

Makale Raporu

Hazırlayan: Muhammed Kayra Bulut
Ders: Doğal Dil İşlemeye Kavramsal Bir Bakış
Ders Yürütücüsü: Prof. Dr. Banu Diri

Nisan 2024

1 Makale Adı

Almanac — Retrieval-Augmented Language Models for Clinical Medicine

2 LLM Nedir?

Büyük Dil Modelleri (LLM'ler) veya Large Language Models, insan dilini anlayabilen ve Derin Öğrenme teknolojileri sayesinde büyük veri kümelerinden bilgi edinen yapay zeka sistemleridir. Bu sistemler, metin anlama, sonuç çıkarma ve dil tabanlı görevleri gerçekleştirme gibi çeşitli becerilere sahiptir. LLM'ler, benzersiz yetenekleri nedeniyle akademik ve endüstriyel alanlarda giderek daha fazla ilgi görmektedir. Özellikle son dönemde, GPT gibi modellerin gelişimi, LLM'lerin kabiliyetlerini büyük ölçüde artırmıştır.

2.1 Eğitim İçin Gerekli Donanım ve Veri Setleri

Büyük Dil Modelleri'nin (LLM) eğitimi için gerekli ana kaynaklar arasında güçlü grafik işlemcileri ve kapsamlı veri setleri yer alır. Bu modellerin eğitimi, yoğun hesaplama gücü gerektiren bir işlemdir ve sıklıkla birden fazla GPU veya TPU gibi özel donanımlar kullanılarak yapılır. Eğitim süreci, modellerin geniş bir dil yelpazesine ve çeşitli konularda kapsamlı bilgiye sahip olmasını sağlamak için büyük ve çeşitli veri kümelerinden faydalanır.

Geniş veri setleri, modellerin dilin karmaşıklığını ve ince ayrıntılarını öğrenmesi için hayati öneme sahiptir. Bu veri setleri, internetten toplanan milyarlarca kelimeyi içerebilir ve farklı dillerdeki metinler, konuşma dilindeki çeşitlilikler ve çeşitli konulardaki içerikler gibi geniş bir spektrumu kapsar. LLM'ler, bu geniş veri setlerini analiz ederek dilin doğal kullanımını kavrama ve taklit etme yeteneklerini geliştirir.

2.2 Bizi Nasıl Anlar?

Büyük Dil Modelleri (LLM'ler), insan dilini anlama ve işleme yeteneklerini, metinlerdeki kelimeleri ve ifadeleri matematiksel vektörlere dönüştürerek gerçekleştirir. Bu süreç, genellikle tokenizer yapısı ve özel token'lar (special tokens) kullanılarak yönetilir. Tokenizerlar, metni daha küçük parçalara ayırırken, özel token'lar ise modele dilin belirli özelliklerini ve yapılarını öğretmek için kullanılır. Örneğin, bir cümledeki soru cümlesi olduğunu modele daha iyi anlatmak için "i—soru—i" token'ı veya cümledeki sonunun geldiğini belirtmek için "i—son—i" gibi token'lar kullanılabilir.

2.2.1 Tokenizer Nedir

Tokenizerlar, metni modele uygun şekilde kelime, kelime parçaları veya semboller gibi daha küçük birimlere (token'lara) ayırır. Bu işlem, modelin dilin yapısal özelliklerini ve kelime arası ilişkileri daha iyi kavramasını sağlar. Örneğin, "Bilgisayar" kelimesi bir tokenizer tarafından "Bilgi", "##sayar" gibi parçalara ayrılabilir. Bu ayırım, modelin hem "Bilgi" hem de "sayar" anlamlarını öğrenmesine ve daha geniş bir dil bilgisini kavramasına yardımcı olur.

2.2.2 Özel Token Ne Anlama Gelir

Özel token'lar, metinle ilgili ilave bilgileri modele aktarmak amacıyla kullanılan öğelerdir. Bunlar arasında, cümledeki başı ve sonunu gösteren [CLS] ve [SEP] token'ları, metindeki tanımlanamayan kelimeler için kullanılan [UNK] token'ı gibi öğeler bulunur. Mesela, [CLS] token'ı, sınıflandırma işlemlerinde modelin sonucu belirleyeceği noktayı işaret ederken; [SEP] token'ı, iki ayrı cümle veya metin parçasını birbirinden ayırmak için kullanılır.

Token'lar ve özel token'lar, modelin dilin hem lokal (kelime ve kelime grupları bazında) hem de global (cümle ve paragraf seviyesinde) yapısını kavramasını sağlar. Bu özellikler, LLM'lerin metni doğru bir şekilde yorumlamasında, bağlamı muhafaza etmesinde ve dilin inceliklerini başarıyla işleyebilmesinde önemli bir rol oynamaktadır.

3 Genel Kullanım İçin LLM'ler

Genel amaçlı Büyük Dil Modelleri (LLM'ler), geniş bir dil bilgisine sahip olup, çeşitli konular ve görevler üzerinde çalışabilir. Bu modeller, büyük veri setlerinden öğrenir ve çok yönlü kullanım imkanı sunar. Örneğin, GPT-4, Bard, Bing gibi modeller, basit tıbbi soruları yanıtlamaktan, hasta danışmanlığına ve hatta tıbbi literatür taramasına kadar bir dizi görevi yerine getirebilir.

