

# Makale Raporu

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

Nisan 2025

## Makale Adı

LLM-MedQA: Enhancing Medical Question Answering through Case Studies in Large Language Models

## 1 Büyük Dil Modelleri (LLM)

Büyük Dil Modelleri (Large Language Models, LLM), yapay zeka alanında devrim yaratan, milyarlarca parametre içeren ve doğal dil anlama ve üretme yeteneklerine sahip derin öğrenme sistemleridir. Bu modeller, insanlarla gerçeğe yakın etkileşimler kurabilme, karmaşık metinleri anlayabilme ve üretebilme kapasiteleriyle öne çıkmaktadır. Bu bölümde, LLM'lerin tanımı ve temel kavramları, tarihsel gelişimi ve yapay zeka ekosistemindeki rollerini inceleyeceğiz.

### 1.1 LLM'lerin tanımı ve temel kavramlar

Büyük Dil Modelleri (LLM'ler), insan dilini anlama ve üretme kapasitesine sahip, milyarlarca parametreye sahip yapay sinir ağlarıdır. Bu modeller, büyük ölçekli metin verilerinde eğitilerek, doğal dildeki karmaşık kalıpları ve ilişkileri öğrenirler. LLM'lerin çalışma prensipleri şu temel kavramlara dayanır:

#### 1.1.1 Transformer Mimarisi

LLM'lerin temelinde, 2017 yılında tanıtılan "Attention is All You Need" makalesindeki Transformer mimarisi bulunmaktadır. Bu mimari, önceki dizi-işleme modellerinin aksine, paralel işleme yeteneği ve uzun mesafeli bağımlılıkları daha etkili yakalama özelliğiyle öne çıkar. Transformer'lar, öz-dikkat (self-attention) mekanizması vesilesiyle bir metindeki kelimelerin birbirleriyle olan alakaları doğrudan modelleyebilir.

#### 1.1.2 Tokenizasyon

Tokenizasyon, metin verilerinin model tarafından anlaşılabilir birimler olan "token"lara dönüştürülmesi işlemidir. Her token, bir kelime, kelime parçası veya

karakteri temsil edebilir. Örneğin, "tokenizasyon" kelimesi "token", "iza", "syon" gibi alt parçalara ayrılabilir. LLM'ler, metni önce bu tokenlar halinde işler, ardından her tokeni bir vektöre dönüştürür ve bu vektörler model içinde işlenir.

### 1.1.3 Gömmeler (Embeddings)

Tokenlerin sayısal vektörlere dönüştürülmesi, LLM'lerin metinle çalışabilmesi için öneme hâiz bir adımdır. Bu vektörler, kelimelerin veya alt-kelime birimlerinin anlamsal özelliklerini çok boyutlu bir uzayda temsil eder. Benzer anlamlı kelimeler, bu vektör uzayında birbirine yakın konumlanır.

### 1.1.4 Ön-eğitim ve İnce Ayar

LLM'ler tipik olarak iki aşamalı bir eğitim sürecinden geçer:

- **Ön-eğitim (Pre-training):** Model, internet, kitaplar ve diğer metin kaynaklarından toplanan devasa miktarda veriyle, dil yapısını ve genel bilgiyi öğrenmek için eğitilir.
- **İnce ayar (Fine-tuning):** Ön-eğitimli model, belirli görevlere veya özel alanlara (örneğin tıbbi metin analizi) adapte olmak için daha küçük ve özelleştirilmiş veri setleri üzerinde yeniden eğitilir.

### 1.1.5 Zero-shot ve Few-shot Öğrenme

Modern LLM'lerin en etkileyici özelliklerinden biri, hiç örnek görmeden (zero-shot) veya çok az örnek görerek (few-shot) yeni görevleri gerçekleştirebilme yeteneğidir. Mesela, bir tıbbi LLM, doğrudan tıbbi sorulara cevap verebilir veya semptomlar üzerinden hastalık tahmininde bulunabilir.

### 1.1.6 Bağlam Penceresi (Context Window)

Bağlam penceresi, modelin bir kerede işleyebileceği azamî token sayısını ifade eder. Bu, modelin "hafızasının" ne kadar uzağa gidebildiğini belirler. Günümüzde gelişmiş modeller, binlerce veya on binlerce token işleyebilmekte, bu da uzun metinleri ve karmaşık görevleri ele alabilmelerini sağlamaktadır.

## 1.2 Tarihsel gelişim ve önemli modeller

LLM'lerin gelişimi, doğal dil işleme alanında kademeli bir ilerleme ve bir dizi önemli atılımın sonucudur. Bu tarihsel gelişim süreci şöyle özetlenebilir:

### 1.2.1 Erken Dönem Dil Modelleri (2000'lerin başı)

İlk dil modelleri, n-gram tabanlı istatistiksel modellerdi. Bunlar, belirli bir kelime dizisinin olasılığını hesaplamak için önceki kelimelerin geçmiş sıklıklarını kullanırdı. Bu modeller basit ve hesaplama açısından verimli olsa da, uzun mesafeli dilbilgisel ilişkileri yakalamakta zorlanıyorlardı.

### 1.2.2 Sinir Ağı Tabanlı Dil Modelleri (2010-2017)

Tekrarlayan sinir ağları (RNN), uzun-kısa vadeli bellek ağları (LSTM) ve kapılı tekrarlayan birimler (GRU) gibi mimari yaklaşımlar, n-gram modellerinin sınırlamalarını aşmak için geliştirildi. Bu modeller metni sırayla işleyerek bağlamsal ilişkileri daha iyi yakalayabiliyordu, ancak uzun metinlerde bilgi kaybı yaşanmaktaydı.

### 1.2.3 Transformer Devri ve BERT (2017-2019)

2017’de tanıtılan Transformer mimarisi, DDİ alanında büyük bir dönüm noktası oldu. Google tarafından 2018’de geliştirilen BERT (Bidirectional Encoder Representations from Transformers), çift yönlü bağlam modelleme yeteneğiyle birçok DDİ görevinde çığır açtı.

### 1.2.4 GPT Ailesi (2018-2023)

OpenAI tarafından geliştirilen Generative Pre-trained Transformer (GPT) modelleri, dil modellemede yeni standartlar belirledi:

- **GPT-1 (2018):** 117 milyon parametre ile ilk büyük ölçekli ön-eğitilmiş dil modeli.
- **GPT-2 (2019):** 1.5 milyar parametre ile daha güçlü metin üretimi yeteneği.
- **GPT-3 (2020):** 175 milyar parametre ile zero-shot ve few-shot öğrenme yetenekleri.
- **GPT-4 (2023):** Çok daha büyük ve çok modlu yeteneklere sahip model.

### 1.2.5 Açık Kaynaklı Modeller ve Özel Modeller (2022-2025)

- **LLaMA Ailesi:** Meta’nın geliştirdiği LLaMA modelleri, açık kaynaklı ve verimli LLM’ler sunarak araştırma topluluğuna katkı sağladı. Makalede bahsedilen LLaMA3.1:70B, bu ailenin gelişmiş bir üyesidir.
- **Claude:** Anthropic tarafından geliştirilen ve güvenlik odaklı dil modeli.
- **Mistral, Gemma, Falcon:** Farklı kuruluşlar tarafından çeşitli boyutlarda ve yeteneklerde geliştirilen açık kaynaklı veya kısmen açık kaynaklı modeller.
- **Özelleştirilmiş Tıbbi LLM’ler:** Med-PaLM, MedLLaMA gibi özel olarak tıp alanı için uyarlanmış modeller geliştirilmiştir.

