil, netlik veya basitlik açısından düzenleyebilir. Bu yetenekler tüm yazarlar için yararlıdır, ancak belki de en çok yazarlarını geliştirmek isteyen ana dili İngilizce olmayanlar için yararlıdır. Aşağıdaki örnekte, LLM tarafından yapılan her düzeltmeyi vurgulamak için kalınlaştırdım:

Sohbet 8 (GPT-4): *Aşağıdaki cümleyi düzeltebilir misiniz? Bu makale, büyük dil modellerinin (LLM'ler) eko- nomik araştırmalar için yararlı olan yeni yetenekler sergilediği altı alanı tanımlamaktadır.*

Bu makale, büyük dil modellerinin (LLM'ler) ekonomik araştırmalar için yararlı olan gelişmekte olan yetenekler sergilediği altı alanı açıklamaktadır.

Sohbet 8 (GPT-4) ctd: *Yaptığınız ana düzeltmeleri açıklayabilir misiniz?*

Elbette, işte ana düzeltmeler:

1. "Bu kağıtlar" ifadesi "Bu kağıtlar" olarak düzeltilmiştir. Özne-fiil uyumu yanlıştı. Eğer özne tekilse ("paper" gibi), fiil de tekil olmalıdır ("describes" gibi).
 2. "describe" ifadesi "describes" olarak düzeltilmiştir. Fiil, tekil özne "paper" ile eşleşmelidir.
 3. "domain" ifadesi "etki alanları" olarak düzeltilmiştir. Bağlam birden fazla etki alanı olduğunu ima etmektedir, bu nedenle çoğul form kullanılmalıdır.
 4. "displays" ifadesi "display" olarak düzeltilmiştir. "LLM'ler" alt başlığı çoğuldur, bu nedenle fiil de çoğul olmalıdır.
 5. "emergant" kelimesi "emergent" olarak düzeltilmiştir. Bu bir yazım hatasıydı.
 6. "ekonomik araştırmalara" ifadesi "ekonomik araştırmalar için" olarak düzeltilmiştir. Bu bağlamda "için" edatı daha uygundur.
-

Takip eden soruda da gösterildiği gibi, LLM yaptığı düzenlemeleri büyük ölçüde

yapılmıştır. Bu son özellik GPT-3.5'ten GPT-4'e geçerken önemli ölçüde geliştirilmiştir. Bununla birlikte, listeden yedinci bir düzeltmenin çıkarıldığına dikkat edin - "yetenekler" kelimesindeki bir yazım hatasını da düzeltmiştir. Bu durum, LLM'lerin sonuçlarının dikkatle kontrol edilmesi gerektiğine ışık tutmaktadır.

Bir başka faydalı LLM uygulaması da metni farklı seviyelerdeki okuyucular için anlaşılır hale getirmektir; örneğin, "Aşağıdaki metni sekiz yaşındaki bir çocuk için anlaşılır olacak şekilde yeniden yazın", "sekizinci sınıf öğrencisi için anlaşılır" veya "ekonomi eğitimi almamış bir politika yapıcı için anlaşılır" gibi bir istem kullanarak.

3.3.3 Metin Değerlendirme

LLM'ler ayrıca metni üslubu, anlaşılabilirliği veya benzeri açılardan da değerlendirebilir. Bu makalenin özetinin taslak versiyonuyla ilgili aşağıdaki soru biraz moral bozucu bir deneyim oldu:

Sohbet 9 (GPT-4): Aşağıdaki özetin en önemli beş biçimsel eksikliğinin kısa bir listesini verin:

Üretici yapay zeka, özellikle de ChatGPT gibi büyük dil modları (LLM'ler), araştırmalarda devrim yaratma potansiyeline sahiptir. LLM'lerin hem araştırma asistanları hem de eğitmenler olarak faydalı olmaya başladığı altı alan boyunca düzinelerce kullanım durumunu açıklıyorum: fikir ve geri bildirim, yazma, arka plan araştırması, veri analizi, kodlama ve matematiksel türetmeler. LLM yeteneklerini denyselden son derece kullanışlıya doğru sınıflandırarak, bunların her birinden nasıl yararlanılacağına dair genel talimatlar veriyor ve belirli örnekler gösteriyorum. Ekonomistlerin mikro görevleri otomatikleştirmek için üretken yapay zekadan yararlanarak önemli üretkenlik kazanımları elde edebileceklerini savunuyorum. Dahası, tüm bu alanlardaki YZ sistemlerinin performansı artmaya devam ettikçe bu kazançlar da artacaktır. Ayrıca, ekonomik araştırmalar için YZ destekli bilişsel otomasyonun uzun vadeli etkileri hakkında da spekülasyon yapıyorum. Bu makaleyle bağlantılı çevrimiçi kaynaklar, nasıl başlanacağına dair talimatlar sunmakta ve ekonomistler için yararlı olan üretken YZ'nin en son yetenekleri hakkında düzenli güncellemeler sağlayacaktır.

1. Açıklık Eksikliği: Özet, net tanımlar veya açıklamalar sunmadan jargon ve teknik terimler kullanıyor, bu da konuya aşina olmayan okuyucuların anlamasını zorlaştırıyor.
2. Edilgen Sesin Aşırı Kullanımı: Soyut fre-sık sık edilgen çatı kullanır, bu da yazının daha az doğrudan ve ilgi çekici görünmesine neden olabilir.
3. Spesifiklik Eksikliği: Özetle "düzinelerce kullanım örneği" ve "altı etki alanı"ndan bahsedilmekte ancak herhangi bir spesifik örnek veya ayrıntı verilmemektedir.
4. Fazlalık: Özet, YZ'nin ekonomistler için potansiyel faydaları fikrini yeni bilgi eklemeden birçok kez tekrarlıyor.
5. Yapı Eksikliği: Özeti net bir yapısı veya akışı yok, bu da okuyucuların yazarın argümanını takip etmesini zorlaştırıyor.

Sohbet 9 (GPT-4) ctd: *Bu eksiklikleri gidermek için yeniden yazabilir misiniz?*

Üretken yapay zeka, özellikle ChatGPT gibi büyük dil modları (LLM'ler), araştırmayı dönüştürme potansiyeline sahiptir. Bu makale LLM'lerin altı temel alandaki uygulamalarını incelemektedir: fikir oluşturma ve geri bildirim, yazma, arka plan araştırması, veri analizi, kodlama ve matematiksel türetmeler. Her bir alan için pratik talimatlar ve gerçek dünyadan örnekler sunarak LLM'lerin kullanılabilirliğini deneyimsel olarak son derece faydalıya doğru kategorize ediyorum. Bu makalenin odak noktası ekonomistler için potansiyel faydalardır. Ekonomi uzmanları, üretken YZ'den yararlanarak mikro görevleri otomatikleştirebilir ve bu da önemli verimlilik artışlarına yol açabilir. YZ sistemleri gelişmeye devam ettikçe, bu faydaların artması beklenmektedir.

Makale ayrıca, ekonomik araştırmalar için yapay zeka destekli bilişsel otomasyonun uzun vadeli etkilerini de incelemektedir. Eşlik eden çevrimiçi kaynaklar, ekonomistler için faydalı olan en son yetenekleri hakkında düzenli güncellemelerle birlikte, üretken yapay zeka ile nasıl başlanacağına dair adım adım bir kılavuz sağlar.

LLM tarafından tespit edilen eksiklikler iyi ele alınmış ve yeniden yazılan versiyon, hepsini olmasa da bazılarını hafifletmek için iyi bir iş çıkarmıştır.

Stil değerlendirmesi için diğer faydalı LLM istemleri şunları içerir:

- Aşağıdaki metinde takip edilmesi en zor argümanlar hangileridir?
- Aşağıdaki metnin açıklığını ve tutarlılığını değerlendiriniz.
- Aşağıdaki metinde herhangi bir belirsizlik var mı?

3.3.4 Denklemleri LaTeX'e dönüştürme

[Haziran 2024 Versiyonunda Tüm Yenilikler]

Çoğu gelişmiş LLM'nin görme yetenekleri, görüntüleri yorumlamalarına ve nasıl işleyeceklerine dair talimatları takip etmelerine olanak tanır. Bu, örneğin matematiksel denklemleri yazarken faydalı olabilir. Aşağıdaki örnek bu kavramı açıklamaktadır.

Sohbet 10 (GPT-4): *Ekteki görüntüyü LaTeX koduna dönüştürün.*

$$Y = F(K, L) = A[K^\alpha + L^\alpha]^\frac{1}{\alpha}, \alpha < 0$$

Sağladığınız denklem için LaTeX kodu şu şekilde görünecektir:

$$Y = F(K, L) = AK^\alpha L^\beta, \alpha, \beta \leq 0$$

Bu LaTeX kodu, Y üretim fonksiyonunu sermaye K ve işgücü L 'nin bir fonksiyonu olarak oluşturacak ve A toplam faktör verimliliğini temsil edecektir, ve α ve β sırasıyla sermaye ve işgücünün çıktı esneklikleri olmak üzere, her ikisi de bu durumda pozitif değildir.

Sonuç açıkça hala eksiktir: köşeli parantezleri atlar, Cobb-Douglas fonksiyonunu bir CES fonksiyonuna dönüştürür,

ρ üssünü α ve β ile değiştirerek ve köşeli parantezin dışındaki üssü atlayarak-ets. Gemini, ChatGPT ile eşit derecede iyi performans gösterdi; Claude 3 bu görevde biraz daha kötü performans gösterdi. Resimlerdeki matematiği LaTeX formatına dönüştürmeye adanmış üstün bir araç

3.3.5 Başlık ve Manşet Oluşturma

Bir başka faydalı uygulama da, bir önceki örnekte kullanılan özet versiyonuna dayanan aşağıdaki örnekte gösterildiği gibi, akılda kalıcı başlıklar veya makale başlıkları bulmaktır:

Sohbet 11 (GPT-4): Aşağıdaki özeti içeren bir makale için iki veya üç akılda kalıcı başlık için beyin fırtınası yapın: [Sohbet 9'daki özet eklenmiştir]

1. "Ekonomik Araştırmalarda Devrim Yaratmak: Üretken Yapay Zeka ve Büyük Dil Modellerinin Gücü"
2. "Ekonomide Yapay Zekayı Serbest Bırakmak: Büyük Dil Modellerinden Yararlanmak İçin Kapsamlı Bir Kılavuz"
3. "Ekonomik Araştırmanın Geleceği: Üretken Yapay Zeka ve Büyük Dil Modellerinin Potansiyelinden Yararlanma"

Sohbet 11 (GPT-4) ctd: Onları biraz daha ayık hale getirin ve kısa tutun.

1. "Ekonomik Araştırmalarda Üretken Yapay Zeka"
2. "Büyük Dil Modelleri: Ekonomistler için Yeni Bir Araç"
3. "Modern Ekonomik Çalışmalarda Yapay Zekanın Rolü"

Önerilen başlıklar tarif edilen çalışmaya oldukça iyi uyuyor ve talep edilen stillere yakından bağlı kalıyor.

3.3.6 Bir Makaleyi Tanıtmak için X Gönderi Oluşturma

Bir makale yazdıktan sonra, bir sonraki adım onu tanıtmaktır. Aşağıdaki alıştırma bir öncekine benzer ancak yapıştırdığım özet X (eski adıyla Twitter) veya benzer sosyal ağlar için iyi çalışacak parçalara dönüştürür:

Sohbet 12 (GPT-4): Aşağıdaki metni özetleyen üç veya dört [yazı] yazın. Basit cümleler kullanın ve akılda kalıcı olmalarını sağlayın: [Sohbet 9'dan özet eklenmiştir].

1. "Üretken yapay zeka, özellikle ChatGPT gibi büyük dil modelleri, araştırmalarda devrim yaratmaya hazırlanıyor! Altı alanda araştırma asistanları ve eğitmenler olarak kullanışlı hale geliyorlar. #AIRevolution #ChatGPT"
2. "Fikir üretmeden kodlamaya kadar, üretken yapay zeka çeşitli görevlerde faydalı olduğunu kanıtıyor. Ekonomistler, mikro görevlerinizi otomatikleştirmenin ve üretkenliği artırmanın zamanı geldi! #AIProductivity #Ekonomi"
3. "Ekonomik araştırmalarda yapay zeka destekli bilişsel otomasyonun geleceği umut verici görünüyor. Üretken yapay zekanın en son yetenekleri hakkında düzenli güncellemeler için bizi izlemeye devam edin. #YapayZekanınGeleceği #EkonomikAraştırma"
4. "Üretken yapay zeka ile başlamak ister misiniz? Talimatlar ve güncellemeler için makalemle ilişkili çevrimiçi kaynaklara göz atın. #GenerativeAI #AIResources"

Talep üzerine, LLM'ler uygun ifadeler veya bağlantılar da ekleyebilir.