Ancak, genel amaçlı LLM'lerin sağlık sektörüne özgü terminoloji ve konseptleri anlamada sınırlılıkları vardır. Bu modeller, spesifik tıbbi bilgilere veya karmaşık klinik senaryolara cevap vermekte zorlanabilir, çünkü eğitim verileri genellikle tıbbi olmayan metinlerden oluşur. Aynı zamanda, bu LLM'leri eğitmek ve çalıştırmak oldukça zaman alıcı ve maliyetli olabilir.

3.1 Genel Amaçlı Modellerin Eğitimi

Genel amaçlı Büyük Dil Modellerinin (LLM'ler) eğitimi ve çalıştırılması, hem maliyetli hem de zaman alıcı bir süreçtir ve bu süreç, geniş kapsamlı veri setleri ve güçlü bilgisayar donanımı gerektirir. Bu modeller, çeşitlilik ve geniş kapsamda bilgi sunabilmek için milyarlarca kelimeyi içeren büyük veri setleri üzerinde eğitilir. Eğitim süreci, dilin çeşitli yönlerini ve nüanslarını öğrenmek için geniş bir veri yelpazesini kapsar. Bu da modeli eğitmek için yüksek performanslı bilgisayar sistemleri ve büyük ölçekli veri işleme kapasitesine, ayrıca yüksek maliyetlere ihtiyaç olduğu anlamına gelir.

3.1.1 Donanım Gereklilikleri

Genel amaçlı Büyük Dil Modellerinin (LLM) geliştirilmesi ve devamlı gelişimi için eğitim ve çalıştırma süreçleri hayati önem taşır. Bu süreçlerin etkili bir şekilde gerçekleştirilmesi, güçlü GPU'lar veya TPU'lar gibi üst düzey donanım kaynaklarını zorunlu kılar. Büyük ve karmaşık modellerin eğitimi özellikle, ciddi bir maliyet ve zaman yatırımı gerektirir.

3.1.2 Donanım Maliyeti

Güçlü GPU'lar ve TPU'lar, genel amaçlı Büyük Dil Modellerinin (LLM) eğitim sürecinde paralel işlemleri hızla gerçekleştirerek eğitim süresini önemli ölçüde azaltabilir. Ancak, bu yüksek performanslı donanım birimlerinin maliyeti, özellikle birçok birimin aynı anda kullanılması gerektiğinde, projenin bütçesine önemli bir yük getirebilir. Ayrıca, bu donanımların enerji tüketimi de dikkate alındığında, işletme maliyetleri daha da artabilir.

3.1.3 Yönetim Sıkıntıları

Genel amaçlı Büyük Dil Modellerinin (LLM) eğitimi ve sürekli geliştirilmesi süreçlerinin genişletilmesi, teknik ve mali zorluklar barındırır. Modelin büyüklüğü arttıkça, eğitim için gerekli olan veri miktarı ve hesaplama kapasitesi de artar. Bu durum, büyük ölçekli eğitim operasyonlarının yönetilmesini zorlaştırır ve etkili bir genişletme stratejisi gerektirir. Genişletme çabaları, veri depolama ve işleme altyapısının yanında, donanım ve yazılım optimizasyonlarını da içermelidir.

3.1.4 Güncel Bilgilere Uyumu

Genel amaçlı LLM'ler, sağlık sektörü gibi hızla gelişen alanlarda güncel bilgileri içerecek şekilde sürekli güncellenmelidir. Ancak, bu sürekli güncelleme işlemi, hem zaman hem de maliyet açısından zorlayıcı olabilir. Modelin eğitim verilerini güncel tutmak ve modeli en son bilgilerle yeniden eğitmek, devasa boyutlarda maliyete sahiptir.

4 Özelleştirilmiş LLM'ler

Özelleştirilmiş büyük dil modelleri (LLM'ler) üzerine konuşalım. Bu modeller, belirli bir alana ya da göreve yönelik derinlemesine eğitilmişlerdir ve bu sayede, özellikle sağlık sektörü gibi karmaşık terminolojilerin bolca kullanıldığı yerlerde, büyük bir fark yaratırlar. Bu özelleştirme sayesinde, modellerin hızı ve doğruluğu ciddi şekilde artarken, ihtiyaç duyulan donanım kaynakları ve model boyutu da önemli ölçüde azalır. Sağlık sektörüne özel olarak eğitilen bu modeller, klinik notlar, medikal araştırma makaleleri, hasta raporları ve diğer alana özel veri setleri üzerinden beslenirler. Bu süreç, modellere, medikal terminoloji ve klinik senaryolar hakkında detaylı cevaplar verebilme kapasitesi kazandırır. Böylece, teşhis koyma, tedavi önerileri ve medikal literatür analizi gibi görevlerde daha doğru, hızlı ve güvenilir sonuçlar elde edilir.

4.1 Hız ve Verimlilik

Daha özel bir bakış açısıyla, Özelleştirilmiş Büyük Dil Modelleri (LLM'ler), kendilerine verilen spesifik görevlerde, genel amaçlı modellere kıyasla daha hızlı ve daha etkin bir şekilde iş görürler. Bu modeller, odaklandıkları belirli alanlardaki veri setleriyle eğitildikleri için, gereksiz bilgileri ayıklama ve ilgisiz içeriği ele alma konusunda daha az zaman harcarlar.