## 1.3 Yapay zeka ekosistemindeki rolleri

LLM’ler, yapay zeka ekosisteminde çok yönlü ve dönüştürücü bir rol oynamaktadır. Bu modellerin etkisi, birçok alanda hissedilmektedir:

### 1.3.1 Genel Doğal Dil İşleme Görevleri

LLM’ler, metin sınıflandırma, duygu analizi, isim varlık tanıma, metin özetleme, makine çevirisi gibi geleneksel DDİ görevlerinde yüksek performans göstermektedir. Tek bir büyük modelin farklı görevlerde uzmanlaşma yeteneği, alan özelinde çoklu model geliştirme ihtiyacını azaltmıştır.

### 1.3.2 İçerik Üretimi ve Yaratıcı Uygulamalar

LLM’ler, makale yazımı, şiir oluşturma, senaryo yazımı, reklam metni üretimi gibi yaratıcı içerik oluşturma görevlerinde kullanılmaktadır. Bu yetenekler, medya, pazarlama ve eğlence sektörlerinde yeni iş modelleri ve uygulamalar doğurmuştur.

### 1.3.3 Eğitim ve Öğretim Desteği

LLM’ler, öğrencilere kişiselleştirilmiş öğrenme deneyimleri sunma, karmaşık konuları açıklamaya, ödevlere yardımcı olma ve çeşitli ders materyalleri hazırlama gibi konularda eğitim sektörüne katkı sağlamaktadır.

### 1.3.4 Araştırma Hızlandırıcı

Bilimsel literatürü tarama, hipotez üretme, deney tasarımı önerme ve sonuçları yorumlama konularında araştırmacılara yardımcı olarak, bilimsel keşif süreçlerini hızlandırmaktadır.

### 1.3.5 Sağlık ve Tıp Alanındaki Uygulamalar

Makalede de vurgulandığı gibi, LLM’ler tıp alanında giderek daha önemli bir rol oynamaktadır. Özellikle:

- **Tıbbi Soru Cevaplama:** LLM-MedQA makalesinde belirtildiği gibi, özelleştirilmiş LLM’ler tıbbi sorulara cevap vermede kullanılmaktadır. LLaMA3.1:70B gibi modeller, çoklu-ajan mimari yapıları ile karmaşık tıbbi sorulara yanıt verebilmektedir.
- **Klinik Karar Destek Sistemleri:** Doktorlara tanı, tedavi ve prognoz konularında destek sağlayan sistemler geliştirilmektedir.
- **Tıbbi Literatür Analizi:** Büyük hacimli tıbbi literatürü tarayıp özetleyerek, klinisyenlere güncel bilgileri ulaştırma.
- **Hasta-Hekim İletişim Desteği:** Hasta şikayetlerini anlama ve hastaların anlayabileceği dile çevirme.

### 1.3.6 Ölçeklenebilir Kişisel Asistanlar

LLM’ler, akıllı asistanlar ve sanal ajanlara daha insan benzeri etkileşim yetenekleri kazandırmış, kullanıcıların daha karmaşık talimatları anlaşılır ve doğal bir şekilde iletmesini mümkün kılmıştır.

### 1.3.7 Kod Geliştirme ve Yazılım Mühendisliği

Programlama dillerinde kod yazma, hata ayıklama, kod açıklama ve refaktör etme gibi yazılım geliştirme görevlerinde programcılara destek olmaktadır. Bu, yazılım geliştirme süreçlerinin verimliliğini artırmaktadır.

### 1.3.8 Çok Modlu Sistemlere Entegrasyon

En son LLM'ler, metin, görüntü, ses ve video gibi farklı veri modalitelerini birleştiren çok modlu sistemlerin merkezinde yer almaktadır. Bu entegrasyon, daha zengin ve kapsamlı yapay zeka uygulamalarını mümkün kılmaktadır.

### 1.3.9 Zorluklar ve Etik Meseleler

LLM'lerin etkisi büyük olmakla birlikte, hallüsinasyon (doğru olmayan bilgiler üretme), önyargı, gizlilik sorunları ve güvenlik riskleri gibi önemli zorluklar da mevcuttur. Özellikle tıp gibi kritik alanlarda, LLM'lerin kararlarının doğruluğu ve güvenilirliği hayati önem taşımaktadır. LLM-MedQA makalesinde bahsedilen çoklu-ajan mimarisi ve vaka üretimi, bu zorlukların bazılarını ele alma girişimlerinden biridir.

## 2 LLM'lerin Teknik Altyapısı

### 2.1 Transformer mimarisi ve çalışma prensibi

Büyük Dil Modelleri'nin (LLM) temelinde, 2017 yılında tanıtılan "Attention is All You Need" makalesindeki Transformer mimarisi bulunmaktadır. Bu mimari, önceki dizi-işleme modellerinin aksine, paralel işleme yeteneği ve uzun mesafeli bağımlılıkları daha etkili yakalama özelliğiyle öne çıkar. Transformer'ların en önemli bileşeni, öz-dikkat (self-attention) mekanizmasıdır. Bu mekanizma sayesinde bir metindeki kelimelerin birbirleriyle olan ilişkileri doğrudan modellenilebilmekte ve uzun mesafeli dilbilgisel ilişkileri daha etkili şekilde yakalanabilmektedir. Öz-dikkat mekanizması, her token'ın diğer tüm token'larla olan ilişkisini hesaplayarak, metnin bağlamsal temsilini oluşturur.

### 2.2 Tokenizasyon

Tokenizasyon, metin verilerinin model tarafından anlaşılabilir birimler olan "token"lara dönüştürülmesi işlemidir. Her token, bir kelime, kelime parçası veya karakteri temsil edebilir. Örneğin, "tokenizasyon" kelimesi "token", "iza", "syon" gibi alt parçalara ayrılabilir. LLM'ler, metni önce bu tokenlar halinde işler, ardından her tokeni bir vektöre dönüştürür ve bu vektörler model içinde işlenir. Tokenizasyon, modelin kelime dağılımını belirler ve modelin işleyebileceği metin biçimini şekillendirir. Etkili bir tokenizasyon, dil modelinin performansını doğrudan etkileyebilir ve farklı diller için farklı stratejiler gerektirebilir.

## 2.3 Eğitim yöntemleri

LLM'ler tipik olarak iki aşamalı bir eğitim sürecinden geçer:

- **Ön-eğitim (Pre-training):** Model, internet, kitaplar ve diğer metin kaynaklarından toplanan devasa miktarda veriyle, dil yapısını ve genel bilgiyi öğrenmek için eğitilir. Bu aşamada model, dil yapısını ve genel dünya bilgisini kazanır.
- **İnce ayar (Fine-tuning):** Ön-eğitilmiş model, belirli görevlere veya özel alanlara (örneğin tıbbi metin analizi) adapte olmak için daha küçük ve özelleştirilmiş veri setleri üzerinde yeniden eğitilir.

Modern LLM'lerin en etkileyici özelliklerinden biri, hiç örnek görmeden (zero-shot) veya çok az örnek görerek (few-shot) yeni görevleri gerçekleştirebilme yeteneğidir. Bu, modellerin geniş ön-eğitim sürecinde kazandıkları bilgiyi farklı bağlamlara transfer edebilmelerini sağlar.

## 2.4 Model ölçeklendirme ve parametre sayısının etkileri

LLM'lerin gelişimi, model boyutunun ve parametre sayısının artmasıyla karakterize edilmiştir. GPT ailesi bu gelişimi açıkça göstermektedir:

- GPT-1 (2018): 117 milyon parametre
- GPT-2 (2019): 1.5 milyar parametre
- GPT-3 (2020): 175 milyar parametre
- GPT-4 (2023): Çok daha büyük ve çok modlu yeteneklere sahip

Parametre sayısının artması, modelin daha karmaşık dil kalıplarını öğrenebilmesini ve daha geniş bağlamları anlayabilmesini sağlar. Mesela, LLaMA3.1:70B gibi büyük modeller, karmaşık tıbbi sorulara başarıyla yanıt verebilmektedir. Ancak, model büyüklüğü arttıkça hesaplama gereksinimleri ve kaynak ihtiyacı da artmaktadır. Bu nedenle, model ölçeklendirme ile verimlilik arasında bir denge kurmak gerekmektedir. Bağlam penceresi (context window) de model ölçeklendirmenin önemli bir yönüdür ve modelin bir kerede işleyebileceği maksimum token sayısını belirler.