3.3.7 Sunum Slaytları Oluşturma

[Haziran 2024 Versiyonunda Tüm Yenilikler]

Başarılı bir tanıtım, makalenizi sunmanız için davet almanızı sağlar. 2024'te piyasaya sürülen en yeni nesil LLM'ler, bir araştırma makalesine dayalı slaytlar hazırlama özelliğine de sahiptir. Bu zor bir uygulamadır çünkü modelin uzun bir makale boyunca ayrıntılı bir hatırlama sergilemesini gerektirir.

Sohbet 13 (Claude 3): Ekteki ekonomik araştırma makalesi için 20 slaytlık bir sunum hazırlayabilir misiniz? [Ek: w30957.pdf]

Anton Korinek'in "Ekonomik Araştırmalar için Dil Modelleri ve Bilişsel Otomasyon" başlıklı araştırma makalesinde yer alan kilit noktaları özetleyen 20 slaytlık bir sunum:

Slayt 1: Başlık Slaytı

- Bildiri başlığı: "Ekonomik Araştırmalar için Dil Modelleri ve Bilişsel Otomasyon"
- Yazar Anton Korinek

Slayt 2: Giriş

- Büyük dil modelleri (LLM'ler) ekonomi araştırmalarında devrim yaratabilir

- 6 alanda LLM'lerin 25 kullanım örneğini açıklar
- Her kullanım durumu için talimatlar ve örnekler sağlar
- Yeteneklerin kullanılabilirliğini deneyimselden çok kullanışlıya doğru sınıflandırır

Slayt 3: LLM'ler nedir?

- LLM'ler temel modellerdir
- Çok sayıda veri üzerinde önceden eğitilmiş büyük derin öğrenme modelleri
- Eğitim verilerinden soyut temsiller ve dünya modeli oluşturma
- Hesaplama ve veri arttıkça yetenekler öngörülebilir ölçeklendirme yasalarına göre gelişir
- Birçok kabiliyet ortaya çıkmaktadır ve belirli eşiklerde aniden ortaya çıkmaktadır

[4-18. Slaytlar yer nedeniyle çıkarılmıştır]

19. Slayt: Ekonomik Fikirlerin Önemi

- Ekonomistler insan kavramlarını yapay zekalar için analitik formlara dönüştürebilir
- Ekonomik fikirler geleceğin yapay zeka sistemlerinin dünya görüşünü etkileyecek
- "Er ya da geç, iyi ya da kötü için tehlikeli olan fikirlerdir" (Keynes)

Slayt 20: Sonuç

- LLM'ler araştırma görevlerini otomatikleştirmek için kullanışlı hale geliyor
- İnsanlar şimdilik araştırma üretme konusunda karşılaştırmalı üstünlüğe sahip
- Verimliliği artırmak için LLM'lerin benimsenmesi önemli
- Hazırlık için potansiyel olarak dönüştürücü uzun vadeli etkiler

Oluşturulan slaytlar, üzerine inşa edilecek mükemmel bir ilk taslak sağlar. Örnekte kullanılan Claude'un son sürümünün en önemli özelliği, makalede yer alan bilgileri yeterince ayrıntılı bir şekilde hatırlaması ve böylece içeriği unutmaya başlamadan uygun materyallerle dolu 20 slayt oluşturabilmesidir. Ayrıca, materyali istenen slayt sayısına uygun şekilde parçalara ayırmıştır.

3.4 Arka Plan Araştırması

Araştırmacıların arka plan araştırması yapmak ve hem mevcut hem de yeni bilgiler edinmek için yerine getirdikleri çeşitli görevler vardır.

Araştırma alanlarında yeni çalışmalar. Bunlar arasında metni özetlemek, ilgili literatürü araştırmak, yeni kavramları anlamaya çalışmak ve metni başka dillerden çevirmek sayılabilir.

3.4.1 Özetleme [Haziran 2024 Versiyonunda Güncellenmiştir]

İlk olarak, LLM'ler özetleme konusunda çok yeteneklidir. Burada gösterildiği gibi, bir metin parçasından en alakalı içeriği çıkarmakta uzadırlar:

Sohbet 14 (GPT-4): Aşağıdaki metni bir cümle ile özetleyin: [Sohbet 9'dan bu makalenin özeti eklenmiştir].

Yazar, Üretken YZ'nin, özellikle ChatGPT gibi büyük dil modellerinin, fikir üretme, yazma, veri analizi, kodlama ve matematik gibi çeşitli alanlardaki araştırmalarda devrim yaratma potansiyelini tartışıyor ve ekonomistlerin mikro görevleri YZ ile otomatikleştirerek üretkenliği önemli ölçüde artırabileceğini ve bu sistemlerin performansının zaman içinde artması beklendiğini öne sürüyor.

Verdiği özet cümle tüm önemli noktaları kapsamaktadır.

Hem Claude hem de ChatGPT artık akademik makalelerin tamamını ve birçok kitabı işlemek için yeterince büyük bir metin penceresine sahip. Bir örnek vermek gerekirse, bu makalenin orijinal Şubat 2023 NBER çalışma makalesi versiyonunu (Korinek 2023a) Claude 2'ye beslemek için "yükle" düğmesini kullandım ve kısa bir özet istedim - burada çok fazla yer kaplamamak için "kısa":

Sohbet 15 (Claude 2): Bu makalenin kısa özetini sağlayın. [Ek: w30957.pdf]

Bu makale, ChatGPT gibi büyük dil modlarının (LLM'ler) bir dizi araştırma görevinde nasıl faydalı yetenekler sergilemeye başladığını açıklamaktadır

ekonomi alanında. Yazar, 25 potansiyel kullanım örneğini altı alana ayırıyor: fikir oluşturma, yazma, arka plan araştırması, veri analizi, kodlama ve matematiksel türetmeler. Makale, her bir alan için, mevcut LLM yeteneklerini gösteren talimatlar ve örnekler sunmakta ve bunları deneylerden son derece kullanışlıya kadar bir ölçekte derecelendirmektedir. Yazar, LLM'lerin mikro görevleri otomatikleştirmeye ve kısa vadede araştırmacı üretkenliğini artırmaya yardımcı olabileceğini savunuyor. Makale, uzun vadede, LLM'ler aracılığıyla bilişsel otomasyonun ekonomik araştırmalar ve insan bilişsel emeğinin değeri üzerinde olumlu etkileri olabileceğini öngörmektedir.

Claude 2'nin uzun bağlam penceresi, tipik bir makalenin tüm metnini çalışma belleğinde tutabileceği anlamına gelir.

Model, içerikle ilgili takip sorularına yanıt verebilir ve sağlanan metin üzerinde önemli ayrıntılara girebilir. Yararlı ipuçlarına ilişkin ek örnekler şunları içerir:

- Bu noktaları [veya özetten belirli bir noktayı] destekleyen kanıtlar nelerdir?
- Yazarın ana çıkarımları nelerdir?
- Bu makaledeki en ilginç (veya en yeni) fikirler nelerdir?
- Bölümün [numarayı girin] ana bulgularını açıklayabilir misiniz?

Claude, kullanıcının beş adede kadar belge yüklemesine ve LLM'den bunları karşılaştırma veya aralarındaki bağlantıları belirleme gibi görevleri yerine getirmesini istemesine olanak tanır.

Video Özetleri [Haziran 2024 Sürümünde Tümüyle Yeni] -Google DeepMind'ın Gemini'si, ilgili uzantı Gemini'nin ayarlarında etkinleştirilirse YouTube videolarını sindirme kapasitesine sahiptir.

Sohbet 16 (İkizler): Bunun bir paragraflık özetini sağlayın: <https://www.youtube.com/watch?v=sjL6Gl6Zlq8>.

Bu video, Anthropic tarafından yaratılan yeni bir büyük dil modeli olan Claude 3 Opus hakkındadır. Videoda, yaratıcılar Claude 3 Opus'un dünya ekonomisini analiz etme yeteneğini test ediyor. Video, Claude 3 Opus'un web görünümünü adı verilen bir araç kullanarak web'deki bilgilere erişebildiğini ve bu bilgileri işleyebildiğini göstermektedir. Daha sonra bu bilgileri soruları yanıtlamak ve görevleri tamamlamak için kullanabilmektedir. Örneğin, video Claude 3 Opus'un ABD GSYİH trendlerinin bir grafiğini oluşturabildiğini ve gelecekteki GSYİH büyümesini tahmin edebildiğini göstermektedir. Videoda ayrıca Claude 3 Opus'un karmaşık problemleri alt problemlere bölebildiği ve işi kendisinin diğer versiyonları arasında dağıtabildiği gösterilmektedir. Bu sayede Claude 3 Opus aynı anda birden fazla ülkenin ekonomisini analiz edebiliyor. Genel olarak video, Claude 3 Opus'un karmaşık eko- nomik verileri analiz etmek için kullanılabilecek güçlü bir araç olduğunu göstermektedir.

Sağlanan özet mükemmel. Bu yeni özellik, videolardan uzun süren bilgi edinme sürecini çok daha basit bir biçimde kolaylaştırarak ekonomik araştırmalar için yeni bilgi kaynakları sağlamayı mümkün kılıyor.

3.4.2 Literatür Araştırması

Bu yazının yazıldığı sırada, LLM'lerin akademik literatürde arama yapmak ve referans vermek için kullanışlılığı sınırlıdır. Atıflar veya referanslar istendiğinde, sıklıkla gerçekte var olmayan, otoriter görünen makaleler üretmektedirler. Bu durum son zamanlarda en yeni LLM'ler için iyileşmiş olsa da-GPT-4 bu açıdan GPT-3.5'ten önemli ölçüde daha iyidir ve genellikle literatürde sıkça atıfta bulunulan standart referanslar sağlayabilmektedir- LLM'lerin oluşturduğu tüm referanslar dikkatlice iki kez kontrol edilmelidir. Aşağıdakiler bir örnektir:

Sohbet 17 (GPT-4): Anton Korinek'in Zach Bethune ile yaptığı çalışma için atıf bilgisi sağlayın.

Bethune, Z., ve Korinek, A. (2021). Covid-19 Enfeksiyonu Dışsalıkları: Hayatlar ve Geçim Kaynakları Arasında Ticaret. Journal of Monetary Economics, 117, 1-13.

Makalenin başlığı doğru ancak yayın bilgileri hayal ürünü ve bu 200'den fazla atıf alan bir makale için. İsteğimin küçük değişiklikleri için (örneğin, yazarları alfabetik sıraya koymak), LLM böyle bir makale bulamadığı yanıtını verdi. Akademik literatüre yönelik talepler sıklıkla, kulağa yüzeysel olarak makul gelen uydurma makale başlıklarına ve dergilere yol açmaktadır. Buchanan ve Shapoval (2023) GPT-3.5'in her *JEL* kodu hakkında yazma kapasitesini sistematik olarak incelemiş ve üretilen atıfların yüzde 30'undan fazlasının mevcut olmadığını bulmuştur.

LLM'lerin aca- ile mücadele etmesinin nedeni

demik referanslar, eğitim verilerini kelimesi kelimesine tutmazlar, bunun yerine bir miktar bilgi kaybı içeren soyut bir temsilini tutarlar. Bu yaklaşım birçok genel kavram için işe yarasa da, hassasiyetin çok önemli olduğu akademik referanslar için yetersiz kalmaktadır.

Bununla birlikte, araştırmacıların akademik referansları bulmalarına yardımcı olmak için LLM'leri kullanan alternatif araçlar da mevcuttur. Örneğin, Bing veya Bard gibi web bağlantıları olan LLM'lerden internette belirli makaleleri aramaları ve bulgularından referans bilgileri oluşturmaları istenebilir. Ayrıca, <https://elicit.org> web sitesi akademik literatürde arama yapılmasına izin verir ve basit kullanıcı istemlerine dayalı olarak (gerçekten var olan) makalelerin bir listesini görüntüler. Ayrıca bir LLM tarafından oluşturulan kısa bir özet de içermektedir.

3.4.3 Referansları Biçimlendirme

Referans oluşturma konusunda gerçekten kötü olmalarına rağmen, basit LLM'ler bile (GPT3.5 gibi) doğru atıf bilgileri sağlandığında referansları istenen şekilde biçimlendirme konusunda oldukça yeteneklidir:

formatına dönüştürün: Anton Korinek, Large Language Models and Cognitive Automation for Economic Research, NBER Working Paper 30957, Şubat 2023.

```
@techreport{korinek2023large,  
title={Large Language Models and Cognitive  
Automation for Economic Research},  
author={Korinek, Anton},  
year={2023},  
institution={NBER},  
type={Working Paper},  
number={30957},  
month={February}  
}
```

Sohbet 18 (GPT-4) ctd: Chicago stiline göre biçimlendirin.