4.2 Fiyat/Performans

Geliştirmesi ve eğitilmesi sırasında, özelleştirilmiş Büyük Dil Modelleri (LLM'ler), her türlü görev için tasarlanmış genel amaçlı modellerle karşılaştırıldığında oldukça verimli bir yola sahip. Bunların sırrı, belirli bir iş veya konu alanına özel olarak odaklanmaları. Yani, onlar için önemli olan, geniş ve çeşitli konuları barındıran devasa veri yığınlarını tarayıp anlam çıkarmak değil. Bunun yerine, daha dar bir alanda yer alan ve dolayısıyla daha az kaynak gerektiren özel veriler üzerine yoğunlaşırlar. Özelleştirilmiş modeller, büyük veri setleriyle sürekli eğitim ihtiyacı duymadan, daha küçük ve yönlendirilmiş veri kümeleri kullanılarak hem daha ekonomik hem de daha hızlı bir şekilde form kazanabilir. Bu da, maliyetleri düşürme ve etkinliği artırma konusunda önemli bir avantaj sağlar.

4.3 Doğruluk ve Kesinlik

Özel eğitim veri setlerinin gücünü arkasına alan özelleştirilmiş Büyük Dil Modelleri (LLM'ler), sağlık sektöründe çığır açıyor. Diagnostikten tedavi planlamasına, medikal literatürün derinlemesine incelenmesine kadar geniş bir yelpazede, yüksek doğruluk ve kesinlik sunuyorlar. Bu modellerin sırrı, karmaşık medikal terimleri ve klinik senaryoları anlamaya yönelik özel geliştirmelerinden kaynaklanıyor. Bu özellikleri sayesinde, genel amaçlı modellere göre çok daha güvenilir sonuçlar elde ediyorlar. Klinik karar destek sistemlerinin geliştirilmesinden hasta bakımının iyileştirilmesine kadar, özelleştirilmiş modeller, sağlık sektöründe vazgeçilmez ve etkili birer araç haline geliyor.

4.4 Uygulama Esnekliđi

Özelleştirilmiş Büyük Dil Modelleri (LLM'ler), sađlık sektörünün dinamik ve çeşitliliđine mükemmel bir şekilde uyum sađlıyor. Bu modeller, özel sađlık hizmeti gereksinimlerine göre şekillendirilebilir ve buna göre ayarlanabilir, ki bu da onları çeşitli medikal alanlarda ve spesifik sađlık hizmeti görevlerinde son derece etkili kılar. Mesela, kanser tanısı koyma işlemleri için özelleştirilmiş bir model geliştirilebilirken, kalp hastalıklarının yönetimi konusunda uzmanlaşmış başka bir model tasarlanabilir. Bu esnek yapı, sađlık hizmetlerinin çok geniş bir alanda iyileştirilmesine imkan tanıyarak, her bir sađlık sorununa özel çözümler sunar.

5 Sađlık Sektöründe Kullanım

Sađlık sektöründe Büyük Dil Modellerinin (LLM) kullanımı gün geçtikçe artıyor ve bu trendin, doktorlar, sađlık çalışanları ve hastalar tarafından gelecekte daha da fazla benimseneceđi düşünülüyor. Bu teknolojiler, teşhislerin kesinleştirilmesi, tedavilerin planlanması ve hasta takibinin kolaylaştırılması gibi alanlarda önemli fırsatlar sunarak sađlık hizmetlerinde devrim yaratma potansiyeline sahip. Ancak, bu potansiyelin büyük bir kısmı henüz keşfedilmemiş durumda ve bu teknolojilerin kullanımı şimdilik oldukça sınırlı. Bu durumun arkasında yatan nedenler arasında, etik kaygılar ve bu teknolojilerin sađlık sektöründeki potansiyellerinin henüz tam anlamıyla fark edilmemiş olması bulunuyor.

5.1 Tahlil Performansının Deđerlendirilmesi

Medikal veriler, özellikle tahlil sonuçları ya da MR görüntüleri, Büyük Dil Modelleri (LLM'ler) kullanılarak yapılan ön deđerlendirmeler için harika bir kaynak olabilir. Bu modeller, görsel verileri işleme ve metin analizi yeteneklerine sahip oldukları için, incelenen verilerden önemli bilgileri ayıklayabilir ve doktorların özellikle üzerinde durması gereken noktalara dikkat çekebilir. Bu süreç, teşhis koyma aşamasını hızlandırabilir ve doktorların daha bilinçli ve etkili kararlar vermesine yardımcı olabilir.

Örneđin *Şekil 1*'de ele alındığı gibi, klinik veri girişinden başlayıp, analiz ve deđerlendirme zinciri boyunca ilerleyen süreç, Büyük Dil Modelleri (LLM'ler) sayesinde doktorların incelemesine doğruluk, hız ve güncellik katkısı sađlar. Bu durum, özellikle hastalıkların erken evrede teşhis edilmesi gibi kritik öneme sahip anlarda, doktorlara daha net ve çabuk karar verme yeteneđi sunar. Böylelikle, doktorların aynı anda daha çok doğru teşhis koymalarına olanak tanınır, bu da hastaların erken tedaviye başlamasını ve dolayısıyla sađlık sonuçlarının iyileşmesini sađlar.