# 3 Sağlık ve Tıp Alanında LLM'ler

## 3.1 Mevcut uygulamalar ve kullanım alanları

Sağlık ve tıp alanında LLM'lerin kullanımı hızla yaygınlaşmaktadır. Makalelerde belirtildiği gibi, özellikle aşağıdaki alanlarda etkili uygulamalar geliştirilmektedir:

- **Tıbbi Soru Cevaplama (MedQA):** LLM-MedQA sistemleri ve LLaMA3.1:70B gibi özelleştirilmiş modeller, karmaşık tıbbi sorulara yanıt verebilmektedir. Bu sistemler, çoklu-ajan mimarisi kullanarak farklı tıbbi uzmanlık alanlarından gelen bilgileri sentezleyebilmekte ve doğru teşhis için klinik vakalar üretebilmektedir.
- **Klinik Karar Destek Sistemleri:** Hekimlere tanı, tedavi ve prognoz konularında destek sağlayan sistemler geliştirilmektedir. Bu sistemler, hekimlerin karar verme süreçlerini hızlandırmakta ve klinik verileri daha etkili bir şekilde değerlendirmelerini sağlamaktadır.
- **Tıbbi Literatür Analizi:** LLM'ler, büyük hacimli tıbbi literatürü tarayıp özetleyerek, klinisyenlere güncel bilgileri ulaştırmaktadır. Bu, sürekli artan tıbbi bilgi birikiminin takip edilmesini kolaylaştırmakta ve kanıta dayalı tıp uygulamalarını desteklemektedir.
- **Hasta-Hekim İletişim Desteği:** LLM'ler, hasta şikayetlerini anlama ve hastaların anlayabileceği dile çevirme konusunda yardımcı olmaktadır. Bu, hasta-hekim iletişimini iyileştirmekte ve tedavi uyumunu artırmaktadır.
- **Elektronik Dokümantasyon:** Hasta kayıtlarının elektronik ortamda dokümantasyonu ve idari görevlerin otomatikleştirilmesi, iş akışı verimliliğini artırmaktadır. LLM'ler, klinik notların oluşturulması, hastalık kodlaması ve tıbbi raporların hazırlanmasında kullanılmaktadır.

### 3.2 Fırsatlar ve potansiyel faydalar

LLM'lerin sağlık ve tıp alanında sunduğu fırsatlar ve potansiyel faydalar şunlardır:

- **Teşhis Doğruluğunda İyileşme:** LLM-MedQA makalesinde gösterildiği gibi, doğru yapılandırılmış LLM sistemleri, teşhis doğruluğunu artırabilmektedir. Özellikle çoklu-ajan mimarisi, karmaşık tıbbi problemlerde
- **Kişiselleştirilmiş Tıp:** LLM'ler, hastaya özgü faktörleri dikkate alarak kişiselleştirilmiş tedavi önerileri sunabilmektedir. Bu, her hastanın bireysel ihtiyaçlarına göre optimize edilmiş sağlık hizmetleri anlamına gelmektedir.
- **Sağlık Profesyonelleri İçin Zaman Tasarrufu:** Literatür tarama, dokümantasyon ve rutin görevlerin otomatikleştirilmesi, sağlık çalışanlarının hastalarına daha fazla zaman ayırabilmelerini sağlamaktadır.
- **Tıbbi Bilgiye Erişimin İyileştirilmesi:** LLM'ler, tıbbi bilgiye erişimi demokratikleştirerek, hem sağlık profesyonelleri hem de hastalar için bilgilerin daha anlaşılabilir ve erişilebilir olmasını sağlamaktadır.
- **Az Hizmet Alan Bölgelere Destek:** Uzman sağlık personelinin sınırlı olduğu bölgelerde, LLM tabanlı sistemler temel sağlık hizmetlerine erişimi artırabilir ve uzaktan konsültasyon imkanları sunabilir.

### 3.3 Zorluklar ve sınırlamalar

LLM’lerin sađlık ve tıp alanındaki kullanımında karşılaşılan başlıca zorluklar ve sınırlamalar şunlardır:

- **Alana Özgü Terminoloji Zorlukları:** LLM-MedQA makalesinde belirtildiđi gibi, tıbbi terminolojinin karmaşıklığı ve uzmanlık gerektirmesi, modellerin dođru anlama ve yanıt üretme yeteneklerini sınırlandırmaktadır.
- **Tıbbi Akıl Yürütme Gereksinimleri:** Tıbbi tanı ve tedavi kararları, karmaşık ve çok aşamalı akıl yürütme gerektirir. LLM’lerin bu tür akıl yürütme süreçlerinde dođru sonuçlara ulaşması her zaman mümkün olmamaktadır.
- **Halüsinasyon ve Faktüel Hatalar:** LLM’ler bazen var olmayan bilgileri üretebilmekte veya yanlış tıbbi bilgiler sunabilmektedir. Bu durum, özellikle hayati önem taşıyan tıbbi kararlarda ciddi risklere yol açabilir.
- **Bađlam Anlama Sınırlamaları:** Tıbbi vakaların karmaşık bađlamını tam olarak anlamak, mevcut LLM’ler için hala zorlayıcı bir görevdir. Hastaya özgü faktörlerin tam olarak değeriendirilememesi, yanlış teşhis ve tedavi önerilerine yol açabilir.
- **Mevcut Sađlık Sistemlerine Entegrasyon:** LLM tabanlı sistemlerin mevcut elektronik sađlık kayıt sistemleri ve klinik iş akışlarına entegrasyonu, teknik ve organizasyonel zorluklar içermektedir.

### 3.4 Etik hususlar ve hasta gizliliđi

LLM’lerin sađlık alanında kullanımı, önemli etik hususları ve hasta gizliliđi konularını gündeme getirmektedir:

- **Veri Gizliliđi ve Güvenliđi:** Hasta verilerinin LLM eğitimi ve kullanımı için toplanması ve işlenmesi, gizlilik ve güvenlik risklerini artırmaktadır. Hassas sađlık bilgilerinin korunması, yasal ve etik bir zorunluluktur.
- **Sorumluluk ve Hesap Verebilirlik:** LLM tabanlı sistemlerin tıbbi kararlarından kimin sorumlu olacağı (geliştirici, hekim veya sađlık kurumu) belirsizlik taşımaktadır. Yanlış teşhis veya tedavi önerilerinden kaynaklanan zararların sorumluluğunun belirlenmesi gereklidir.
- **Şeffaflık ve Açıklanabilirlik:** LLM’lerin nasıl sonuçlara ulaştıklarını açıklama zorluğu, tıbbi karar süreçlerinde şeffaflığı sınırlandırmaktadır. LLM-MedQA makalesinde önerilen vaka üretimi gibi yaklaşımlar, açıklanabilirliği artırmaya yönelik çabalardır.
- **Önyargı ve Adalet:** LLM’lerin eğitim verilerindeki mevcut önyargılar, belirli hasta gruplarına yönelik ayrımcı veya yetersiz öneriler sunma riskini taşımaktadır. Bu durum, sađlık eşitsizliklerini derinleştirebilir.



- **Düzenleyici Uyum:** LLM tabanlı sağlık uygulamaları, farklı ülkelerdeki sağlık düzenlemelerine, veri koruma yasalarına ve etik standartlara uyum sağlamalıdır. Bu düzenlemeler ülkeden ülkeye değişiklik gösterebilmekte ve uluslararası uygulamaları zorlaştırabilmektedir.