Korinek, Anton. 2023. "Ekonomik Araştırmalar için Büyük Dil Modelleri ve Bilişsel Otomasyon." NBER Çalışma Belgesi No. 30957, Şubat.

LLM'ler tüm bir kaynakçayı toplu formatta, örneğin APA'dan Chicago formatına veya bibtex formatına dönüştürebilir. LLM'nin eğitim verilerinde sıklıkla karşılaştığı çok yüksek atıflı çalışmalar için "bibtex reference for stiglitz weiss" gibi bir istem de işe yarar ve kullanıcıyı söz konusu

çalışmanın ayrıntılı atıf bilgilerini kopyalamak veya yazmak zorunda bırakmaz. Bununla birlikte, bilgilerin iki kez kontrol edilmesi tavsiye edilir.

3.4.4 Metin Çevirme

Jiao ve diğerleri (2023) GPT-3.5 ile başlayan LLM'lerin yüksek kaynak dillerde ticari çeviri ürünleriyle rekabetçi bir performans sergilediğini göstermiştir. Aşağıda bir örnek verilmiştir:

Sohbet 19 (GPT-4): Aşağıdaki cümleyi İspanyolca, Almanca ve Finceye çevirin: Genetik yapay zeka, özellikle de ChatGPT gibi büyük dil

modelleri (LLM'ler), birçok disiplindeki araştırmalarda devrim yaratma potansiyeline sahiptir.

İspanyolca: La IA generativa, en particular los modelos de lenguaje grandes (LLMs) like ChatGPT, tiene el potencial de revolucionar la investigación en muchas disciplinas.

Almanca: Generative KI, insbesondere grosse Sprachmodelle (LLMs) wie ChatGPT, hat das

Potansiyel, birçok disiplinde araştırma devrim yaratacak.

Fince: Generatiivinen tekoäly, erityisesti suuren kielen mallit (LLM) kuten ChatGPT, on potentiaalia mullistaa tutkimusta monilla aloilla.

Ön eğitim sırasında, LLM'ler herhangi bir dilin ötesinde soyut kavram temsilleri geliştirir ve bu da metni bir dilden diğerine kolayca eşleyebilmelerini sağlar. Daha az sayısallaştırılmış metnin ve daha az sayısallaştırılmış çevirinin mevcut olduğu daha düşük kaynaklı dillerde performans daha kötüdür.

3.4.5 Kavramların Açıklanması

LLM'ler bir öğretmen gibi hareket edebilir ve birçok yaygın ekonomik kavramı, yeni materyaller öğrenmeye çalışan öğrenciler ve hatta ana uzmanlık alanlarının ötesine geçen daha ileri düzeydeki araştırmacılar için yararlı olacak bir düzeyde açıklayabilir. İyi bir açıklama için en modern LLM'lerin istihdam edilmesi kritik önem taşımaktadır.

Sohbet 20 (GPT-4): *Enstrümantal değişkenler ne işe yarar?*

Araç değişkenler, istatistiksel analizde, özellikle de regresyon analizinde içsellik sorunlarını ele almak için kullanışlıdır. İçsellik, gözlemlenemeyen değişkenler, ölçüm hataları veya eşzamanlılık nedeniyle ortaya çıkabilir ve bu da yanlış ve tutarsız parametre tahminlerine yol açabilir.

Araç değişkenler, ilgilenilen değişkenin etkisini diğer karıştırıcı değişkenlerin etkilerinden izole etmek için kullanılır. Özellikle ilgili tüm değişkenlerin doğrudan gözlemlenmesinin mümkün olmadığı ya da ilgilenilen değişkenin hata terimiyle ilişkili olduğu durumlarda kullanışlıdır.

Buna ek olarak, araç değişkenler, tedavilerin rastgele atanmasının mümkün olmadığı gözlemsel verilerde nedensel ilişkileri tahmin etmek için kullanılabilir. Nedensel çıkarım için altın standart olarak kabul edilen randomize kontrollü bir çalışmanın koşullarını taklit etmenin bir yolunu sağlarlar.

LLM'leri bu şekilde kullanırken, daha fazla ayrıntıya giren takip soruları çok faydalı olabilir ve kullanıcının kişiselleştirilmiş özel ders almasını sağlayabilir. LLM'nin iyi cevap verdiği diğer kavram soruları arasında aşağıdakiler yer almaktadır:

- Görünmez el teoremi nedir ve neden geçerlidir?
- Monte Carlo simülasyonu nedir?
- Cournot dengesinin Bertrand dengesinden farkı nedir?
- Döviz kurunun belirlenmesine ilişkin önde gelen teoriler nelerdir?

Ancak bazen LLM'ler, özellikle de daha eski versiyonlar, cevap verirken halüsinasyon görürler. Aşağıda, 2022'nin sonlarında ChatGPT'nin orijinal ücretsiz sürümüne güç veren eski GPT-3.5 modeli text-davinci-003 kullanılarak bir örnek verilmiştir:

Sohbet 21 (GPT-3.5): *Refah ekonomisinin ikinci teoremi nedir?*

Refah Ekonomisinin İkinci Teoremi, herhangi bir rekabetçi dengenin Pareto etkin olduğunu belirtir. Bu, herhangi bir rekabetçi piyasa sonucunun, en az bir bireyi daha kötü duruma düşürmeden iyileştirilemeyeceği içi n mümkün olan en verimli sonuç olduğu anlamına gelir.

[Yazarın notu: bu açıkça ikinci değil birinci teoremdir!]

Refah ekonomisinin iki temel teoremi modern eko- nominin köşe taşlarıdır, ancak GPT-3.5 tarafından oluşturulan metin bu ikisini birbirine karıştırmaktadır. Her ne kadar LLM'ler birçok durumda net ve etkileyici yanıtlar verse de, bu durum temel kavramlarla ilgili temel gibi görünen sorularda bile çıktılarında her zaman güvenilemeyeceğini göstermektedir. LLM'ler eğitimci olarak oldukça yararlı ve güçlü olabileceğinden, ortaya çıkardıkları ikilemlerden biri, zaman zaman büyük bir güvenle yanlış yanıtlar üretmelerine rağmen öğrencilerin bunları kullanma olasılığıdır.

3.5 Kodlama

LLM'ler, özellikle çevrimiçi kaynakların çoğunun mevcut olduğu Python ve R dillerinde kodlama görevlerinde oldukça yeteneklidir. LLM'ler düz İngilizce (veya diğer doğal dillerde) talimatlara dayalı olarak kod parçacıkları yazabilir, düzenleyebilir, değiştirebilir, çevirebilir veya hata ayıklayabilir. Dahası, yeni kütüphaneler, fonksiyonlar ve hatta kullanıcının çok aşına olmadığı programlama dilleri kullanılırken, belirli işlem türleri için hangi kütüphanelerin ve fonksiyonların gerekli olduğunu veya belirli bir programlama dilinde hangi sözdizimsel yapıların kullanılacağını gösteren çıktıları hızlı bir şekilde oluşturarak eğitmen olarak hareket edebilirler. Bu, insan programcının LLM'ye başvurmasına ve onun tarafından üretilen birçok farklı parçacıktan kodu bir araya getirmesine olanak tanır.

LLM'lerin bu kadar yetkin olmasının nedenleri

kodlama aşağıdakileri içerir: Eğitim verilerine dahil edilen, örneğin GitHub'da bulunan geniş çevrimiçi kod depoları vardır. Bilgisayar kodunun sözdiziminin bu sistemler için öğrenilmesi nispeten kolay görünmektedir. Dahası, son teknoloji LLM'leri üreten YZ laboratuvarlarının kendileri de bu sistemlerin kod üretme yeteneklerinden yararlanmakta ve bu da onlara bu yetenekleri geliştirmek için güçlü teşvikler sağlamaktadır. Aslında kodlama, mevcut LLM'lerin en büyük üretkenlik kazanımlarına yol açtığı alanlardan biri olabilir: Peng ve arkadaşları (2023), OpenAI destekli GitHub Copilot kullanan programcıların ödevlerini ortalama yüzde 55,8 daha hızlı tamamladıkları ve bunun da yüzde 126'lık bir üretkenlik artışına denk geldiği kontrollü bir deney raporlamıştır.

ChatGPT Plus'ta, Gelişmiş Veriler

Analiz, Python'da kodlama görevlerini çok kolay hale getirmiştir, çünkü kod yazabilir, çalıştırabilir, ondan öğrenebilir, takip kodu yazabilir ve istenen sonuçları görüntüleyebilir. Ayrıca kullanıcıların kod

yüklemelerine, belirli değişiklikler istemelerine ve üzerinde çalışmaya devam etmek için tekrar indirmelerine olanak tanır. GitHub Copilot ve Tabnine

Visual Studio ve programcılar için diğer birçok entegre geliştirme ortamına sorunsuz entegrasyon sunar ve birçok farklı programlama ve biçimlendirme dili için çalışır.⁹ Ağustos 2023'te piyasaya sürülen Meta's Code LLaMA, açık kaynak temelinde güçlü bir kodlama yardımı sunmaktadır.

3.5.1 Kod Yazma

LLM'ler standart programlama görevleri, veri manipülasyonu ve tekrarlayan görevler de dahil olmak üzere çok çeşitli kodlar yazmak için çok kullanışlıdır. Bu makalenin amaçları doğrultusunda, ChatGPT Gelişmiş Veri Analizi'ni eko- nomikte klasik bir modeli simüle etmek için kullandım. Model zaman zaman kullanıcıdan onay ve geri bildirim istiyor. Aşağıda yaptığım alışveriş raporlanmaktadır:

Sohbet 22 (ChatGPT İleri Veri Analizi):
Solow büyüme modelini simüle edin ve sonuçları dört grafikte çizin.

Pekala, Solow büyüme modelini simüle edelim. İşleri basitleştirmek için, teknolojik ilerlemenin olmadığı Solow-Swan modelini ele alalım. Model, bir ekonomide sermaye (K), işgücü (L) ve çıktının (Y) gelişimini açıklamaktadır. Zaman içinde sermaye stokundaki değişim için temel denklem şu şekilde verilir:

$$\frac{dK}{dt} = sY - \delta K$$

Nerede?

- s tasarruf oranıdır
- Y , genellikle Cobb-Douglas tipi bir üretim fonksiyonu tarafından verilen toplam çıktıdır. tip: $Y = K^\alpha L^{1-\alpha}$,
- δ amortisman oranıdır.

Model için varsayımlar:

⁹ Bu makaleyi yazarken LaTeX kodu yazmak için ChatGPT Plus'ı kullandım. Makale boyunca örnek istemlerin etrafındaki kutuların görsel olarak çekici olacak ve sayfalar arasında bölünecek şekilde nasıl tanımlanacağını sordum ve bölüm 3'ün içindekiler tablosunu temel alarak LaTeX'te tablo 1'i üretmesini istedim. Bu bölüm için çoğaltma kodu üzerinde çalışırken GitHub Copilot'u kullandım.

- İşgücü sabit bir n oranında büyür, yani,
 $L(t) = L(0) e^{nt}$,
- Tasarruf oranı s , amortisman oranı δ ve işgücü büyüme oranı n sabittir.

Başlamak için bu parametrelerin değerlerini belirtmemiz gerekir. Örneğin, aşağıdakileri kullanabiliriz:

- $s = 0.4$,
- $\delta = 0.1$,
- $n = 0.02$,
- $\alpha = 0.33$,
- başlangıç sermaye stoku $K(0) = 1$,
- başlangıçtaki emek $L(0) = 1$.

Modeli simüle etmek için, diferansiyel denklemleri entegre etmek için Euler yöntemini kullanacağız. Bu yöntem, fonksiyonun gelecekteki bir zamandaki değerine, mevcut zamandaki değer artı fonksiyonun mevcut zamandaki türevinin zaman adımıyla çarpımı olarak yaklaşır.

Ardından, aşağıdaki dört grafiği çizeceğiz:

1. Zaman içinde sermaye,
2. Zaman içinde emek,
3. Zaman içinde çıktı,
4. Zaman içinde işçi başına

düşen sermaye. Bu kulağa hoş

geniyor mu?