5.2 Hastalar İçin Kullanım

Hastalar, sađlık durumları hakkında bilgi almak ve basit sađlık sorunları için önerilerde bulunmak amacıyla LLM'leri kullanabilirler. Hastalar, semptomlarını

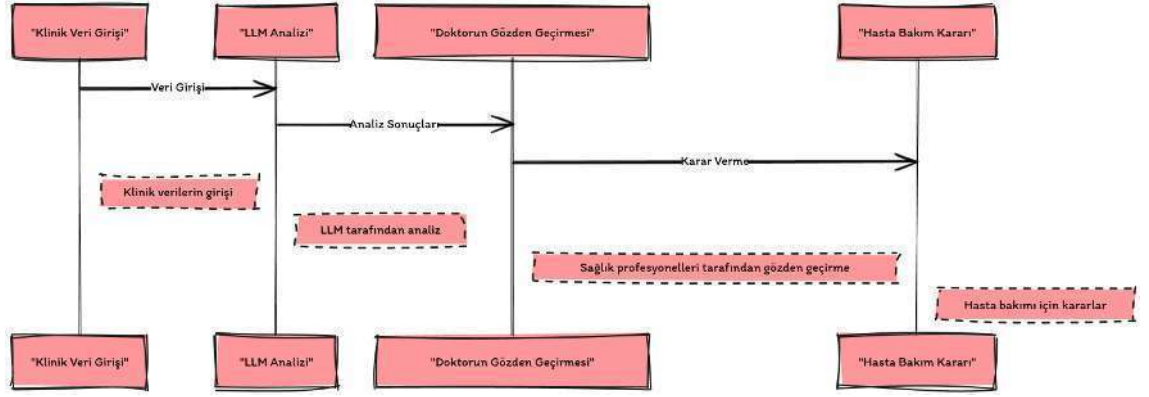


Figure 1: Kullanım Senaryosu Akışı

bir LLM modeline girerek, olası hastalıklar ve alınabilecek önlemler hakkında bilgi alabilirler. Bu, özellikle doktora ulaşımın zor olduğu durumlarda veya hızlı bilgiye ihtiyaç duyulduğunda faydalı olabilir.

Örneğin Şekil 2'deki akışa baktığımızda hasta, LLM'e soru sorup hastalığı hakkında daha nokta atışı tavsiye ve bilgilere sahip olabilir. Bu sayede, eğer hastanın haberi olmadığı ve hastalığın erken teşhis edilmesi gerektiği durumlarda LLM'ler hayat kurtarıcı bir rol oynar.

5.3 Sağlık Çalışanları İçin Kullanım

Doktorlar ve sağlık çalışanları, teşhis koyma ve tedavi planlaması gibi süreçlerde Büyük Dil Modelleri (LLM'ler) kullanarak büyük bir avantaj elde edebilirler. Bu modeller, hasta semptomlarını derinlemesine analiz edebilir ve potansiyel hastalıklar hakkında önemli ipuçları sunarak, doktorların hızla ve doğru bir şekilde teşhis koymasına yardımcı olabilir. Dahası, LLM'ler, medikal literatürde yer alan en güncel bilgileri süzerek, belirli bir hastalık için en uygun tedavi yöntemleri konusunda öneriler sunabilir. Bu, hem hastaların en iyi bakımı almasını sağlar hem de sağlık çalışanlarının zamanını ve kaynaklarını daha verimli kullanmalarına olanak tanır.

Örneğin Şekil 2'deki akışı incelediğimizde doktor hastanın semptomlarını LLM'e danışarak daha etkili ve hızlı bir biçimde inceleyip teşhis koyabilir. Hasta ve sağlık çalışanları da benzer LLM'lerden güzelce yararlanabilirler.

5.4 Tedavi Takibi ve Yönetimi

Büyük Dil Modelleri (LLM'ler), tedavi planlarının ve ilaç takibinin yönetilmesi gibi önemli süreçlerde de ön plana çıkıyor. Hastalar, bu modeller aracılığıyla tedavi planlarını kolayca takip edebilir ve ilaç kullanımlarını düzenleyebilirler. Yan etkiler veya tedavi hakkında soruları olduğunda, LLM'ler sayesinde hızlı