## 4 Tıbbi Soru Cevaplama (MedQA) Sistemleri

### 4.1 Tanım ve önemi

Tıbbi Soru Cevaplama (MedQA) sistemleri, tıbbi alanlardaki sorulara doğru ve güvenilir cevaplar sağlamak için tasarlanmış özelleştirilmiş doğal dil işleme sistemleridir. LLM-MedQA makalesinde belirtildiği gibi, "doğru ve verimli soru cevaplama sistemleri, tıp alanında yüksek kaliteli hasta bakımının sağlanması için kritik öneme sahiptir." MedQA sistemleri, sağlık profesyonellerinin karmaşık tıbbi sorulara hızla yanıt bulmasına, tıbbi literatürün analizine ve klinik karar verme süreçlerine destek olmasına yardımcı olur.

Tıbbi soru cevaplama sistemlerinin önemi şu noktalarda ortaya çıkmaktadır:

- **Klinik Karar Verme Desteği:** Doktorlar ve sağlık profesyonelleri, teşhis ve tedavi kararları alırken kritik bilgilere hızlıca erişebilir.
- **Tıbbi Bilgiye Erişimin Demokratikleştirilmesi:** Tıbbi bilginin karmaşıklığı ve hacmi göz önüne alındığında, MedQA sistemleri geniş tıbbi bilgi tabanını daha erişilebilir hale getirir.
- **Uzman Kaynaklarının Optimizasyonu:** Rutin soruların otomatikleştirilmesi, uzman hekimlerin daha karmaşık vakalara odaklanmasını sağlar.
- **Tıbbi Eğitim Desteği:** Tıp öğrencileri ve sağlık personelinin eğitiminde yardımcı araç olarak kullanılabilir.

### 4.2 Geleneksel yaklaşımlar

Tıbbi soru cevaplama sistemlerinin gelişiminde, LLM'lerin ortaya çıkışından önce çeşitli geleneksel yaklaşımlar kullanılmıştır. Bu yaklaşımlar genellikle şu özelliklere sahiptir:

- **Kural Tabanlı Sistemler:** Önceden tanımlanmış kurallar ve hiyerarşik karar ağaçları kullanılarak tıbbi bilgi modellenmiştir. Bu sistemler genellikle belirli tıbbi alanlarda iyi performans gösterse de, esnek olmayan yapıları nedeniyle kompleks sorguları anlamakta zorlanmışlardır.
- **İstatistiksel Modeller:** Olasılık tabanlı modeller ve makine öğrenimi teknikleri (SVM, Karar Ağaçları, vb.), tıbbi metinleri işlemek ve soru-cevap eşleştirmesi yapmak için kullanılmıştır.
- **Bilgi Çıkarımı Yaklaşımları:** Tıbbi literatürden bilgi çıkarımı yaparak yapılandırılmış bilgi tabanları oluşturulmuş ve sorgular bu veritabanlarında aranmıştır.

- **Ontoloji Tabanlı Yöntemler:** Tıbbi ontolojiler (SNOMED CT, UMLS gibi) kullanılarak kavramlar arasındaki ilişkiler modellenmiş ve sorular bu ilişkilere göre cevaplanmıştır.
- **Bilgi Geri Getirimi Sistemleri:** Anahtar kelime tabanlı veya vektör uzayı modelleri kullanılarak tıbbi dokümanlarda arama yapılmış ve ilgili bilgiler kullanıcıya sunulmuştur.

Bu geleneksel yaklaşımların ortak sınırlamaları arasında, doğal dil varyasyonlarını anlamada zorluk yaşamaları, tıbbi terminolojinin karmaşıklığını tam olarak kavrayamamaları ve bağlamsal anlama yeteneklerinin sınırlı olması sayılabilir.

### 4.3 LLM tabanlı modern yaklaşımlar

LLM'lerin gelişimiyle birlikte, tıbbi soru cevaplama sistemleri önemli bir dönüşüm geçirmiştir. LLM-MedQA makalesinde vurgulandığı gibi, modern yaklaşımlar şu özelliklere sahiptir:

- **Büyük Dil Modeli Entegrasyonu:** LLaMA3.1:70B gibi milyarlarca parametreye sahip büyük dil modelleri, geniş tıbbi bilgi içeren metinlerden ön-eğitim alarak tıbbi sorgulara cevap verebilmektedir.
- **Çoklu-Ajan Mimarisi:** Makalede önerilen çoklu-ajan yaklaşımında, farklı uzmanlık alanlarından (soru uzmanları, seçenek uzmanları) ajanlar işbirliği yaparak karmaşık tıbbi soruları yanıtlamaktadır. Bu mimaride her ajan, problemin farklı yönlerine odaklanarak kapsamlı bir analiz sağlar.
- **Vaka Üretimi Bileşeni:** LLM-MedQA'nın en yenilikçi özelliklerinden biri, verilen soruya ve seçeneklere dayalı olarak otomatik klinik vaka üretimi yapabilmesidir. Bu vakalar, sistemin tahminlerini destekleyen ve açıklayan gerçekçi senaryolar sunmaktadır.
- **Zero-shot ve Few-shot Öğrenme:** Modern MedQA sistemleri, ek eğitim verileri gerektirmeden (zero-shot) veya çok az örnek görerek (few-shot) tıbbi sorulara cevap verebilme yeteneğine sahiptir.
- **Düşünce Zinciri (Chain-of-Thought):** LLM'lerin adım adım akıl yürütme yeteneklerini kullanarak, tıbbi sorulara verilen yanıtların gerekçelendirilmesini sağlayan yaklaşımlar geliştirilmiştir.
- **Kendini-Tutarlılık (Self-Consistency):** Farklı akıl yürütme yolları üretilip en tutarlı cevabı seçerek, modelin güvenilirliğini artıran yöntemler kullanılmaktadır.

LLM-MedQA makalesine göre, bu modern yaklaşımlar geleneksel yöntemlere kıyasla doğruluk ve F1 skorunda

#### 4.4 Performans deęerlendirme metrikleri

Tıbbi soru cevaplama sistemlerinin performansı, doęruluk, gvenilirlik ve yorumlanabilirlik gibi eřitli faktrlere dayanarak deęerlendirilir. LLM-MedQA makalesinde ve performans deęerlendirme literatrnde yaygın olarak kullanılan metrikler řunlardır:

- **Doęruluk (Accuracy):** Sistemin doęru yanıtladıęı soruların toplam soru sayısına oranını gsteren temel metrik. LLM-MedQA’da, multi-ajan yaklařımıyla %77.2 doęruluk elde edilmiřtir.
- **Makro-Kesinlik (Macro Precision):** Tm sınıflar zerinden kesinlik skorlarının ortalaması. Bu metrik, sistemin yanlış pozitif sonuları ne kadar iyi nledięini gsterir.
- **Makro-Duyarlılık (Macro Recall):** Tm sınıflar zerinden duyarlılık skorlarının ortalaması. Bu metrik, sistemin doęru yanıtları ne kadar iyi yakaladıęını gsterir.
- **Makro F1-Skoru:** Kesinlik ve duyarlılık arasında bir denge saęlayan harmonik ortalama. F1 skoru, zellikle dengesiz veri setlerinde performansı deęerlendirmek iin nemlidir.
- **Yorumlanabilirlik (Interpretability):** Sistemin verdięi yanıtların gerekesini ne kadar iyi aıklayabildięini ler. Modern MedQA sistemlerinde, zellikle oklu-ajan mimari ve vaka retimi gibi bileřenler, yorumlanabilirlięi artırmak iin kullanılmaktadır.
- **Gvenilirlik (Reliability):** Sistemin tutarlı ve gvenilir yanıtlar retme yeteneęi. Bu, zellikle tıbbi karar vermede kritik neme sahiptir.
- **Klinik Alaka Dzeyi (Clinical Relevance):** Verilen yanıtların klinik uygulamada ne kadar kullanışlı olduęunu deęerlendirir. Bu genellikle uzman deęerlendirmesi gerektirir.