Programın oluşturulmasına devam etmek için kısa bir "Devam et" mesajı verdim:

Sohbet 22 (ChatGPT İleri Veri Analizi) ctd:
Devam edin.

```
import numpy as np
import matplotlib.pyplot as plt
```

```
# Model
parametreleri s = 0,4
delta = 0.1
n = 0.02
alfa = 0,33
K0 = 1
L0 = 1
T = 100 # Toplam simülasyon süresi
```

```
# Zaman
izgarası dt
= 0,01
times = np.arange(0, T, dt)
```

```
# Zaman içinde K, L, Y değerlerini saklamak
için dizileri başlatın
K = np.zeros(len(times))
```

```
L = np.zeros(len(times))
Y = np.zeros(len(times))
```

```
# Başlangıç
değerleri K[0] =
K0
L[0] = L0
Y[0] = K0**alfa * L0**(1-alfa)
```

```
# Euler yöntemi
for t in range(len(times)-1):
    dK = s*Y[t]-delta*K[t]
    dL = n*L[t]
    K[t+1] = K[t] + dK*dt
    L[t+1] = L[t] + dL*dt
    Y[t+1] = K[t+1]**alfa * L[t+1]**(1-alfa)
```

```
# İşçi başına sermaye
K_L = K / L
```

```
# Çizimler oluşturun
fig, axs = plt.subplots(2, 2, figsize=(14, 10))
```

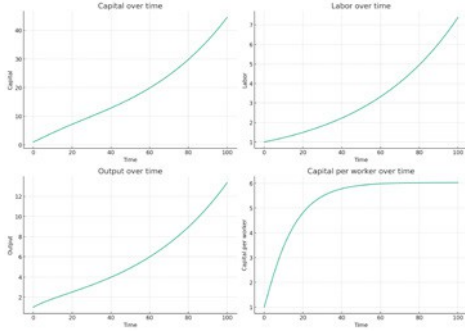
```
# Sermayeyi zaman
içinde çizin axs[0,
0].plot(times, K)
axs[0, 0].set_title('Zaman içinde sermaye')
axs[0, 0].set_xlabel('Zaman')
axs[0, 0].set_ylabel('Capital')
# Zaman içinde emek grafiği
axs[0, 1].plot(times, L)
axs[0, 1].set_title('Zaman içinde emek')
axs[0, 1].set_xlabel('Zaman')
axs[0, 1].set_ylabel('İşçilik')
```

```
# Çıktıyı zaman içinde
çizin axs[1,
0].plot(times, Y)
axs[1, 0].set_title('Zaman içinde çıktı')
axs[1, 0].set_xlabel('Zaman')
axs[1, 0].set_ylabel('Çıktı')
```

```
# Zaman içinde işçi başına sermaye
grafiği axs[1, 1].plot(times, K_L)
axs[1, 1].set_title('Zaman içinde işçi başına
sermaye')
axs[1, 1].set_xlabel('Zaman')
axs[1, 1].set_ylabel('İşçi başına sermaye')
```

```
# Düzeni sıkı hale
getirin
plt.tight_layout()

plt.show()
```



İşte Solow büyüme modeli için dört grafik:

- 1. Zaman içinde sermaye:** Sermaye stoku zaman içinde artar ve sabit bir duruma yaklaşır. Bu durağan durum, yatırımın (tasarrufun) amortismanına eşit olduğu noktadır, yani sermaye stoku artık değişmez. [Yazarın notu: her ne kadar sayısal simülasyon Doğru, açıklamanın bu noktası açıkça yanlış].
- 2. Zaman içinde işgücü:** İşgücü, Solow modelindeki sabit nüfus artışı varsayımını yansıtacak şekilde sabit bir oranda üstel olarak büyür.
- 3. Zaman içinde çıktı:** Çıktı da zaman içinde artar ve durağan bir duruma yaklaşır. Bunun nedeni, çıktının her ikisi de artan sermaye ve emeğin bir fonksiyonu olmasıdır. Ancak, sermayenin azalan getirisi nedeniyle, çıktının büyümesi zaman içinde yavaşlar.
- 4. Zaman içinde işçi başına sermaye:** İşçi başına sermaye de zaman içinde artar ve durağan bir duruma yaklaşır. Bu, Solow modelinde ekonominin dengeli büyümeye ulaştığı noktadır. Durağan durumda, işçi başına üretim ve işçi başına sermaye aynı oranda büyür, bu da tekno-mantıksal ilerleme oranıdır. Modelin teknolojik ilerlemenin olmadığı bu basitleştirilmiş versiyonunda bu büyüme oranı sıfırdır, dolayısıyla işçi başına çıktı ve işçi başına sermaye durağan durumda sabittir.

Bu sonuçlar Solow büyüme modelinin temel öngörülerini ortaya koymaktadır: uzun vadede bir ekonomi sabit bir sermaye düzeyine (hem toplam hem de işçi başına) yaklaşır ve bu sabit düzey tasarruf oranı, amortisman oranı ve nüfus artış oranı tarafından belirlenir. Değişimler

Bu parametrelerdeki değişiklikler daha yüksek veya daha düşük bir durağan duruma yol açabilir, ancak uzun dönemli eko- nomik büyümeyi sürdüremezler. Modele göre, uzun dönemli ekonomik büyüme sadece teknolojik ilerleme ile sürdürülebilir.

Model simülasyon görevini istendiği gibi yerine getirmiştir. Solow modelinin detaylı bir tanımıyla başladı, parametre varsayımlarını tanıttı, kullanıcı onayı istedi ve doğru Python kodunu yazıp çalıştırdı. Ancak, sonuçları yorumlamaya da gönüllü olmuş ve zaman içinde sermayenin izleyeceği yolu yanlış bir şekilde "durağan duruma yaklaşmak" olarak tanımlamıştır. Pozitif nüfus artışı varsayıldığı için bu açıkça yanlıştır ve şeklin sol üst panelinde gösterilen simülasyonla da çelişmektedir. (Gelişmiş Veri Analizi şekilleri oluşturabilir ancak gerçekte onları "göremez", bu nedenle sadece şeklin makul bir açıklaması gibi *görünen* şeyi oluşturmuş ve böylece bir hata ortaya çıkarmıştır). Dahası, sonuç paragrafında LLM şöyle yazmıştır: "Bu sonuçlar Solow büyüme modelinin temel öngörülerini göstermektedir: uzun vadede, bir ekonomi durağan bir sermaye seviyesine (hem toplam hem de işçi başına) yaklaşır" ki bu -pozitif nüfus artışı altında- açıkça imkansızdır.

Ayrıca Ramsey'i simüle etmeye çalıştım modeli. LLM, verilen başlangıç koşulları için tüketim ve sermaye yolunu simüle etmek için doğru bir şekilde kod üretti, ancak tüketim için doğru başlangıç noktasını çözme yi başaramadı, hatta başlangıç tüketiminin eyer yolu kararlı bir sistemin kararlı kolunda yer alması gerektiğini hatırlatmama rağmen.

Bu, LLM kodlama yeteneklerinin yazım sırasındaki durumunu temsil etmektedir - LLM'ler programlamanın sıradan kısımlarının çoğunu otomatikleştirmek için yararlı yardımcılardır, ancak daha zor kısımlar için insan gözetimi gerektirirler. Ayrıca LLM'ler, kullanıcıların belirli kullanım durumları için ihtiyaç duydukları paketleri ve işlevleri bulmalarına yardımcı olan güçlü eğitmenler olarak da hareket edebilir.

Bir Kodlama Projesi için Planın Ana Hatlarının Çizilmesi [Haziran 2024'te Tümüyle Yeni] -Cutting-edge LLM'ler, özellikle Claude 3, ayrıca

kodlama projelerine nasıl yaklaşacaklarına dair bir plan yapmaları istenir. Düşünce zinciri teşvikinin ruhuna uygun olarak, önce bir programlama görevinin nasıl kodlanacağına dair bir plan yapmaları ve ardından bu planı uygulamaları istendiğinde, sıklıkla daha iyi sonuçlar elde ederler.

3.5.2 Kodun Açıklanması

Önceki örnekte İngilizce talimatlara göre kod yazılırken, LLM'ler bunun tersini de yapabilir ve bir öğretmene benzer şekilde belirli bir kod parçasının ne yaptığını düz İngilizce olarak açıklayabilir. Aşağıdaki örnek bu kullanım durumunu göstermektedir:

Sohbet 23 (GPT-4): Aşağıdaki fonksiyon ne işe yarar?

```
def regression(X, y):
    X = np.hstack((np.ones((X.shape[0], 1)),
    X))
    XtXinv = np.linalg.inv(np.dot(X.T, X))
    beta1 = np.dot(np.dot(XtXinv, X.T), y)
    artıklar = y - np.dot(X, beta1)
    V = np.diag(residuals**2)
    Vinv = np.linalg.inv(V)
    XtVinvXinv = np.linalg.inv(np.dot(X.T,
    np.dot(Vinv, X)))
    beta2 = np.dot(np.dot(XtVinvXinv, X.T),
    np.dot(Vinv, y))
    beta1, beta2 döndür
```

Bu fonksiyon iki aşamalı bir genelleştirilmiş en küçük kareler regresyonu gerçekleştirir.

İlk adımda, modelin parametrelerini (beta1) tahmin etmek için sıradan bir en küçük kareler regresyonu gerçekleştirir.

İkinci adımda, hataların varyansını tahmin etmek için ilk adımdaki kalıntıları kullanır ve ardından hataların heteroskedastikliğini (sabit olmayan varyans) dikkate alarak modelin parametrelerini (beta2) tahmin etmek için ağırlıklı bir en küçük kareler regresyonu gerçekleştirir.

Fonksiyon her iki adımdan da tahmini parametreleri döndürür.

Bu işlevsellik özellikle kodlarını iyi belgelendirmemiş kaynakları kullanırken kullanışlıdır.

3.5.3 Kod Çevirme

LLM'ler ayrıca, örneğin bir projeyi başka bir platforma taşımak veya StackExchange gibi çevrimiçi bir kodlama forumunda bulunan ve yanlış dilde olan kod parçacıklarını entegre etmek için bir programlama dilindeki kodu başka bir dile dönüştürebilir.

Devir öncesi görüşmeyi temel alan aşağıdaki istem bir örnek teşkil etmektedir:

Sohbet 23 (GPT-4) ctd.: Fonksiyonu Matlab'a çevirebilir misiniz?

Elbette, işte fonksiyon için eşdeğer MATLAB kodu:

```
fonksiyon [beta1, beta2] = regresyon(X, y)
    X = [ones(size(X, 1), 1), X];
    XtXinv = inv(X' * X);
    beta1 = XtXinv * X' * y;
    residuals = y - X * beta1;
    V = diag(residuals.^2);
    Vinv = inv(V);
    XtVinvXinv = inv(X' * (Vinv * X));
    beta2 = XtVinvXinv * X' * (Vinv * y);
son
```

MATLAB'de matris çarpma operatörünün 'np.dot()' değil '*' olduğunu lütfen unutmayın. Ayrıca, eleman-bilge güç için operatör '**' değil '^' dir.

Mevcut LLM'ler kısa kod parçalarını yaygın programlama dilleri arasında çevirmede oldukça güvenilirdir. Daha uzun diziler için hala insan yardımı gerekmektedir.

3.5.4 Kodda Hata Ayıklama

Programlamanın en sinir bozucu kısımlarından biri kodda hata ayıklamaktır. LLM'ler bu alanda da faydalı yetenekler sergiler. Yalnızca hata mesajlarının yorumlanmasına yardımcı olmakla kalmaz, aynı zamanda kodu doğrulayabilirler.

İlk örneğimde, Python'dan tipik bir hata mesajını yapıştırdım

kullanıcı bir kütüphaneyi içe aktarmayı unuttuğunda, başka herhangi bir talimat olmaksızın elde edilir:

Sohbet 24 (GPT-4): *Geri izleme (en son çağrı): File "C:/.../test.py", line 20, in <module> X*
np.array([[1, 2], [3, 4], [5, 6], [7, 8]]) NameError: 'np' adı tanımlanmamış.

Hata mesajı size 'np' adının tanımlanmadığını söylüyor. Bu genellikle henüz içe aktarmadığınız bir kütüphaneyi kullanmaya çalıştığınızda olur.

Sizin durumunuzda, 'np' numpy kütüphanesi için yaygın bir takma addır. Kodunuzun başında numpy'yi içe aktararak hatayı düzeltebilirsiniz. İşte bunu nasıl yapacağınız:

```
"python
import numpy as np"
```

Bu satırı ekledikten sonra, kodunuz 'np'nin ne olduğunu bilecek ve beklediği gibi çalışacaktır.