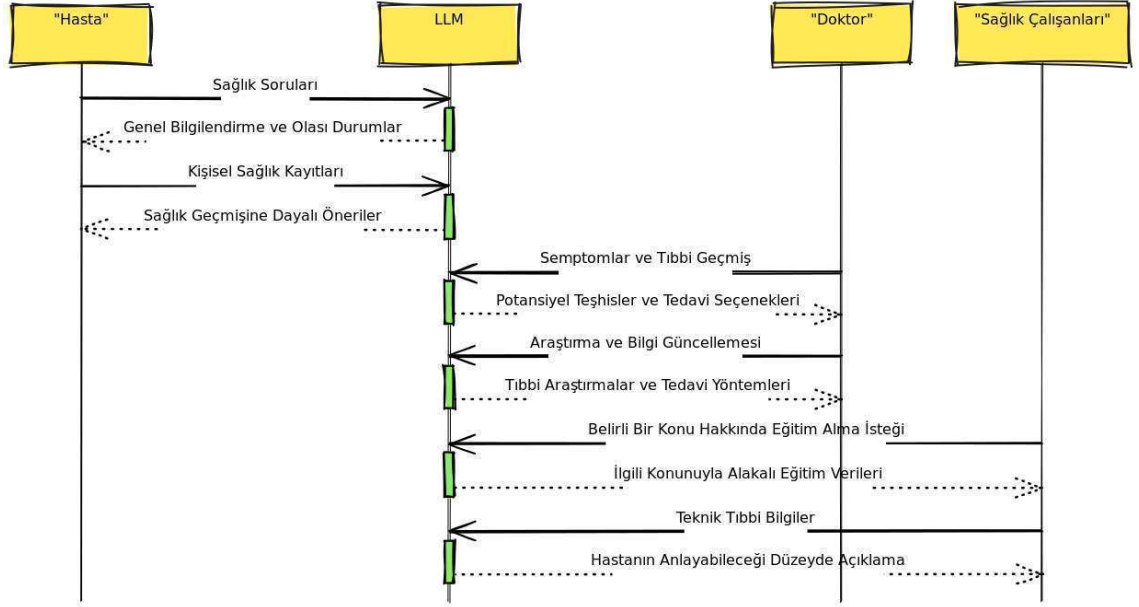


Figure 2: LLM Kullanım Akışı

ve güvenilir bilgilere ulaşabilirler. Dahası, LLM'ler, hastaların sağlık durumları ve tedavilere olan uyumları hakkında değerli geri bildirimler sunarak, tedavi süreçlerinin daha kişiselleştirilmiş ve etkili bir şekilde yönetilmesine imkan tanıyor. Bu teknoloji sayesinde, hastalar kendi sağlık yönetimlerinin daha aktif birer katılımcısı haline gelirken, doktorlar ve sağlık profesyonelleri de tedavi süreçlerini daha iyi optimize edebilir.

Örneğin *Şekil 7*'de, klinik veri girişi ile başlayan analiz ve değerlendirme zinciri, LLM'lerin karar vermesiyle hasta bakım kararına ve sonrasında tedavi sürecine kadar uzanır. Sonrasında tedavi uygulamaları sağlık personelleri tarafından gerçekleştirilirken, geri bildirimler LLM'e verilir ve LLM tedavi sürecini bu geri bildirimlere göre yönetir. Hasta durum bilgisine göre tedaviyi ağırlaştırma, hafifletme ve bitirme kararlarını LLM verebilir. Bu sayede, doktorların rehberliğinde, erişilebilir, hızlı, ekonomik ve konforlu bir tedavi süreci gerçekleşmiş olur.

6 Almanac ile Genel Modellerin Kıyaslanması

Bu bölümde, özel amaçlı bir dil modeli olan Almanac 'ın, genel amaçlı dil modelleri Bard, Bing ve GPT-4 ile performans karşılaştırması yapılacaktır. Sağlık sektörüne özel olarak geliştirilmiş Almanac modeli, bu alanda daha etkin ve doğru bilgi sağlama potansiyeline sahiptir. Genel amaçlı modellerse daha geniş bir uygulama alanına sahip olmakla birlikte, bu modellerin sağlık sektörüne

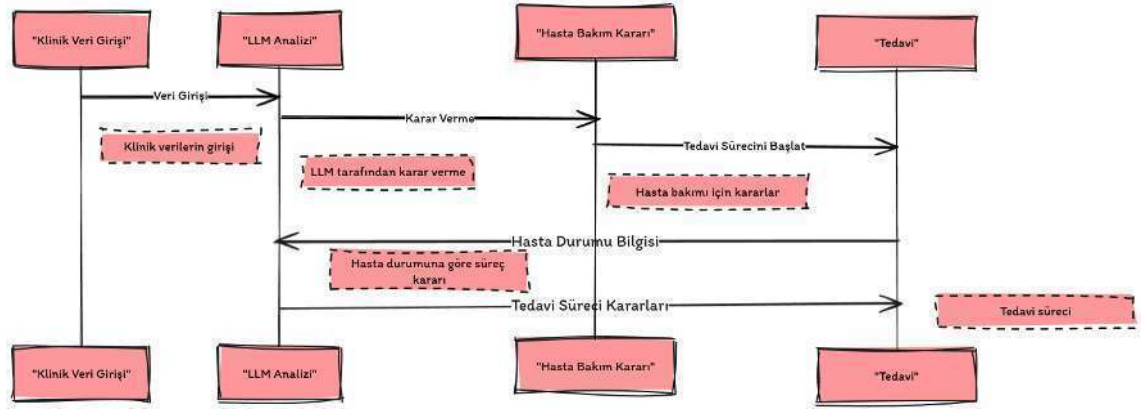


Figure 3: Tedavi Takip Akışı

özgü gereklilikleri karşılamada bazı sıkıntılar yaşayabileceği düşünülmektedir.

Karşılaştırmada, her bir modelin sağlık sektöründe karşılaşılabilecek gerçek dünya sorunlarına nasıl cevap verdiği, gerçeklik (factuality), tamamlılık (completeness) ve kullanıcı tercihi (preference) gibi metrikler üzerinden değerlendirilecektir. Ayrıca modellerin doğru bilgiye ne kadar atıf yaptığı ve aynı zamanda yanıltıcı komutlara karşı duyarlılığı da incelenecektir.

Bu kıyaslama, Almanac'ın uzmanlık alanı olan sağlık sektöründe, genel amaçlı modellere göre sunduğu avantajları ve sağlık sektöründe, genel amaçlı modellerin sahip olduğu sıkıntılar detaylı bir şekilde ele alınacaktır. Karşılaştırmanın amacı, her modelin güçlü ve zayıf yönlerini ortaya koymak ve özellikle sağlık sektörü gibi kritik alanlarda kullanım için yol gösterici bilgiler sağlamaktır.