MedQA sistemlerinin deęerlendirilmesinde, sayısal metriklerle birlikte, sistemin klinik ortamlarda kullanılabilirlięi ve tıbbi uzmanlara gerek deęer katıp katmadıęı da dikkate alınmalıdır.

### 5 LLM-MedQA Makalesinin Analizi

#### 5.1 Ana hedefler ve katkılar

LLM-MedQA makalesi, tıbbi soru cevaplama sistemlerinin etkinlięini artırmak iin yeniliki bir yaklařım sunmaktadır. alıřmanın ana hedefleri ve katkıları řunlardır:

- **Temel Amaç:** Makale, "tıp alanında yüksek kaliteli hasta bakımının sağlanması için doğru ve verimli soru cevaplama sistemlerinin kritik öneme sahip olduğunu" vurgulayarak, LLM'lerin tıbbi terminoloji anlama ve karmaşık tıbbi akıl yürütme zorluklarını aşmayı hedeflemektedir.
- **Çoklu-Ajan Mimarisi:** Araştırmacılar, tıbbi soruları çözmek için LLaMA3.1:70B modelini kullanan yeni bir çoklu-ajan çerçevesi geliştirmiştir. Bu yapı, farklı tıp uzmanlık alanlarından bilgiyi entegre etmek için tasarlanmıştır.
- **Vaka Üretimi Bileşeni:** Makalenin sunduğu en önemli yeniliklerden biri, soruları ve cevap seçeneklerini destekleyecek klinik vakalar üreten bir modül geliştirmektir. Bu, sistemin yorumlanabilirliğini ve güvenilirliğini artırmaktadır.
- **Performans İyileştirmesi:** Geliştirilen sistem, MedQA veri seti üzerinde mevcut kıyaslama modellerine kıyasla doğruluk ve F1 skorunda %7'lik bir iyileşme sağlamıştır.
- **Zero-Shot Öğrenme:** Yaklaşım, ek eğitim verileri gerektirmeden, modelin doğal tıbbi bilgi ve akıl yürütme yeteneklerinden yararlanmaktadır.

## 5.2 Çok ajanlı mimari yapısı

LLM-MedQA'nın çok ajanlı mimarisi, altı ana bileşenden oluşmaktadır:

- **Ajan Üretimi:** Bu ilk aşamada, sistem iki tür uzman üretir:
  - **Soru Uzmanları:** Pulmonoloji, Endokrinoloji, Acil Tıp gibi belirli tıbbi alanlarda uzmanlaşmış ve klinik problemleri analiz edebilen ajanlardır.
  - **Seçenek Uzmanları:** Her bir cevap seçeneğinin geçerliliğini ve ilgisini değerlendiren uzmanlardır.
- **Önerme Analizi:** Bu aşamada uzmanlar, sorunu ve seçenekleri ayrıntılı olarak inceler:
  - Soru uzmanları, belirtileri, olası tanıları ve kritik özellikleri tanımlayarak yapılandırılmış bir analiz sunar.
  - Seçenek uzmanları, her seçeneği bağımsız olarak değerlendirir ve bunların klinik duruma uygunluğunu tartışır.
- **Vaka Üretimi:** Sistem, soruya ve seçeneklere dayalı olarak destekleyici klinik vakalar üretir.
- **Rapor Özeti:** Tüm analizleri ve üretilen vakaları tutarlı bir rapora entegre eder. Bu rapor iki ana bölümden oluşur:
  - **Toplam Analiz:** Klinik senaryonun kapsamlı bir sentezi
  - **Anahtar Bilgiler:** En önemli tanısal ipuçları ve klinik bağlam

- **Oylama Mekanizması:** Tüm uzmanlar üretilen rapor üzerinde oy kullanır ("Evet" veya "Hayır"). Herhangi bir uzman "Hayır" oyu verirse, düzeltme önerileri toplanır ve rapor yeniden düzenlenerek tekrar oylamaya sunulur.
- **Karar Verme:** Oybirliğiyle onaylanan nihai rapor, doğru cevabın seçilmesi için kullanılır.

Bu mimari, farklı tıbbi uzmanlık alanlarından gelen bilgileri sentezleyerek, ortak bir klinik akıl yürütme sürecini simüle etmektedir.

### 5.3 Vaka üretimi bileşeni ve önemi

Vaka üretimi bileşeni, LLM-MedQA sisteminin en yenilikçi özelliklerinden biridir. Bu bileşen, verilen tıbbi soruya ve seçeneklere dayalı olarak gerçekçi klinik vakalar üretir, sistemin kararlarını destekler ve açıklanabilirliğini artırır.

- **Vaka Üretim Yapısı:** Her üretilen vaka üç temel bileşenden oluşmaktadır:
  - **Bağlam (Context):** Belirtiler, tıbbi geçmiş ve tanısal bulgular gibi ayrıntılı bir klinik senaryo sunar.
  - **Anahtar Mekanizma/Akıl Yürütme:** Klinik bulgular ile doğru tanı arasındaki bağlantıyı açıklayarak, seçilen seçeneğin neden doğru olduğunu gerekçelendirir.
  - **Tarafsızlık Kontrolü:** Seçilen seçenek hakkında abartılı iddialardan kaçınarak ve uygun olduğunda ilgili alternatifleri kısaca tanıyarak objektifliği sağlar.
- **Önem ve Katkıları:**
  - **Yorumlanabilirliği Artırma:** Üretilen vakalar, soyut tıbbi kavramları somut klinik senaryolarla açıklayarak sistemin kararlarını daha anlaşılır hale getirir.
  - **Güvenilirliği Güçlendirme:** Gerçekçi klinik vakalar, sistemin önerilerini destekleyen kanıtlar sunar.
  - **Performans İyileştirmesi:** Ablasyon çalışmasında görüldüğü üzere, vaka üretimi modülü doğruluk oranını %1-2 artırmaktadır.
  - **Klinik Alaka:** Teorik tıbbi bilgi ile pratik klinik uygulama arasında köprü kurarak, önerileri daha ilgili hale getirir.
- **Entegrasyon:** Üretilen vakalar, nihai raporda diğer analizlerle birleştirilerek, bütünsel bir klinik değerlendirme sağlar. Makalede belirtildiği gibi, "üretilen vakalar bağımsız çıktılar değil, problem ve seçenek uzmanlarının analizleriyle sinerjik olarak çalışan" bileşenlerdir.

## 5.4 Deneysel metodoloji ve kullanılan veri kümeleri

LLM-MedQA çalışmasında kullanılan metodoloji ve veri kümeleri şu şekilde özetlenebilir:

- **Veri Kümesi:** Çalışmada, tıbbi çoktan seçmeli sorular içeren MedQA veri kümesi kullanılmıştır. Her örnek, bir klinik soru, beş cevap seçeneği ve doğrulama için doğru cevap içermektedir. Makalede, MedQA veri kümesinin "tıbbi bilginin özelleşmiş doğası ve gerekli akıl yürütmenin karmaşıklığı nedeniyle benzersiz zorluklar" sunduğu belirtilmektedir.
- **Model Seçimi:** Araştırmacılar, açık kaynaklı olması, nispeten düşük hesaplama gereksinimleri ve maliyet etkinliği nedeniyle LLaMA3.1:70B modelini seçmiştir.
- **Deney Düzenlemesi:**
  - Tüm deneyler zero-shot (örneksiz) ayarda gerçekleştirilmiştir.
  - Veri kümesinden rastgele 300 örnek üç kez seçilmiştir.
  - Sonuçlar, üç çalıştırmanın ortalaması olarak raporlanmıştır.
  - Çıkarım parametreleri: sıcaklık=0, frekans cezası=0, mevcudiyet cezası=0, top p=1
  - Her örnek için çıkarım süresi yaklaşık 1.5 dakikadır.
- **Değerlendirme Metrikleri:**
  - **Doğruluk (Accuracy):** Doğru tahminlerin genel oranı.
  - **Makro Kesinlik (Macro Precision):** Tüm sınıflar üzerinden kesinlik skorlarının ortalaması.
  - **Makro Duyarlılık (Macro Recall):** Tüm sınıflar üzerinden duyarlılık skorlarının ortalaması.
  - **Makro F1-Skoru:** Tüm sınıflar üzerinden F1 skorlarının ortalaması.
- **Karşılaştırma Yaklaşımları:**
  - **Doğrudan Çıkarım:** Soruyu ve seçenekleri doğrudan LLM'ye verme.
  - **Düşünce Zinciri (CoT):** Adım adım akıl yürütme yaparak cevaba ulaşma.
  - **CoT+Kendini-Tutarlılık (SC):** Birden fazla akıl yürütme yolu üreterek en tutarlı cevabı seçme.
- **Ablasyon Çalışması:** Farklı model boyutları (8B vs 70B parametreler) ve vaka üretimi bileşeninin etkisi değerlendirilmiştir. Sonuçlar, hem daha büyük model boyutunun hem de vaka üretiminin performansı iyileştirdiğini göstermiştir.

Deneyler, önerilen LLM-MedQA sisteminin tüm karşılaştırma yaklaşımlarından daha iyi performans gösterdiğini, yaklaşık %77 doğruluğa ulaştığını (diğer yöntemlerin %70 civarında olduğu) ve hem doğrulukta hem de F1 skorunda %7'lik bir iyileştirme sağladığını göstermiştir.

## 6 LLM-MedQA Yaklaşımının Sonuçları ve Değerlendirmesi

### 6.1 DeneySEL sonuçlar ve başarıml ölçütleri

### 6.2 Diğer yaklaşımlarla karşılaştırma

### 6.3 Güçlü yanlar ve yenilikçi özellikler

### 6.4 LLM-MedQA Yaklaşımının Sonuçları ve Değerlendirmesi

### 6.5 DeneySEL sonuçlar ve başarıml ölçütleri

LLM-MedQA sistemi, MedQA veri seti üzerinde gerçekleştirilen kapsamlı deneylerde etkileyici sonuçlar elde etmiştir. Değerlendirme için kullanılan temel metrikler şunlardır:

- **Doğruluk (Accuracy):** LLM-MedQA yaklaşımı, test edilen örneklerde yaklaşık %77.2 doğruluk oranına ulaşmıştır. Bu oran, diğer karşılaştırma yaklaşımlarının %70 civarındaki değerlerinden belirgin şekilde yüksektir.
- **Makro Kesinlik (Macro Precision):** Sistem, tüm sınıflar üzerinden %77.1'lik bir makro kesinlik değeri elde etmiştir.
- **Makro Duyarlılık (Macro Recall):** Tüm sınıflar üzerinden duyarlılık skorlarının ortalaması %77.2 olarak ölçülmüştür.
- **Makro F1-Skoru:** Kesinlik ve duyarlılık arasındaki dengeyi gösteren F1 skoru %77.1 olarak gerçekleşmiştir.

Deney tasarımında önemli noktalar:

- Tüm deneyler zero-shot (örneksiz) ayarda gerçekleştirilmiş, yani model ek eğitim verileri olmadan doğrudan sorulara cevap vermiştir.
- Veri kümesinden rastgele 300 örnek üç kez seçilmiş ve sonuçlar bu üç çalıştırmanın ortalaması olarak raporlanmıştır.
- Çıkarım parametreleri optimize edilmiştir (sıcaklık=0, frekans cezası=0, mevcudiyet cezası=0, top p=1).
- Sistem, LLaMA3.1:70B modelini temel almıştır.

Ablasyon çalışması sonuçları, hem model boyutunun hem de vaka üretimi bileşeninin performans üzerinde önemli etkisi olduğunu göstermiştir. 70B parametrelili model, 8B parametrelili modele göre yaklaşık %20 daha iyi performans sergilemiştir. Ayrıca, vaka üretimi bileşeninin eklenmesi, 70B modelde %2'lik ek bir performans artışı sağlamıştır.

## 6.6 Diğer yaklaşımlarla karşılaştırma

LLM-MedQA sistemi, çeşitli karşılaştırma yaklaşımlarıyla değerlendirilmiş ve hepsinden daha üstün performans göstermiştir:

- **Doğrudan Çıkarım (Direct Inference):** Bu temel yaklaşım, soruyu ve seçenekleri doğrudan modele verip cevap almayı içerir. Zero-shot ayarda %71.4 doğruluk elde etmiştir.
- **Düşünce Zinciri (Chain of Thought, CoT):** Adım adım akıl yürütme yaparak cevaba ulaşan bu yaklaşım, zero-shot ayarda %69.8 doğruluk elde etmiştir.
- **CoT+Kendini-Tutarlılık (Self-Consistency, SC):** Birden fazla akıl yürütme yolu üreterek en tutarlı cevabı seçen bu yaklaşım, zero-shot ayarda %71.9 doğruluk sağlamıştır.
- **Few-Shot Yaklaşımlar:** Az sayıda örnek kullanılan few-shot öğrenme ayarları da test edilmiş, ancak bunlar da LLM-MedQA'dan daha düşük performans göstermiştir. En iyi few-shot sonucu, CoT+SC yaklaşımıyla %72.7 doğruluk olmuştur.

LLM-MedQA, hem doğrulukta hem de F1 skorunda diğer yaklaşımlara kıyasla yaklaşık %7'lik bir iyileşme sağlamıştır. Bu önemli fark, çoklu-ajan mimarisi ve vaka üretimi bileşenlerinin etkili entegrasyonundan kaynaklanmaktadır. Özellikle zero-shot ayarda elde edilen yüksek performans, sistemin ek eğitim verileri olmadan karmaşık tıbbi sorulara doğru yanıt verme yeteneğini göstermektedir.

## 6.7 Güçlü yanlar ve yenilikçi özellikler

LLM-MedQA yaklaşımının öne çıkan güçlü yanları ve yenilikçi özellikleri şunlardır:

- **Çoklu-Ajan Mimarisi:** Farklı tıp uzmanlık alanlarından ajanların işbirliği yapmasına olanak tanıyan bu mimari, tıbbi bilgiyi daha kapsamlı ve çok boyutlu bir şekilde değerlendirmektedir. Soru uzmanları ve seçenek uzmanları gibi farklı rollere sahip ajanlar, klinik problemleri farklı perspektiflerden analiz ederek daha bütünsel bir değerlendirme sağlamaktadır.
- **Vaka Üretimi Bileşeni:** Makalenin en yenilikçi özelliklerinden biri olan vaka üretimi modülü, soruya ve seçeneklere dayalı olarak gerçekçi klinik vakalar otomatik olarak üretmektedir. Bu vakalar üç bileşenden oluşur:
  - Bağlam (Context): Detaylı klinik senaryo
  - Anahtar Mekanizma/Akıl Yürütme: Bulgular ile tanı arasındaki bağlantının açıklaması
  - Tarafsızlık Kontrolü: Objektifliği sağlamaya yönelik kontrol



- **Oylama Mekanizması:** Sistem, uzmanlar arasında fikir birliği sağlamak için tasarlanmış bir oylama mekanizması kullanmaktadır. Bu mekanizma, herhangi bir uzmanın raporun bir yönüne itiraz etmesi durumunda düzeltici geri bildirimlere olanak tanır ve nihai rapor ancak tüm uzmanların onayı alındıktan sonra oluşturulur.
- **Yüksek Yorumlanabilirlik:** Sistem, sadece doğru cevabı seçmekle kalmaz, aynı zamanda kararın arkasındaki gerekçeyi açıklayan kapsamlı bir rapor üretir. Üretilen vakalar, soyut tıbbi kavramları somut klinik senaryolarla ilişkilendirerek, sistemin kararlarını daha anlaşılır ve güvenilir hale getirir.
- **Zero-Shot Öğrenme Yeteneği:** LLM-MedQA, ek eğitim verileri gerektirmeden karmaşık tıbbi sorulara cevap verebilmektedir. Bu, özellikle tıbbi veri etiketleniminin pahalı ve zaman alıcı olduğu durumlarda önemli bir avantaj sağlar.