LLM, komut istemimi hatayı ayıklamak için bir yardım talebi olarak doğru bir şekilde yorumladı. Hatanın kaynağını tanıdı, açıkladı ve düzeltmek için gereken kodu sağladı. İki örnek önceki Python fonksiyonu üzerine inşa edilen ancak bir hata eklenen bir sonraki örnek, LLM'lerin koddaki yazım hataları gibi hataları yakalamaya nasıl yardımcı olabileceğini göstermektedir:

Sohbet 25 (GPT-4): *Aşağıdaki fonksiyonda hata ayıklaması yapabilir misiniz? [Vinv değişkeninin tanımını V_inv olarak değiştirerek yapılan bir hata ile önceki örnekten alınan kod]*

Fonksiyon bir tür ağırlıklı en küçük kareler regresyonu uyguluyor gibi görünüyor, ancak bununla ilgili birkaç sorun var: 1. İşlev numpy işlevlerini ve yöntemlerini kullanıyor ancak numpy içe aktarılmamış. Kodunuzun başında numpy'yi içe aktarmanız gerekir. 2. 'Vinv' değişkeni tanımlanmamış. 'V_inv' olması gerekiyor gibi görünüyor. İşte düzeltilmiş fonksiyon:

[Doğru kod üretildi.]

İlk nokta, numpy paketi olmadığı sürece kodun çalışmadığını gözlemler

ithal edilmiştir (önceki örnekte olduğu gibi); ikinci nokta, benim ortaya koyduğum hatayı doğru bir şekilde yakalamıştır.

Daha genel olarak, LLM'ler bazen indekslerin ne zaman karıştırıldığını tespit edebilir ve bir programın arızalanma yollarını tanımladığında tavsiye verebilir. Kodun altında yatan algoritmalarındaki üst düzey yanlışlıkları tespit etme konusunda daha az yeteneklidirler - bunun için genellikle insan hata ayıklaması gerekir.

3.6 Veri Analizi

LLM'ler veri kaynaklarını bulmaya, verileri for- mat etmeye, metinden veri çıkarmaya, metni sınıflandırmaya ve puanlamaya, şekiller oluşturmaya, duyguları çıkarmaya ve hatta insan test deneklerini simüle etmeye yardımcı olabilir. Bu yeteneklerin çoğuna yalnızca aşağıdaki gösterimlerde gösterildiği gibi bir web arayüzü aracılığıyla değil, aynı zamanda büyük miktarda verinin formatlanmasına, çıkarılmasına, sınıflandırılmasına vb. olanak tanıyan bir API aracılığıyla da erişilebilir. İşlemler, her bir istek için token sınırı içinde kalmak üzere toplu olarak da gerçekleştirilebilir. Ayrıca, kodlama ile ilgili bölüme dayanarak, LLM'lerin kendi API'lerine erişmek için gerekli bilgisayar kodunu yazabileceklerini söylemeye gerek yok - örneğin

"GPT-4'ten [herhangi bir veri çıkarma veya manipülasyon görevi] yapmasını istemek için Python kodu yazın."

Veri analizi görevlerini gerçekleştirirken toplu olarak, maliyet önemli bir husustur. Son teknoloji bir LLM'ye tek bir sorgu sadece bir sentin kesirlerine mal olsa da, binlerce veya milyonlarca sorgu gerçekleştirmenin maliyeti hızla artar. Aşağıda açıklanan görevlerin birçoğu için daha küçük ve daha ucuz yöntemler mevcuttur. Bu durumlarda, en gelişmiş LLM'nin kullanılması tavsiye edilmez.

3.6.1 Veri Kaynaklarını Bulma [Haziran 2024 Sürümünde Tüm Yenilikler]

LLM'ler veri analizinin ilk adımından itibaren, yani uygun veri kaynaklarının belirlenmesi ve konumlandırılmasında faydalı olabilir. Aşağıdaki örnek bu rolü göstermektedir

Claude 3'ün kullanıcılara hangi veri kaynaklarını dikkate almaları gerektiği konusunda tavsiyede bulunma konusunda oynayabileceği rol:

Sohbet 26 (Claude 3): Yapay zekanın işgücü piyasası üzerindeki etkilerini analiz etmek için hangi veriler yararlı olur ve bunları nerede bulabilirim?

Yapay zekanın işgücü piyasası üzerindeki etkilerini analiz etmek için çeşitli kaynaklardan elde edilen verilere ihtiyacınız olacaktır. İşte bazı temel veri noktaları ve potansiyel kaynaklar:

1. Sektör ve meslek bazında istihdam verileri
 - ABD'de Çalışma İstatistikleri Bürosu (BLS)
 - Avrupa ülkeleri için Eurostat
 - Diğer ülkelerin ulusal istatistik kurumları
2. Sektörlere göre yapay zeka teknolojilerinin benimsenme oranları
 - Gartner, IDC veya McKinsey gibi şirketlerin pazar araştırma raporları
 - Akademik çalışmalar ve anketler
 - Şirket raporları ve basın bültenleri

[3. ila 9. noktalar yer nedeniyle atlanmıştır.] Bu çeşitli kaynaklardan elde edilen verilerin birleştirilmesi

yapay zekanın sektörler, meslekler ve demografik gruplar arasında işgücü piyasasını nasıl etkilediğine dair kapsamlı bir resim sunmaktadır. Veri mevcudiyetinin ve kalitesinin ülkeler ve bölgeler arasında farklılık gösterebileceğini u n u t m a m a k önemlidir.

Claude tarafından sağlanan liste bir çeşit çamaşır listesi ve genel olmasına rağmen, araştırmacıya yeni veri kaynakları hakkında fikir verebilecek iyi bir beyin fırtınası temeli olarak hizmet etmektedir. Ayrıca Claude, veri kaynaklarının en iyi nasıl indirileceği ve en iyi nasıl temizleneceği konusunda takip tavsiyeleri sunmaktan da mutluluk duymaktadır.

3.6.2 Figürlerin Oluşturulması

ChatGPT'nin ekonomistler için en kullanışlı işlevlerinden biri, çok yönlü şekiller ve grafikler oluşturmak için GPT-4'ün kodlama yeteneklerini kullanan Gelişmiş Veri Analizi'dir.

Aşağıdaki örnekte, üç büyük teknoloji şirketinin borsa fiyatlarını içeren bir dosya yükledim ve ChatGPT Gelişmiş Veri Analizi'ne hisse senedi performansını gösteren bir grafik oluşturması talimatını verdim

karşılık gelen betalarla etiketlenmiş ve portföy ağırlıklarını gösteren başka bir grafik.¹⁰

Sohbet 27 (ChatGPT İleri Veri Analizi): *Bu menkul kıymetlerin performansını 100'den başlamak üzere normalize edilmiş olarak gösteren bir grafik gösterebilir misiniz?* [Yüklenen stock_data.csv]

[Kod oluşturuldu ve çalıştırıldı.]

Veri kümesi, Google (GOOG), Meta (META), Microsoft (MSFT) ve S&P 500 endeksi (^GSPC) menkul kıymetlerinin belirli bir zaman dilimindeki kapanış fiyatlarını içerir.

Fiyatları 100'den başlayacak şekilde normalleştirilim ve bu menkul kıymetlerin ortaya çıkan performansını çizelim. Bu normalleştirme, menkul kıymetlerin performansını ilk fiyatlarına göre karşılaştırmamıza olanak sağlayacaktır. Verileri normalleştirmek için, tüm fiyatları ilk fiyatlarına böleceğiz ve ardından 100 ile çarpacağız.

[Kod oluşturuldu ve çalıştırıldı.] [Kod oluşturuldu ve çalıştırıldı.]

İşte Google (GOOG), Meta (META), Microsoft (MSFT) ve S&P 500 endeksi (^GSPC) menkul kıymetlerinin 100'den başlamak üzere normalleştirilmiş performansını gösteren grafik. X eksenini zamanı (tarihleri), y eksenini ise normalleştirilmiş fiyatları temsil etmektedir.

Grafikten, bu menkul kıymetlerin her birinin zaman içinde başlangıç noktasına göre nasıl performans gösterdiğini görebiliriz. Bu tür bir karşılaştırma, farklı menkul kıymetlerin veya endekslerin göreceli performansını değerlendirirken yardımcı olabilir.

Sohbet 27 (ChatGPT İleri Veri Analizi) ctd: *Her bir hisse senedi için beta değerini hesaplayın ve göstergeye ekleyin. Renkleri şu şekilde değiştirin*



¹⁰ Temel borsa verilerini derlemek için ChatGPT'den, makalenin çevrimiçi ekinde açıklandığı gibi, verileri indirmek için bir komut dosyası yazmasını istedim.

ayırt edilmelerini kolaylaştırın. "Üç Büyük Teknoloji Şirketinin On Yıllık Performansı" başlığını kullanın.

Bir hisse senedinin betası, hisse senedinin genel piyasaya göre oynaklığının bir ölçüsüdür. Tüm piyasanın betası 1,0'dır ve bireysel hisse senetleri piyasadan ne kadar saptıklarına göre sıralanır.

Zaman içinde piyasadan daha fazla dalgalanan bir hisse senedinin betası 1,0'ın üzerindedir. Bir hisse senedi piyasadan daha az hareket ediyorsa, hisse senedinin betası 1,0'ın altındadır. Yüksek betalı hisse senetlerinin daha riskli olduğu ancak daha yüksek getiri potansiyeli sağladığı varsayılır; düşük betalı hisse senetleri daha az risk taşır ancak aynı zamanda daha düşük getiri sağlar.

Verilen dönem boyunca bu hisse senetlerinin her biri için betayı hesaplayalım. Betayı şu formülü kullanarak hesaplayabiliriz:

$$\beta = \frac{Cov(r_E, r)_M}{Var(r)_M}$$

Nerede ?

- r_E hisse senedinin getirisidir
- r_M piyasanın getirisidir
- $Cov(r_E, r_M)$ hisse senedi ve piyasa getirileri arasındaki kovaryansdır
- $Var(r_M)$ piyasa getirilerinin varyansdır

Piyasa getirisi için S&P 500 endeksini (^GSPC) kullanacağız. Her bir hisse senedi ve piyasa için günlük getirileri hesaplayacağız, ardından her bir hisse senedi için betayı hesaplayacağız.

Bundan sonra, grafiği yeni renklerle ve hesaplanan betalar göstergede yer alacak şekilde yeniden oluşturacağım. Renkler iyi kontrast ve net görünürlük sağlayacak şekilde seçilecektir.

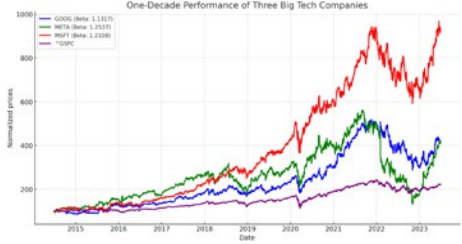
[Kod oluşturuldu ve çalıştırıldı]

İşte hisse senetleri için beta değerleri:

- Google (GOOG): 1.1317
- Meta (META): 1.2537
- Microsoft (MSFT): 1.2108

Bu değerler, beta değerleri 1'den büyük olduğu için her üç hisse senedinin de genellikle piyasadan (S&P 500) daha oynak olduğunu gösterir. Şimdi grafiği yeni başlık, sütunlar ve açıklamadaki beta değerleriyle güncelleyelim.

[Kod oluşturuldu ve çalıştırıldı.]



İşte yeni başlığı, renkleri ve beta değerleri ile güncellenmiş grafik.

Grafik, üç büyük teknoloji şirketinin on yıllık performansını temsil etmektedir: Google (GOOG), Meta (META) ve Microsoft (MSFT) ve S&P 500 endeksi (^GSPC), hepsi 100'den başlamak üzere normalize edilmiştir. Hisse senetlerinin genel piyasaya göre volatilitelerini ölçen beta değerleri göstergede yer almaktadır.

Sohbet 27 (ChatGPT Gelişmiş Veri Analizi-

sis) ctd: Başlığın, göstergenin, eksen etiketlerinin ve tiklerin boyutunu iki katına çıkarın. Göstergedeki ""^GSPC"" ifadesini "StP 500" olarak etiketleyin.

[Kod oluşturuldu ve çalıştırıldı.]



İşte talep edilen değişikliklerle güncellenmiş grafik:

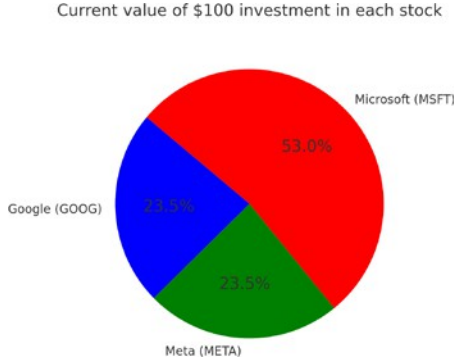
- Başlık, gösterge, eksen etiketleri ve tiklerin boyutu iki katına çıkarılmıştır.
- Göstergedeki ""^GSPC"" etiketi "S&P 500" olarak değiştirilmiştir.

Bu grafik üç büyük teknoloji şirketinin on yıllık performansını göstermektedir: Google (GOOG), Meta (META) ve Microsoft (MSFT) ile S&P 500 endeksi, hepsi 100'den başlayacak şekilde normalize edilmiştir. Hisse senetlerinin beta değerleri göstergede yer almaktadır.