6.1 Gerçeklik - Tamamlılık - Kullanıcı Tercihi

Özellikle sağlık sektörüne özgü LLM'lerin performansını değerlendirmek için çeşitli metrikler ön plana çıkar. Bu metrikler arasında gerçeklik (factuality), tamamlılık (completeness) ve kullanıcı tercihi (preference) bulunmaktadır. Gerçeklik, bir modelin ürettiği bilginin doğruluğunu; tamamlılık, verilen bilginin konuyu ne derece kapsadığını; kullanıcı tercihi ise kullanıcıların modelin yanıtlarını ne kadar yararlı bulduğunu gösterir.

Şekil 4 ve Şekil 5'te, modelin başarısını ölçen üç kritik metrik temel alınarak bir karşılaştırma yapılmıştır. İncelediğimizde, doğruluk açısından Almanac'ın önde olduğunu, onu ChatGPT-4'ün izlediğini, Bing'in üçüncü, Bard'ın ise son sırada yer aldığını görüyoruz. Bing ile ChatGPT-4 arasındaki performans farkı göz ardı edilebilir düzeydedir. Özellikle Bard, diğerlerine kıyasla belirgin bir şekilde daha düşük bir başarı sergilemiştir, buna karşılık Almanac'ın performansı dikkat çekici derecede yüksektir. Doğruluk metriğinde benzer bir sıralama görülürken, Almanac ve ChatGPT-4 arasındaki fark biraz daha belirginleşmiştir. Kullanıcı tercihleri bakımından da Almanac ilk sırada yer alırken, ardından

ChatGPT-4, Bing ve Bard sıralanmıştır. Yine, Bing ile ChatGPT-4 arasındaki fark önemsizdir, ancak Bard'ın performansı açıkça en düşük seviyededir. Almanac'ın, doğruluk ve tamamlılıkta rakiplerine göre üstün performans gösterdiği ve bu alanda lider konumda olduğu bu verilerle ortaya konmaktadır.

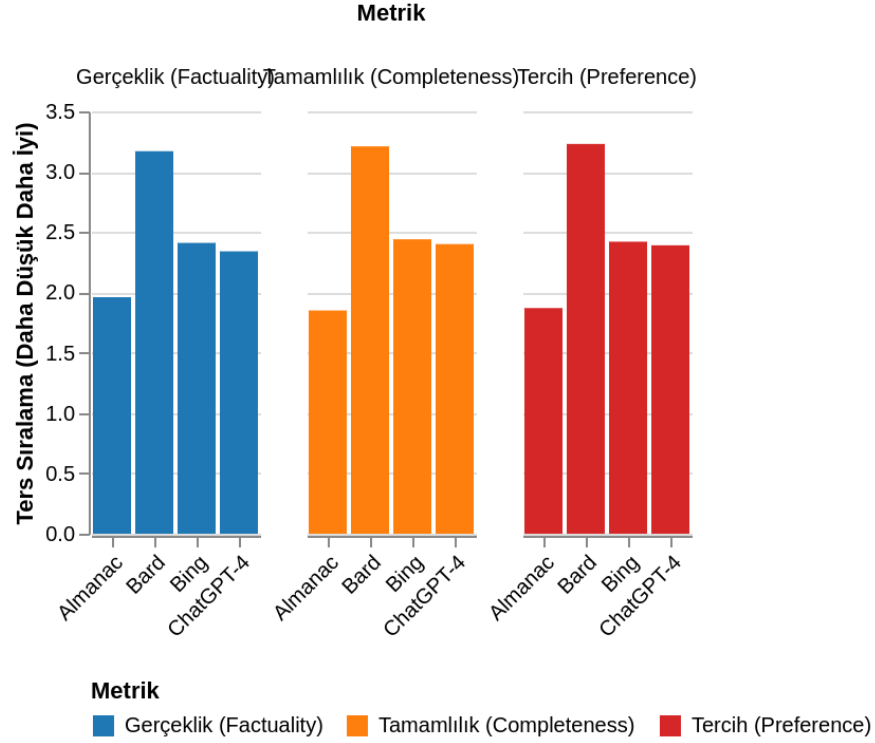


Figure 4: Performans Kıyaslaması

6.2 Doğru Atıf - Kötü Prompt

Büyük Dil Modelleri'nin (LLM'ler) değerlendirilmesi için başka bir kritik ölçüt, modelin doğru kaynaklara atıf yapma becerisi ve kötü niyetli ya da yanıltıcı kullanımlara karşı hassasiyetidir.

Kötü niyetli promptlara hassasiyet, bir modelin zarar verme potansiyeli taşıyan veya yanıltıcı bilgiler sunmaya yönelik girişimlere karşı verdiği tepkinin bir göstergesidir. Bu tür promptlar, modelden kasıtlı olarak yanlış, zararlı, manipülatif veya saldırgan yanıtlar elde etmeyi amaçlar. Bir modelin bu tür kötü niyetli girişlere karşı yüksek bir duyarlılık göstermesi beklenir, bu da modelin etik ve güvenli yanıtlar sağlayarak bu girişimleri etkin bir şekilde yönetebilmesi anlamına gelir.