## 6.8 Sınırlamalar ve iyileştirme alanları

LLM-MedQA yaklaşımının mevcut sınırlamaları ve potansiyel iyileştirme alanları şunlardır:

- **Hesaplama Gereksinimleri:** Çoklu-ajan mimarisi ve vaka üretimi bileşeni, geleneksel yaklaşımlara kıyasla daha fazla hesaplama kaynağı gerektirmektedir. Örnek başına çıkarım süresi yaklaşık 1.5 dakikadır, bu da gerçek zamanlı uygulamalarda bir kısıtlama olabilir. Sistemin hesaplama verimliliğini artırmak için optimizasyon çalışmaları yapılabilir.
- **Model Boyutuna Bağımlılık:** Ablasyon çalışması, sistemin performansının model boyutuna oldukça bağımlı olduğunu göstermiştir. 8B parametrelili model ile 70B parametrelili model arasında yaklaşık %20'lik bir performans farkı vardır. Bu durum, sistemin daha küçük modellerde etkili bir şekilde çalışmasını zorlaştırabilir.
- **Özelleştirilmiş Alan Bilgisi:** Mevcut sistemde, tıbbi alan bilgisi LLM'in ön-eğitim sürecinde edindiği genel bilgiye dayanmaktadır. Daha spesifik tıbbi uzmanlık alanlarında performansı artırmak için alan-özel bilgi tabanlarıyla entegrasyon yapılabilir.
- **Vaka Üretimi Çeşitliliği:** Üretilen vakaların çeşitliliği ve gerçekçiliği daha da geliştirilebilir. Farklı hasta profilleri, nadir hastalıklar ve karmaşık komorbiditeler gibi faktörleri daha iyi temsil eden vakalar üretilebilir.
- **Dinamik Uzman Seçimi:** Mevcut sistemde uzman ajanların seçimi statiktir. Sorunun karmaşıklığına ve tipine göre dinamik olarak uzman seçen daha adaptif bir yaklaşım geliştirilebilir.
- **Açıklanabilirlik:** Sistemin kararlarının gerekçelerini daha detaylı ve hekimlerin anlayabileceği şekilde sunmaya yönelik iyileştirmeler yapılabilir. Bu, klinik uygulama entegrasyonu için kritik öneme sahiptir.

- **Gerçek Dünya Değerlendirmesi:** Sistem şu ana kadar kontrollü bir veri seti üzerinde değerlendirilmiştir. Gerçek klinik ortamlarda ve daha geniş veri setleri üzerinde performans değerlendirmesi yapılması gerekmektedir.

Gelecekteki çalışmalarda, sistem daha geniş bir klinik senaryo yelpazesini destekleyecek şekilde genişletilebilir ve gerçek zamanlı tıbbi ortamlarda uygulanabilirliği artırılabilir. Ayrıca, farklı dillerde tıbbi soru cevaplama yeteneği geliştirilebilir ve çeşitli tıbbi uzmanlık alanlarına özel modüller eklenebilir.

## 7 Gelecek Yönelimler ve Uygulamalar

### 7.1 Sağlık için LLM teknolojilerindeki potansiyel ilerlemeler

Sağlık alanında kullanılan LLM teknolojilerinde gelecekte gerçekleşebilecek potansiyel ilerlemeler şunlardır:

- **Klinik Verilerle Ön-eğitim:** Gelecekteki LLM'ler, daha geniş ve kaliteli tıbbi veri setleri kullanılarak ön-eğitime tabi tutulabilir, böylece tıbbi terminoloji anlama ve tıbbi akıl yürütme yetenekleri geliştirilecektir.
- **Çok Modlu Tıbbi Modeller:** LLM'ler, metin verilerinin yanı sıra görüntü (radyoloji, patoloji), ses (hasta görüşmeleri) ve biyosensör verileri gibi çoklu veri modalitelerini entegre edebilecek şekilde geliştirilebilir. Bu, teşhis ve tedavi planlamasında daha bütünsel bir yaklaşım sunabilir.
- **Daha Gelişmiş Çoklu-Ajan Mimarileri:** LLM-MedQA makalesinde önerilen çoklu-ajan yapısı, farklı tıp alanlarından uzmanların bilgilerini daha etkin birleştirebilen ve kendi kendini düzeltebilen daha karmaşık sistemlere evrilebilir.
- **Bağlam Penceresi İyileştirmeleri:** Daha geniş bağlam pencerelerine sahip modeller, hastaların tüm tıbbi geçmişini ve karmaşık test sonuçlarını aynı anda analiz edebilecek, böylece daha kapsamlı klinik değerlendirmeler yapabilecektir.
- **Tıbbi Halüsinasyon Azaltma:** Özellikle tıp alanında kritik önem taşıyan faktüel doğruluğu artırmak için halüsinasyonları (gerçekte olmayan bilgiler ürete eğilimini) azaltma teknikleri geliştirilecektir.

### 7.2 Araştırma fırsatları

LLM'lerin sağlık ve tıp alanında kullanımı ile ilgili çeşitli araştırma fırsatları bulunmaktadır:

- **Klinik Karar Destek Mekanizmaları:** LLM-MedQA gibi sistemlerin gerçek klinik ortamlarda karar verme üzerindeki etkilerini inceleyen araştırmalar, bu teknolojinin hasta bakımını nasıl iyileştirdiğini veya değiştirdiğini değerlendirebilir.

- **Açıklanabilir LLM Yaklaşımları:** Tıbbi LLM'lerin kararlarını ve önerilerini hekimlere ve hastalara şeffaf ve anlaşılır şekilde açıklayabilen yöntemlerin geliştirilmesi, klinik güvenin artırılması için kritiktir.
- **Alan-Spesifik Tıbbi Dil Modellemesi:** Belirli tıbbi alt alanlara (kardiyoloji, onkoloji, psikiyatri vb.) özelleştirilmiş daha küçük ama daha hedefli modellerin geliştirilmesi, hem performans hem verimlilik açısından araştırılabilir.
- **Kompakt Tıbbi LLM'ler:** Klinik ortamlarda daha uygulanabilir olan, kaynak kullanımı optimize edilmiş ve daha küçük ancak özelleştirilmiş tıbbi LLM'lerin geliştirilmesi üzerine araştırmalar.
- **Kültürel ve Dil Adaptasyonu:** Farklı dillerde ve kültürel bağlamlarda etkili çalışabilen tıbbi LLM sistemleri, küresel sağlık hizmetlerindeki eşitsizliklerin azaltılmasına katkı sağlayabilir.