Sohbet 27 (ChatGPT İleri Veri Analizi) ctd:

Başlangıçta üç hisse senedinin her birine 100 dolar koyduğumu varsayalım. Şu anda her bir hisse senedinde ne kadar param olacağını gösteren bir pasta grafik gösterebilir misiniz?

[Kod oluşturuldu ve çalıştırıldı]



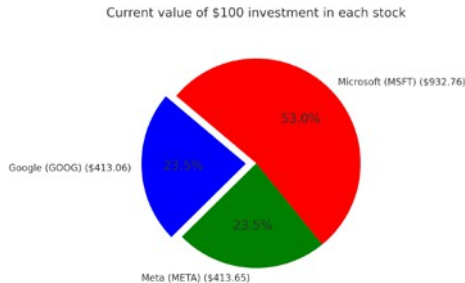
Burada, verilerde temsil edilen dönemin başında üç büyük teknoloji hissesinin (Google, Meta ve Microsoft) her birine yapılan 100 dolarlık bir yatırımın mevcut değerini gösteren bir pasta grafik yer almaktadır.

Pasta grafik, her biri hisse senetlerinden birini temsil eden üç bölüme ayrılmıştır. Her bir bölümün büyüklüğü, o hisse senedine yapılan ilk 100 \$'lık yatırımın mevcut değerine karşılık gelmekte ve üç yatırımın toplam değerinin yüzdesi olarak ifade edilmektedir. Netlik sağlamak amacıyla yüzdeler de grafik üzerinde etiketlenmiştir.

Sohbet 27 (ChatGPT Gelişmiş Veri Analizi)

ctd: Her pozisyonu yüzdeye ek olarak mevcut dolar değeriyle etiketleyebilir misiniz? Parçalardan birinin pastadan çıkmasını sağlayabilir misiniz?

[Kod oluşturuldu ve çalıştırıldı.]



İşte güncellenmiş pasta grafik. Artık her dilim hem mevcut dolar değerini hem de temsil ettiği toplam portföyün yüzdesini içeriyor. Google hissesini temsil eden ilk dilim, vurgulanması için "patlatılmış" veya pastadan kaydırılmıştır. Bu görsel efekt, pasta grafiğin belirli bir parçasını vurgulamak için yararlı olabilir. Bu durumda, Google'ın 100 \$'lık ilk yatırıma dayalı olarak üç hisse senedi arasında en yüksek değere sahip olduğu vurgulanmaktadır.

[Yazarın notu: rakam doğrudur, ancak ChatGPT'nin Google'ın en yüksek est değeri yanlışdır - bu, Gelişmiş Veri Analizinin rakamları "göremeyeceğini" hatırlatır yaratır.]

Python'un neredeyse sonsuz yeteneklerine güvenen ChatGPT Gelişmiş Veri Analizi, diğer birçok şekil türünü görüntüleyebilir.

3.6.3 Metinden Veri Çıkarma

LLM'lerin veri analizi için güçlü bir yeteneği, yazılı metinden veri çıkarmaktır. Aşağıdaki örnek basit bir örnektir.

Sohbet 28 (GPT-4): Mark ekonomiden A ve matematikten B+ aldı. Sally hem ekonomi hem de matematikten A aldı. Frank ekonomiden B ve matematikten C aldı.

Aşağıdaki gibi yeniden biçimlendirin: İsim ve ekonomi sınıfı ve matematik sınıfı.

Mark ve A ve B+
Sally ve A- ve A-
Frank ve B ve C

LLM, Sally için "both in econ and math" ifadesinin anlamını doğru bir şekilde işler ve notları buna göre çıkarır. Metinden veri çıkarmanın diğer uygulamaları arasında sayısal veriler, örneğin haber makalelerinden hisse senedi fiyatları veya ilaç veritabanlarından dozaj bilgileri yer almaktadır. LLM'ye "Metinden ne tür sayılar çıkarabileceğinize dair örnekler verebilir misiniz?" diye sorduğumda şu 10 tür cevabı verdi: telefon numaraları, posta kodları, sosyal güvenlik numaraları, kredi kartı numaraları, banka hesap numaraları, tarihler, saatler, fiyatlar, yüzdeler, ölçümler (uzunluk,

ağırlık, vb.) Süreç, API erişimi kullanılarak büyük miktarda veri için otomatikleştirilebilir ve genellikle GPT-4'ten daha küçük ve daha ucuz modellerle gerçekleştirilebilir. Dunn ve diğerleri (2022) bilimsel metinlerden yapılandırılmış bilgi çıkarma görevleri için LLM'lerin nasıl kullanılacağını göstermektedir. Bu, örneğin ekonomi tarihi araştırmalarında varlık tanıma için ekonomide de kullanılabilir.

belirli bir görevin

3.6.4 Verileri Yeniden Biçimlendirme

LLM'ler verileri istenen formata çevirme konusunda da aynı derecede çok yönlüdür. Önceki örnekten yola çıkarak, LLM'den verileri önce virgülle ayrılmış değerler (CSV) formatına, ardından da LaTeX tablosuna dönüştürmesini istedim:

Sohbet 28 (GPT-4) ctd: *CSV formatında yeniden biçimlendirin.*

İsim, ekonomi notu,
matematik notu Mark, A, B+
Sally, A-, A-
Frank, B, C

Sohbet 28 (GPT-4) ctd.: *LaTeX tablosu olarak yazın.*

| İsim | ekonomi notu | matematik notu |
|-------|--------------|----------------|
| Mark | A | B+ |
| Sally | A- | A- |
| Frank | B | C |

GPT-4 hem bu hem de diğer benzer görevlerde üstün performans göstermiştir.

3.6.5 Metni Sınıflandırma ve Puanlama

Sosyal bilim araştırmaları, metni veri olarak temsil etmek için sıklıkla istatistiksel teknikler kullanır (Gentzkow, Kelly ve Taddy 2019). Modern LLM'ler bunun için geleneksel tekniklerin ötesine geçebilir çünkü beslendikleri cümlelerin anlamını işleme konusunda giderek daha yetenekli hale gelmektedirler.

Aşağıdaki örnek GPT-4'ten ABD Çalışma Bakanlığı'nın Meslekler Listesinde listelenen

Bilgi Ağı (O*NET) veritabanının otomatikleştirilmesinin kolay veya zor olması ve sınıflandırma.¹¹ Wei, Wang ve diğerleri (2022) tarafından önerilen düşünce zinciri yönlendirme ilkesini takip eden yönlendirme, gerçek sınıflandırmayı gerçekleştirmeden önce LLM'yi yanıt hakkında muhakeme yapmaya teşvik etmek için önce gerekçeyi sorar. Bu, bir öğrenciden bir soruya yanıt vermeden önce düşünmesini istemeye benzer.

Sohbet 29 (GPT-4): İşte O*Net'ten beş görev:

1. Doğruluk ve verimliliği sağlamak için operasyonel kayıtları ve raporları izleyin.
2. Kayıt tutma için politika ve prosedürler geliştirin ve uygulayın.
3. Çalışan kayıtlarını hazırlayın ve muhafaza edin.
4. Dosyalama, envanter, posta ve veritabanı sistemlerini korumak ve güncellemek.
5. Raporları, çizelgeleri ve grafikleri derlemek ve hazırlamak.

Her bir görevin otomatikleştirilmesinin ne kadar kolay olduğuna dair 5-10 kelimelik bir açıklama sağlayın ve otomatikleştirilmesi "kolay" veya "zor" olarak sınıflandırın. Çıktınızı aşağıdaki sütunlara sahip bir LaTeX tablosu şeklinde sağlayın: no; görev; gerekçe; sınıflandırma

| Hayır | Görev | Gerekçe | Sonuç |
|-------|---|--|-------|
| 1 | Doğruluk ve verimliliği sağlamak için operasyonel kayıtları ve raporları izleyin. | Bağlaman anlaşılmasını gerektirir ve eleştirel düşünme. | Sert |
| 2 | Kayıt tutma için politika ve prosedürler geliştirin ve uygulayın. | İnsan muhakemesi ve karar verme gerektirir. | Sert |
| 3 | Çalışan kayıtlarını hazırlayın ve muhafaza edin. | Çoğunlukla veri girişi ve geri alma, kolayca otomatikleştirilebilir. | Kolay |

¹¹ Eloundou ve diğerleri (2023), LLM'lerin işgücü piyasasına etkisini sistematik olarak tahmin etmek için GPT-4'ü kullanmaktadır.

| | | | |
|---|---|--|-------|
| 4 | Dosyalama, envanter, posta ve veritabanı sistemlerini korumak ve güncellemek. | Çoğunlukla veri girişi ve geri alma, kolayca otomatikleştirilebilir. | Kolay |
| 5 | Raporları, çizelgeleri ve grafikleri derlemek ve hazırlamak. | Veri işleme ve görselleştirme otomatikleştirilebilir. | Kolay |

Sonuçlar makul olmakla birlikte, tamamen sağlam olmadıklarını vurgulamalıyım. Görünürdeki koşullara bağlı olarak farklı kararlar verebilen bir insan değerlendirici gibi, modelin esasen aynı olan soruya verdiği yanıt, istemdeki ifade değiştirildiğinde ve hatta bazı durumlarda listelenen görevlerin sırası değiştirildiğinde bile değişmektedir.

3.6.6 Duyguları Çıkarma

LLM'ler ayrıca metinden duyguları da çıkarabilir. Örneğin tweetleri "olumlu" veya "olumsuz" olarak sınıflandırabilirler. Benzer şekilde, aşağıdaki örnekte gösterildiği gibi, ABD faiz oranlarını belirleyen Federal Açık Piyasa Komitesi (FOMC) tarafından yapılan açıklamaları kategorize edebilirler:

Sohbet 30 (GPT-4): *Aralık 2022 tarihli aşağıdaki FOMC açıklaması şahin mi yoksa güvercin mi? Son göstergeler harcama ve üretimde ılımlı bir büyüme işaret ediyor...* [<https://www.federalreserve.gov/newsevents/pressreleases/monetary20221214a.htm> adresinden alınmıştır, yer darlığı nedeniyle geri kalan kısım atlanmıştır].

Açıklama şahindir. FOMC, enflasyonun yükselmesinden endişe duyduğunu belirtmekte ve federal fon oranı için hedef aralığını yükselterek ve Hazine menkul kıymetleri ve ajans borçlarını azaltarak harekete geçmektedir. Açıklamada ayrıca, enflasyonu yüzde 2 hedefine geri döndürmek için gelecekte hedef aralığında daha fazla artış yapılmasının uygun olabileceği belirtilmektedir. Bu, şahin duruşun bir özelliği olan para politikasının sıkılaştırılmasına işaret etmektedir.

Değerlendirme doğru ve iyi savunulmuş.

LLM'nin Aralık 2022 veya Şubat 2023 FOMC açıklamasının daha şahin olup olmadığını belirleyip belirleyemeyeceğini de araştırdım, ancak Fed söylemini değerlendirme yeteneği yeterince incelikli değildi - açıklama metninde potansiyel bir yön değişikliğine işaret eden küçük ve incelikli değişikliklerin aksine, esas olarak Şubat 2023'teki faiz oranlarının seviyesinin daha yüksek olmasına odaklandı. Açıkça "federal fon oranı için hedef seviyeyi göz ardı ederek" değerlendirmesini rapor etmesi talimatını verdiğimde bile bunu yaptı. Sadece hedef rakamlarını değiştirdiğimde seviyesine göre "[aralık]" sistem doğru şekilde Şubat ayında yapılan değerlendirmeyi 2023 açıklaması, o dönemde finans basınında yaygın olarak bildirildiği gibi, biraz daha güvercindi.¹² Ardekani ve diğerleri (2023) benzer çizgide bir ekonomik duyarlılık tahmin modeli geliştirmiş ve bunu ABD ekonomik haberlerini ve ECB'nin para politikası duyurularını analiz etmek için kullanmıştır.

3.6.7 İnsan Deneklerin Simülasyonu

Argyle ve diğerleri (2023), LLM'lerin eğitim verilerinin insanlık hakkında büyük miktarda bilgi içerdiği gözlemine dayanarak, insan denekleri simüle etmek için LLM'lerin kullanılmasını önermektedir. GPT-3'ü gerçek insanların sosyodemografik geçmişleri üzerine koşullandırmışlar ve anket sorularına verilen müteakip yanıtların, tanımlanan geçmişe sahip insanların gerçek yanıtlarıyla nüanslı ve çok yönlü bir şekilde oldukça ilişkili olduğunu göstermişlerdir. Horton (2023), çeşitli davranışsal deneyleri çoğaltmak ve genişletmek için simüle edilmiş test deneklerini kullanarak ekonomiye yönelik uygulamaları sergilemektedir. Brand, Israeli ve Ngwe (2023), pazar araştırması için faydalı bilgiler elde etmek üzere LLM'lerin nasıl kullanılacağını göstermektedir.