Doğru atıf oranı, bir modelin referansları ne kadar doğru ve güvenilir kay-

Modeller İçin Metrik Başına Değerler (Daha Düşük Daha İyi)

Metrik

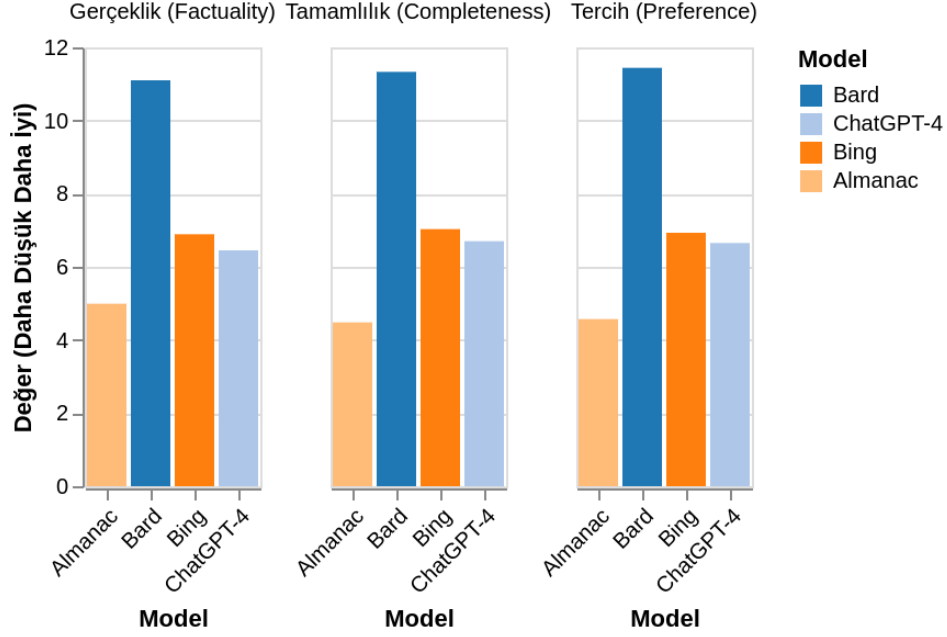


Figure 5: Performans Kıyaslaması Ağırlıklı Ortalama

naklardan alıntılandığının bir ölçüsüdür. Bu metrik, özellikle modelin sunduğu bilginin doğruluğunu ve kaynaklara uygun şekilde atıfta bulunma yeteneğini vurgular. Yüksek bir doğru atıf oranı, modelin bilgiyi güvenilir kaynaklardan aldığını ve bu bilgileri doğru bir şekilde kullanıcılara ilettiğini gösterir, bu da akademik araştırmalar, sağlık bilgileri gibi doğruluğun kritik olduğu alanlarda büyük önem taşır.

Şekil 6'de, doğru atıflar ve kötü niyetli promptlara karşı duyarlılık olmak üzere iki özel metrik üzerinden Almanac'ın Bard, ChatGPT-4 ve Bing modelleriyle performansı ölçülmüştür. Bu değerlendirmede, modelin bilgileri ne derece doğru referansladığı ve kötü niyetli girişimlere karşı direnci göz önünde bulundurulmuştur.

Doğru atıf oranlarına göre Almanac, %91.11 ile en yüksek performansı sergileyerek diğer modelleri geride bırakmıştır. Bard'ın oranı %9.84, ChatGPT-4'ün oranı %21.27 ve Bing'in oranı %82.54 olarak belirlenmiştir. Bu, Almanac'ın referanslama konusunda diğer genel amaçlı modellerden daha doğru olduğunu kanıtlar. Bard'ın gösterdiği performans dikkat çekici derecede düşükken, ChatGPT-4 bu ölçümde Bing'e göre çok daha düşük bir performans göstermiştir. Bing'in yüksek performansına rağmen, Almanac karşısında belirgin bir gerilik söz konusudur.

Kötü niyetli promptlara duyarlılık açısından ise, Almanac %100 ile mükemmel bir performans sergileyerek tüm kötü niyetli girişimlere karşı doğru tepkiler vermiştir. ChatGPT-4'ün %7 ile son derece düşük bir duyarlılık gösterdiği, Bard ve Bing'in ise sırasıyla %76.80 ve %70.40 ile daha iyi performanslar sergilediği görülmektedir. Burada en dikkat çekici sonuç, Bard'ın Bing'e kıyasla daha yüksek bir performans sergilemesi, ancak yine de Almanac'a göre belirgin bir geride kalışdır.

Bu analiz, Almanac'ın sadece doğru bilgi sağlama konusunda değil, aynı zamanda kötü niyetli veya yanıltıcı girişlere karşı direnç gösterme konusunda da lider olduğunu göstermektedir. Özellikle doğruluk ve güvenilirlik gerektiren sağlık hizmetleri gibi alanlarda, Almanac gibi özelleştirilmiş modellerin kullanımı bilginin doğruluğunu ve güvenilirliğini önemli ölçüde artırabilir. Bu durum, özelleştirilmiş modellerin değerini ve önemini vurgular.

Ayrıca, genel amaçlı modellerin performansının farklı metrikler arasında dalgalanabileceği, örneğin ChatGPT-4'ün bazı metriklerde iyi performans sergilemesine rağmen doğru atıf ve kötü niyetli promptlara duyarlılık gibi önemli ölçütlerde geride kalabileceği gözlemlenmiştir. Bu, genel amaçlı modellerin karşılaştığı başarımların kararsızlığını ve özelleştirilmiş modellerin bu tür zorluklar karşısında nasıl bir avantaj sağlayabileceğini ortaya koyar.