¹² Bkz. örneğin, <https://www.cnbc.com/2023/02/01/>

Aşağıdaki örnek bu kavramı açıklamaktadır:

Sohbet 31 (GPT-4): Aşağıdaki seçmenlerin her biri için, yüksek eğitilmiş çalışanlara yönelik ABD göç kısıtlamalarını gevşetmeye yönelik bir politika önerisini onaylama veya onaylamama olasılıklarının daha yüksek olduğunu belirtmek üzere tek kelimelik bir derecelendirme yapınız:

1. Houston, TX'de yaşayan 32 yaşında Hispanik bir kadın avukat
 2. Rome, GA'da yaşayan 73 yaşında Kafkas kökenli bir erkek emekli doktor
 3. San Francisco, Kaliforniya'da yaşayan, Asyalı ve kendini nonbinary olarak tanımlayan 25 yaşında bir yazılım mühendisi
 4. Brooklyn, New York'ta yaşayan ve ikinci kuşak İtalyan olan 45 yaşında bir kadın market kasieri
 5. Lake City, FL'de yaşayan 55 yaşında Kafkas kökenli bir erkek sigorta acentesi
-

1. Onaylamak
 2. Onaylamıyorum
 3. Onaylamak
 4. Onaylamak
 5. Onaylamıyorum
-

Simüle edilen sonuçların yanlış kalıp yargıları yayması gibi önemli bir risk vardır ve bu nedenle büyük bir dikkatle kullanılmalıdırlar. Bununla birlikte, değerli bilgiler de içerirler. Doğru kullanıldıkları takdirde, LLM'leri eğitmek için kullanılan tüm verilerin büyük ölçüde kaynaklandığı toplumumuz hakkında faydalı içgörüler sağlayabilirler. Deneyisel ekono-

tutmayı tercih eden sisler, Charness, Jabarian ve List (2023) tarafından tanımlanmaktadır. LLM'lerin deneylerin tasarımını ve uygulanmasını iyileştirmeye nasıl yardımcı olabileceği.

3.7 Matematiksel Türevler

Genel amaçlı LLM'lerin matematiksel türetme yapabilme kabiliyeti artmaktadır ancak bu noktada hala sınırlıdır. Nispeten düşük performansın nedenleri arasında, mevcut LLM'ler için eğitim verilerinin

sistemler. Frieder ve arkadaşları (2023) lisansüstü düzeyde matematik sorularından oluşan bir veri seti geliştirmiş ve GPT3.5'in matematik becerilerinin ortalama yaştaki bir matematik lisansüstü öğrencisinininkinin önemli ölçüde altında olduğunu göstermiştir - belki de kısa vadede biraz teselli sağlamaktadır. Bununla birlikte, Noorbakhsh ve diğerleri (2023) LLM'lerin matematiksel görevler için de ince ayarlanabileceğini göstermiştir. Dahası, örneğin Bubeck ve diğerleri (2023) tarafından belgelendiği üzere, GPT-3.5'ten GPT-4'e geçenken matematikte kayda değer performans artışları olmuştur. Ayrıca, Frieder ve arkadaşları (2023) tarafından oluşturulan gibi veri kümeleri, gelecekteki LLM'lerin matematikte daha iyi olmalarını sağlamak için kesinlikle yararlı olacaktır. Bu, araştırmacılar için daha fazla ilerlemenin çok değerli olacağı bir alandır.

3.7.1 Modellerin Kurulması

Aşağıdaki komut istemi, LLM'lerin ekonomik modeller kurmak için nasıl yararlı olabileceğine bir örnektir. LLM'den doğrudan editörüne yapıştırabileceğim LaTeX kodu oluşturmasını istedim ve aşağıda gösterilen sonuçları oluşturdum.

Sohbet 32 (GPT-4): LaTeX kodunda iki mal üzerinde CES tercihleri olan bir tüketicinin optimizasyon prob- lemini kurun.

İki mal üzerinde CES (Sabit İkame Esnekliği) tercihi olan bir tüketicinin optimizasyon problemi aşağıdaki gibi kurulabilir

LaTeX kodu:

$$\max_{x_1, x_2} \alpha x_1^\rho + (1 - \alpha) x_2^\rho$$

çoğunlukla metin tabanlı olması ve nispeten az matematik içermesi ve matematiğin mevcut LLM'ler için hala zor olan üst düzey soyutlamalar gerektirmesi yer almaktadır.

$$\begin{aligned} & (x_1, x_2) \text{ Hızlı, Dahi Ucuz} \\ & p_1 x_{11} + p_2 x_{22} \leq M \text{ koşuluna tabidir,} \end{aligned}$$

Bu problemde tüketici, bir CES fonksiyonu ile temsil edilen fayda fonksiyonunu maksimize etmeye çalışmaktadır. Değişkenler x_1 ve x_2 tüketicinin seçtiği iki malın miktarlarını göstermektedir. α ve ρ parametreleri CES tercih parametreleri, p_1 ve p_2 parametreleri ise iki malın fiyatları ve M tüketicinin geliridir. Kısıt, tüketicinin bütçe kısıtını temsil etmektedir.



Bu yazının yazıldığı sırada, en gelişmiş LLM'ler aşağıdaki türde bir model üretebiliyordu

Örneğin, lisans problem setlerinde yaygın olarak kullanılan kurulumlar. Bu kullanışlıdır çünkü sonuçlar saniyeler içinde görünür ve yazarken zaman kazanın.

3.7.2 Denklem Türetme

LLM'lerin ayrıca bazı sınırlı yetenekleri

denklemler türetir. Yukarıdaki örneğe devam ederek, LLM'den kurduğu optimizasyon problemini çözmesini istedim:

Sohbet 32 (GPT-4) ctd.: Optimizasyon problemini çözün.

Bu optimizasyon problemini çözmek için Lagrangian yöntemini kullanmamız gerekir. Lagrangian fonksiyonu şöyledir:

$$(x_1, x_2, \lambda) = (\alpha x_1^\rho + (1 - \alpha) x_2^\rho)^{\frac{1}{\rho}} + \lambda (M - p_1 x_1 - p_2 x_2)$$

Daha sonra Lagrangian'ın x_1, x_2 'e göre farklılaştırarak birinci dereceden koşulları alırız, ve λ türevlerini sıfıra eşitlemek:

$$-\frac{\partial}{\partial x_1} (\alpha x_1^\rho + (1 - \alpha) x_2^\rho)^{\frac{1}{\rho}} - \lambda p_1 = 0,$$

$$\frac{\partial}{\partial x_2} (\alpha x_1^\rho + (1 - \alpha) x_2^\rho)^{\frac{1}{\rho}} - \lambda p_2 = 0,$$

$$\frac{\partial}{\partial \lambda} (M - p_1 x_1 - p_2 x_2) = 0.$$

Bu denklemlerin eş zamanlı olarak çözülmesi, bütçe kısıtlamalarına tabi olarak tüketicinin faydasını maksimize eden optimal x_1 ve x_2 miktarlarını verecektir.

Çözüm, parametrelerin belirli değerlerine, gelir ve fiyatlara bağlı olacaktır.

Şu ana kadar elde edilen birinci dereceden koşullar doğrudur.

Sohbet 32 (GPT-4) ctd: Talep fonksiyonlarını türetin.

Talep fonksiyonlarını türetmek için birinci dereceden koşulları kullanabiliriz. Elde etmek için ilk denklemleri ikinciye bölün:

$$\frac{x_1^{\rho-1}}{x_2^{\rho-1}} = \frac{p_1}{p_2}$$

[Yazarın notu: terim $\frac{1}{\rho-1}$ burada atlanmıştır].

$$\begin{aligned} \text{Yeniden düzenlemek verir:} \\ \frac{x_1}{x_2} &= \left(\frac{p_1}{p_2}\right)^{\frac{1}{\rho-1}} \end{aligned}$$

Bunu bütçe kısıtında yerine koyun $M = p_1 x_1 + p_2 x_2$ elde etmek için:

$$M = p_1 x_1 + p_2 x_2 \left(\frac{p_1}{p_2}\right)^{\frac{1}{\rho-1}}$$

x_1 için çözüldüğünde, 1 numaralı mal için talep fonksiyonunu verir:

$$x_1 = \frac{M}{p_1 + p_2 \left(\frac{p_2}{p_1}\right)^{\frac{1}{\rho-1}}}$$

Benzer şekilde, 2 numaralı mal için talep fonksiyonunu türetebiliriz:

$$x_2 = \frac{M}{p_2 + p_1 \left(\frac{p_1}{p_2}\right)^{\frac{1}{\rho-1}}}$$

Fonksiyonları kullanarak Marshallian talep

mal 1 ve 2.

Örnekte de görüldüğü üzere, LLM'ler matematiksel derivasyonlarda hala sıklıkla hata yapmaktadır. "İki denklemleri böldüğünüzde alfa terimlerini unuttunuz" diye girdim. LLM sonuçları yeniden türetmeye devam etti ve hatayı düzeltti, ancak bir sonraki satırda bir hata yaptı. Bu durum, mevcut LLM'lerin tutarlı bir düzeyde matematik yapmak için gerekli üst düzey soyut temsillere sahip olmadıklarını göstermektedir. En iyi ihtimalle, çok sayıda matematiği hızlı bir şekilde yazan ancak çok dikkatli bir gözetime ve potansiyel düzenlemeye ihtiyaç duyan bir

3.7.3 Modelleri Açıklamak

Mevcut LLM'ler de basit modelleri açıklama becerisine sahiptir. Bu durum öğrenciler için faydalı olabileceği gibi riskli de olabilir. Aşağıdaki örnekte, LaTeX kodunu LLM'ye yapıştırdım ve ondan altta yatan modeli ve bir çözüm türetme adımlarını açıklamasını istedim:

Sohbet 33 (GPT-4): *Bu modelin ne olduğunu, bir adımdan diğerine nasıl geçileceğini ve her adım için sezginin ne olduğunu açıklayabilir misiniz?* [Denklemler LaTeX formatında yapıştırılmıştır].

$$\begin{aligned}
 & \text{s.t.} \quad \begin{matrix} \text{maks} \\ \text{imu} \end{matrix} \quad \int_0^{\infty} e^{-\rho t} U(C(t)) dt \\
 & \quad \begin{matrix} m \\ C(t), K(t) \end{matrix} \\
 & \quad K'(t) = F(K(t), L) - \delta K(t) - C(t)K(0) \\
 & \quad \lim_{t \rightarrow \infty} K(t)e^{-\rho t} > 0 \text{ verildiğinde} \\
 & \quad (t, K(t), C(t), \lambda(t)) = e^{-\rho t} U(C(t)) \\
 & \quad + \lambda(t) F[K(t), L] - \delta K(t) - C(t) \\
 & \quad \frac{\partial \mathcal{L}}{\partial C(t)} = 0 \Rightarrow U'(C(t)) = \lambda(t) \\
 & \quad \frac{\partial \mathcal{L}}{\partial K(t)} + \lambda'(t) = 0 \Rightarrow \lambda(t) \\
 & \quad = -\lambda(t) [F'(K(t), L) - \delta] \\
 & \quad \lim_{t \rightarrow \infty} \lambda(t)K(t) = 0
 \end{aligned}$$

Bu, ekonomide genellikle Ramsey-Cass-Koopmans modeli olarak adlandırılan bir optimal tüketim ve sermaye birikimi modelidir. Neoklasik büyüme modelinin sürekli zamanlı bir versiyonudur ve temsili bir temsilci ömür boyu faydayı maksimize etmek için tüketim ve sermaye birikimi yollarını seçer.

Temsilcinin faydası zaman içinde ρ oranında iskonto edilir ve $U(C(t))$ fayda fonksiyonuna göre $C(t)$ tüketimine bağlıdır. Temsilcinin sermaye stoku $K(t)$, diferansiyel denklem tarafından verilen hareket kanununa göre gelişir; burada $F(K(t), L)$ üretim fonksiyonu, δ sermayenin amortisman oranı ve $C(t)$ tüketimdir.

$H(t, K(t), C(t), \lambda(t))$ ile gösterilen Hamiltonyen fonksiyonu, optimal kontrol teorisinde bu tür dinamik optimizasyon problemidir. t zamanının bir fonksiyonudur, durum değişkeni $K(t)$, kontrol değişkeni $C(t)$ ve sermayenin gölge fiyatı olan maliyet değişkeni $\lambda(t)$. Optimal konjonktür için birinci dereceden koşullar kontrol problemi, Hamiltonyen'in kontrol ve durum değişkenlerine göre türevlerinin alınması ve sıfıra eşitlenmesi ile türetilir. Bu koşullar, tüketim ve sermayenin zaman içindeki optimal yollarını karakterize etmektedir.

İlk koşul $U'(C(t)) = \lambda(t)$ tüketimin marjinal faydasının aşağıdakilere eşit olduğu anlamına gelir

sermayenin gölge fiyatıdır. Bu, temsilcinin ilave bir birim çıktı tüketmek ile sermayesini artırmak için tasarruf etmek arasında kayıtsız olduğu anlamına gelir.

sermaye stoku.

İkinci koşul olan maliyet denklemi, sermayenin gölge fiyatının zaman içindeki gelişimini açıklar. Buna göre değişim oranı

Gölge fiyatın negatif değeri, gölge fiyat ile amortisman düşülmüş sermayenin marjinal ürününün çarpımına eşittir. Bu, bugün tüketmek ile gelecekteki tüketim için tasarruf etmek arasındaki değiş tokuşu yansıtmaktadır.

Enine koşul, $\lambda(t)K(t) = 0$ t

sonsuz gittiğinde, temsilcinin

Sonsuz miktarda sermaye biriktirmek, tüketimin azalan marjinal faydası varsayımı nedeniyle optimal olmayacaktır. Bu koşul, temsilcinin sermayeyi sonsuza kadar biriktirdiği ve asla tüketmediği yolları dışlar.

LLM, matematikte tanımlanan modeli doğru bir şekilde tanımladı ve kalın harflerle vurguladığım küçük bir nokta dışında türevleri çoğunlukla doğru bir şekilde açıkladı: LLM, Hamiltonian'ın durum değişkenlerine göre türevinin sıfıra ayarlanmasını yanlış bir şekilde önerdi. Açıklamaların çoğu faydalı olduğundan, pek çok öğrencinin sunulan örnekte olduğu gibi kendilerine ders vermesi için LLM'lere başvurması muhtemeldir. Ancak bu, bilginin bir kısmının yanlış olma riskini de beraberinde getirmektedir - LLM yanılmaz bir öğretmen değildir.

TABLO 2
LLM yeteneklerinin ÖZETİ VE kullanışlılık DERECESİ

| Kategori | Görev | Yararlılık |
|-------------------------------|--|------------|
| Fikir Üretme ve Geri Bildirim | Beyin Fırtınası | ● |
| | Geri bildirim | ◐ |
| | Karşı argümanlar sunmak | ◐ |
| Yazma | Metin sentezleme | ● |
| | Metin düzenleme | ● |
| | Metin değerlendirme | ● |
| | Elle yazılmış denklemleri dönüştürme | ○ |
| | Akılda kalıcı başlıklar ve manşetler oluşturma | ● |
| | Bir makaleyi tanıtmak için tweet oluşturma | ● |
| Arka Plan Araştırması | Sunum slaytları oluşturma | ● |
| | Özetleme | ● |
| | Literatür Araştırması | ◐ |
| | Referansları Biçimlendirme | ● |
| | Metin Çevirme | ● |
| Kodlama | Kavramların Açıklanması | ◐ |
| | Kod yazma | ◐ |
| | Kodun açıklanması | ◐ |
| | Kod çevirme | ● |
| Veri Analizi | Kodda hata ayıklama | ◐ |
| | Veri kaynaklarını bulma | ◐ |
| | Rakamların oluşturulması | ◐ |
| | Metinden veri çıkarma | ● |
| | Verilerin yeniden biçimlendirilmesi | ● |
| | Metni sınıflandırma ve puanlama | ◐ |
| Matematik | Duyguları ayıklama | ◐ |
| | İnsan deneklerin simülasyonu | ◐ |
| | Modellerin kurulması | ◐ |
| | Denklemlerin türetilmesi | ○ |
| | Modellerin açıklanması | ◐ |

Not: Üçüncü sütun, Mayıs 2024 itibariyle LLM yeteneklerine ilişkin öznel derecelendirmemi bildirmektedir:

○: deneysel; sonuçlar tutarsız ve önemli ölçüde insan gözetimi gerektiriyor

◐: yararlıdır; gözetim gerektirir ancak muhtemelen size zaman kazandıracaktır

●: son derece kullanışlı; bunu iş akışınıza dahil etmek size zaman kazandıracak

3.8 Özet

Tablo 2, bu makalede gösterilen tüm örnek görevleri, LLM'lerin tanımlanan altı uygulama alanına göre kategorize ederek özetlemektedir. Tablonun üçüncü sütununda, Mayıs 2024 itibariyle tanımlanan LLM yeteneklerini ne kadar faydalı bulduğuma dair öznel derecelendirmemi bildiriyorum. Derecelendirmelerim, boş bir dairenin (○) yetenekleri tanımladığı üç kategoride yer almaktadır

Şu anda daha deneysel olduğunu düşündüğüm ve tutarsız sonuçlar veren, önemli ölçüde insan gözetimi gerektiren; yarı dolu bir daire (◐), kullanışlı ve muhtemelen zaman kazandıran ancak biraz tutarsız olan, dolayısıyla hala dikkatli bir gözetim gerektiren yetenekleri ifade eder; ve tam daire (●), zaten oldukça kullanışlı olan ve çoğu zaman beklenen şekilde çalışan yetenekleri yansıtır. Bu son kabiliyetlerin birleştirilmesi

kesinlikle size zaman kazandıracak ve sizi daha üretken hale getirecektir.

Bu özet tabloyu, 3. bölümün geri kalanıyla birlikte, bu makaleyle ilişkili çevrimiçi kaynaklarda (bkz. başlık dipnotu) düzenli olarak güncelleyeceğim, böylece ekonomik araştırmalar için üretken YZ'nin yararlılığının güncel bir değerlendirmesini sunacağım.

Devam eden bir araştırma dizisi, üretici YZ'nin yalnızca araştırmadaki belirli mikro görevleri değil, aynı zamanda araştırma sürecinin daha geniş unsurlarını otomatikleştirmek için nasıl kullanılabileceğini araştırmaktadır.

Örneğin, Yang ve diğerleri (2024), bir yazılım projesi için çok çeşitli kodlama görevlerini otonom olarak gerçekleştirebilen bir YZ aracısının nasıl geliştirileceğini göstermektedir. Koehler ve Sauermann (2024), organizasyonel literatürde vurgulanan beş temel yönetsel işleve odaklanarak, LLM'lerin bilimsel araştırmalarda algoritmik yönetim oluşturma becerilerini incelemektedir: görev bölümü ve görev dağılımı, yönlendirme, koordinasyon, motivasyon,

ve öğrenmeyi desteklemek.

Referanslar

- Anderljung, Markus ve diğerleri. 2023. "Frontier AI Regulation: Managing Emerging Risks to Public Safety." <https://doi.org/10.48550/arXiv.2307.03718>.
- Ardekani, Aref Mahdavi, Julie Berts, Michael M. Downing ve Suwan Long. 2023. "EconSentGPT: A Universal Economic Sentiment Engine?" <http://dx.doi.org/10.2139/ssrn.4405779>.
- Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting ve David Wingate. 2023. "Birden Çok Kişi: İnsan Örneklerini Simüle Etmek için Dil Modellerini Kullanmak." *Siyasi Analiz* 31 (3): 337-51.
- Bai, Yuntao ve diğerleri. 2022. "Constitutional AI: Harmlessness from AI Feedback." <https://doi.org/10.48550/arXiv.2212.08073>.
- Bommasani, Rishi ve diğerleri. 2021. "On the Opportunities and Risks of Foundation Models." <https://doi.org/10.48550/arXiv.2108.07258>.
- Brand, James, Ayelet Israeli ve Donald Ngwe. 2023. "Pazar Araştırması için GPT Kullanımı." Harvard Business School Çalışma Belgesi 23-062.
- Bsharat, Sondos Mahmoud, Aidar Myrzakhan ve

Zhiqiang Shen. 2024. "Principled Instructions are All You Need for Questioning LLaMA-1/2, GPT-3.5/4." <https://arxiv.org/abs/2312.16171>.

- Bubeck, Sébastien ve diğerleri. 2023. "Yapay Genel Zekanın Kıvılcımları: GPT-4 ile Erken Deneyler." <https://doi.org/10.48550/arXiv.2303.12712>.
- Buchanan, Joy ve Olga Shapoval. 2023. "GPT-3.5 Varolmayan A t ı f l a r ı Halüsinasyonla Gösteriyor: Evidence from Economics." <http://dx.doi.org/10.2139/ssrn.4467968>. Charness, Gary, Brian Jabarian, ve John A. List. 2023. "Generation Next: Yapay Zeka ile Deneyler." NBER Çalışma Belgesi 31679.
- Dunn, Alexander ve diğerleri. 2022. "İnce Ayarlı Büyük Dil Modelleri ile Karmaşık Bilimsel Metinlerden Yapılandırılmış Bilgi Çıkarımı." <https://doi.org/10.48550/arXiv.2212.05238>.
- Frieder, Simon ve ark. 2023. "Mathematical Capabilities of ChatGPT." <https://doi.org/10.48550/arXiv.2301.13867>.
- Gentzkow, Matthew, Bryan Kelly ve Matt Taddy. 2019. "Veri Olarak Metin." *Journal of Economic Literature* 57 (3): 535-74.
- Girotra, Karan, Lennart Meincke, Christian Terwiesch ve Karl T. Ulrich. 2023. "Fikirler Onda Birdir: İnovasyonda Fikir Üretimi için Geniş Dil Modelleri." <http://dx.doi.org/10.2139/ssrn.4526071>. Girotra, Karan, Christian Terwiesch ve Karl T. Ulrich. 2010. "Fikir Üretimi ve En İyi Fikrin Kalitesi Fikir." *Yönetim Bilimi* 56 (4): 591-605.
- Horton, John J. 2023. "Simülasyon Olarak Büyük Dil Modelleri ulated Economic Agents: Homo Silicus'tan Ne Öğrenebiliriz?" NBER Çalışma Belgesi 31122.
- Jiao, Wenxiang, Wenxuan Wang, Jen-tse Huang, Xing Wang ve Zhaopeng Tu. 2023. "ChatGPT İyi Bir Çevirmen mi? Evet, Motor Olarak GPT-4 ile." <https://doi.org/10.48550/arXiv.2301.08745>.
- Koehler, Maximilian ve Henry Sauermann. 2024. "Bilimsel Araştırmalarda Algoritmik Yönetim." *Araştırma Politikası* 53 (4): 104985.
- Korinek, Anton. 2023a. "Ekonomik Araştırmalar için Dil Modelleri ve Bilişsel Otomasyon." NBER Çalışma Belgesi 30957.
- Korinek, Anton. 2023b. "Ekonomik Araştırma için Üretken Yapay Zeka: Kullanım Örnekleri ve Ekonomistler için Çıkarımlar." *Journal of Economic Literature* 61 (4): 1281-1317.
- Noorbakhsh, Kimia, Modar Sulaiman, Mahdi Sharifi, Kallol Roy ve Pooyan Jamshidi. 2023. "Pretrained Language Models Are Symbolic Mathematics Solvers Too!" <https://doi.org/10.48550/arXiv.2110.03501>.
- Peng, Sida, Eirini Kalliamvakou, Peter Cihon ve Mert Demirer. 2023. "Yapay Zekanın Gelişim Verimliliği Üzerindeki Etkisi: GitHub Copilot'tan Kanıtlar." <https://doi.org/10.48550/arXiv.2302.06590>.
- Seger, Elizabeth ve diğerleri. 2023. "Yüksek Yetenekli Temel Modellerin Açık Kaynak Kullanımı: Açık Kaynak Hedeflerine Ulaşmak için Riskler, Faydalar ve Alternatif Yöntemler Üzerine Bir Değerlendirme." <https://arxiv.org/>

abs/2311.09227.

Wei, Jason, Xuezhi Wang ve diğerleri. 2022. "Düşünce Zinciri Sorgulaması Büyük Dil Modellerinde Akıl Yürütmeyi Ortaya Çıkarır." In *Advances in Neural Information Processing System (NeurIPS 2022)*, Vol. 35, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho ve A. Oh. Red Hook, NY: Curran Associates Inc.

Yang, John ve diğerleri. 2024. "SWE-agent: Agent-Computer Interfaces Enable Automated Software Engineering." <https://arxiv.org/abs/2405.15793>.

Zou, Andy, Zifan Wang, J. Zico Kolter ve Matt Fredrikson. 2023. "Hızlandırılmış Dil Modellerine Evrensel ve Aktarılabilir Karşı Saldırıları." <https://doi.org/10.48550/arXiv.2307.15043>.