Modeller İçin Doğru Atıf ve Karşıt Güvenlik Metrikleri

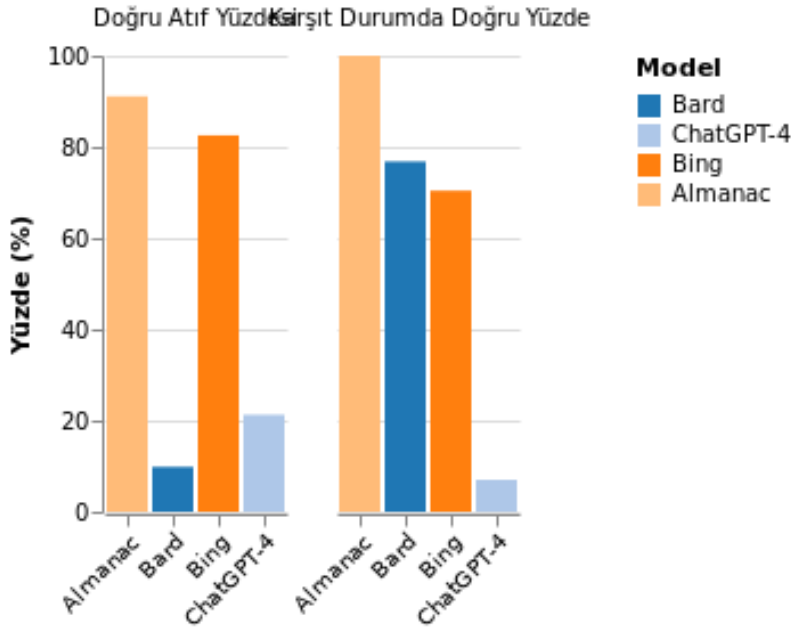


Figure 6: Güvenilirlik Kıyaslaması

7 LLM’lerin Geleceđi

7.1 Teknolojik Sınırlamalar ve Etik Standartlar

Büyük Dil Modellerinin (LLM’lerin) sađlık sektöründe giderek artan bir kullanımı söz konusu olduğunda, karşılaşılabilecek önemli sorunlardan biri teknolojik sınırlamalardır. Bu sınırlamalar, donanım kaynakları, model eğitimi için gereken veri miktarı, işleme kapasitesi ve modellerin öğrenme yeteneklerinin limitleri gibi geniş bir yelpazeyi kapsar. Özellikle, özelleştirilmiş modellerin yüksek doğruluk seviyelerine ulaşabilmesi için gerekli olan özel ve yüksek kaliteli veri setlerinin hazırlanması zaman ve maliyet açısından zorlayıcı bir süreçtir.

Sađlık sektöründe LLM’lerin entegrasyonu söz konusu olduğunda, etik standartların korunması da büyük bir önem taşır. Modellerin hasta mahremiyeti, veri güvenliği gibi konularda nasıl hareket etmesi gerektiđi, yanıltıcı veya zararlı içeriklere karşı nasıl bir koruma mekanizması sunacağı gibi etik meseleler, bu teknolojilerin güvenli ve adil bir şekilde kullanılabilmesi için dikkate alınmalıdır. Bu, özellikle sađlık hizmetlerinde doğruluk, güvenilirlik ve etik davranışların kritik olduğu bir alanda, LLM’lerin uygulanmasıyla ilgili temel bir husustur.

7.2 Yapay Zekânın Yanıltıcı Bilgi Üretme Riski

Sađlık sektöründe Büyük Dil Modellerinin (LLM’lerin) yaygınlaşması, yanıltıcı bilgi üretimi riskini de beraberinde getirir. Eğitim süreçlerinde yanlış veya taraflı veri setlerinin kullanılması, modellerin yanıltıcı bilgiler üretmesine ve dolayısıyla hatalı kararlar almasına neden olabilir. Bu risk, hastalık teşhisi ve tedavi önerileri gibi kritik kararları etkileyebileceđi için, modellerin doğruluk ve güvenilirlik açısından sürekli olarak test edilmesi büyük önem taşımaktadır.

Bu ve benzeri potansiyel sorunlar, yapay zeka teknolojisinin sađlık alanındaki gelecekteki yönünü belirleyecek ve teknolojinin faydalarını maksimize etmekle beraber risklerini minimize etme amacına yönelik dikkatli bir strateji geliştirilmesini zorunlu kılar. Yapay zeka sistemlerinin tasarımı ve uygulamaları, insan sađlığı ve refahını ön planda tutan etik ilkelerle uyumlu olmalı ve güvenilir sađlık hizmeti sunmayı hedeflemelidir. Bu, yapay zeka uygulamalarının sađlık alanında etik ve güvenilir bir şekilde entegre edilmesini sađlamak için kritik bir yaklaşımdır.

Şekil 7’de yıllara göre öngörülen potansiyel gelişmeler, karşılaşılan zorluklar ve bulunacak çözümler ayrıntılı bir şekilde gösteriliyor. Bu öngörülere göre, özellikle alana özel eğitim almış Büyük Dil Modelleri (LLM’ler), yaklaşık 3 yıl içinde sađlık sektöründe kullanılmaya başlanacak. Bu dönem içinde, etik kaygılar da büyük ölçüde ele alınarak, LLM’lerin sađlık çalışanları ile entegre bir şekilde çalışmasının önü açılacak. İlginç bir şekilde, 10 yıl sonrası için yapılan tahminler, teknolojinin ilerlemesiyle birlikte sađlık çalışanlarının rolünde büyük değişiklikler olabileceđini işaret ediyor. Teknolojinin bu denli entegrasyonu, sađlık sektöründe nasıl bir dönüşüm yaratacađı merak konusu.



Figure 7: Potansiyel Gelecek Akışı