### 7.3 Pratik uygulama imkanları ve entegrasyon yolları

LLM tabanlı sağlık sistemlerinin pratik uygulaması ve mevcut sağlık sistemlerine entegrasyonu için:

- **Elektronik Sağlık Kaydı (ESK) Entegrasyonu:** LLM sistemlerinin mevcut ESK sistemleriyle entegrasyonu, hasta verileri üzerinde gerçek zamanlı analiz ve öneri sağlayabilir. Bu, klinik iş akışına minimum kesinti ile maksimum değer katabilir.
- **Mobil Sağlık Uygulamaları:** Taşınabilir cihazlarda çalışabilen hafif LLM-tabanlı sistemler, sağlık profesyonellerinin sahada hızlı klinik destek almasını sağlayabilir.
- **Hibrit İnsan-AI İş Akışları:** Hekimlerin ve LLM sistemlerinin güçlü yönlerini birleştiren iş akışları tasarlanabilir. Örneğin, sistem ön taramayı ve bilgi derlemeyi yaparken, nihai klinik kararlar hekimler tarafından verilebilir.
- **Tele-tıp Platformları:** LLM-MedQA gibi sistemler, uzaktan sağlık hizmetlerini destekleyerek, özellikle uzman erişiminin sınırlı olduğu kırsal veya az hizmet alan bölgelerde sağlık hizmetlerine erişimi iyileştirebilir.
- **Klinik Eğitim Simülatörleri:** Tıp öğrencileri ve asistanlar için, LLM-MedQA'nın vaka üretim bileşenine benzer şekilde gerçekçi klinik senaryolar oluşturabilen eğitim sistemleri geliştirilebilir.

Bu potansiyel ilerlemeler ve uygulamalar, LLM'lerin sağlık alanında güvenli, etik ve etkili bir şekilde benimsenmesini sağlamak için disiplinler arası işbirliği ve sürekli değerlendirme gerektirecektir. LLM-MedQA çalışmasında vurgulandığı gibi, çoklu-ajan mimarileri ve vaka üretimi gibi yenilikçi yaklaşımlar, gelecekteki daha güvenilir ve yorumlanabilir tıbbi yapay zeka sistemlerinin temelini oluşturabilir.

## 8 Sonuç

### 8.1 Ana noktaların özeti

Bu çalışma, LLM tabanlı tıbbi soru cevaplama sistemlerinin en son gelişmelerini ve özellikle LLM-MedQA yaklaşımını incelemiştir. Büyük Dil Modelleri, 2017’de tanıtılan Transformer mimarisiyle başlayan ve günümüzde milyarlarca parametreye sahip modellerle devam eden bir evrim geçirmiştir. Tıbbi soru cevaplama alanında, bu modeller geleneksel yöntemlerin karşılaştığı terminoloji anlama ve karmaşık akıl yürütme zorluklarını aşma potansiyeli göstermektedir.

LLM-MedQA yaklaşımı, tıbbi QA sistemlerini iyileştirmek için iki temel yenilik sunmuştur: (1) Farklı tıbbi uzmanlık alanlarından (soru uzmanları ve seçenek uzmanları) oluşan bir çoklu-ajan mimarisi ve (2) Gerçekçi klinik vakalar üreten yenilikçi bir vaka üretimi bileşeni. Bu yaklaşım, MedQA veri kümesi üzerinde %77.2 doğruluk oranı elde ederek, mevcut en iyi yöntemlerden yaklaşık %7 daha iyi performans göstermiştir.

Deneysel sonuçlar, hem model boyutunun hem de vaka üretimi bileşeninin sistem performansına önemli katkılar sağladığını göstermiştir. LLaMA3.1:70B gibi büyük modeller, daha küçük modellere kıyasla tıbbi sorulara cevap vermede çok daha etkilidir, ve vaka üretimi bileşeni doğruluk oranını %1-2 artırmıştır.

### 8.2 Tıbbi soru cevaplama LLM’lerin önemi

Tıbbi soru cevaplama, sağlık hizmetlerinin kalitesini ve erişilebilirliğini doğrudan etkileyen kritik bir teknolojidir. LLM’ler, bu alanda devrim niteliğinde gelişmeler sunarak, hem sağlık profesyonellerinin hem de hastaların tıbbi bilgiye erişimini demokratikleştirmektedir. LLM-MedQA makalesinde belirtildiği gibi, ”doğru ve verimli soru cevaplama sistemleri, tıp alanında yüksek kaliteli hasta bakımının sağlanması için kritik öneme sahiptir.”

LLM’lerin tıbbi soru cevaplamaadaki benzersiz değeri şu noktalarda ortaya çıkmaktadır:

- **Bilgi Entegrasyonu:** LLM’ler, geniş tıbbi literatürden edinilen bilgiyi entegre ederek, herhangi bir klinisyenin tek başına sahip olabileceğinden çok daha kapsamlı bir bilgi tabanı sunar.
- **Karmaşık Akıl Yürütme:** Çoklu-ajan mimarileri, farklı tıp uzmanlıklarından gelen bilgileri sentezleyerek, tek bir modelin yapabileceğinden daha kapsamlı tıbbi akıl yürütme sağlar.
- **Zero-Shot Öğrenme Yeteneği:** LLM’ler, ek eğitim verileri olmadan karmaşık tıbbi sorulara yanıt verebilmektedir. Bu, nadir hastalıklar veya yeni tedavi yöntemleri gibi eğitim verisi sınırlı olan alanlarda özellikle değerlidir.
- **Erişilebilirlik:** LLM tabanlı tıbbi QA sistemleri, tıbbi bilgiyi daha erişilebilir hale getirerek, coğrafi olarak izole edilmiş veya kaynak sınırlı ortamlardaki sağlık profesyonellerini destekler.

Geleneksel tıbbi QA sistemleri genellikle sınırlı bilgi tabanlarına ve kurallara dayalıydı. LLM’lerin ortaya çıkışı, tıbbi dili ve bağlamı anlama, karmaşık vakaları analiz etme ve çok çeşitli tıbbi sorulara yanıt verme konusunda özgün yetenekler sunmaktadır.

### 8.3 LLM-MedQA yaklaşımının tıp alanına katkıları ve etkisi

LLM-MedQA yaklaşımı, tıp alanına birkaç önemli katkı sağlamıştır:

- **Klinik Karar Destek Mekanizması:** Çoklu-ajan mimarisi, gerçek dünyada görülen konsültasyon süreçlerini modelleyerek, klinisyenlere daha kapsamlı ve güvenilir karar destek mekanizmaları sunmaktadır. Bu, özellikle karmaşık veya nadir vakaların yönetiminde değerlidir.
- **Eğitim ve Öğretim Aracı:** Vaka üretimi bileşeni, tıp öğrencileri ve asistanlar için gerçekçi klinik senaryolar sağlayarak, klinik akıl yürütme becerilerini geliştirmek için etkili bir eğitim aracı olabilir.
- **Yorumlanabilirlik ve Güven:** LLM-MedQA, sadece bir cevap değil, aynı zamanda klinik akıl yürütme sürecini gösteren kapsamlı bir rapor ve destekleyici vakalar da sunarak, sistemin önerilerine olan güveni artırmaktadır.
- **Klinik Uygulama Entegrasyonu:** Çoklu-ajan yaklaşımı ve vaka üretimi, tıbbi LLM’lerin klinik uygulamaya entegrasyonu için daha uygulanabilir bir yol sunmaktadır. Oylama mekanizması, uzman geri bildirimleri ve sistem kalibrasyonu yoluyla sürekli iyileştirmeye olanak tanır.
- **Performans İlerlemesi:** %7’lik performans iyileştirmesi, tıbbi bağlamda önemli bir ilerlemeyi temsil eder, çünkü doğru teşhis ve tedavi için küçük bir iyileşme bile klinik sonuçlarda büyük etki yaratabilir.

LLM-MedQA makalesinde geliştirilen yaklaşım, tıbbi bilgilerin kompleksliğini ve uzman görüşlerinin çeşitliliğini kapsayan, daha bütünsel ve gerçekçi bir tıbbi karar verme süreci sunmaktadır. Bu teknolojinin devamında geliştirilecek sistemler, tıp eğitimini zenginleştirebilir, klinik karar vermeyi destekleyebilir ve nihayetinde sağlık hizmetlerinin kalitesini iyileştirebilir. Yapay zekanın tıp alanında sorumlu ve etik kullanımı sağlandığı takdirde, LLM-MedQA benzeri yaklaşımlar hem tıp mesleğini hem de hasta bakımını olumlu yönde dönüştürme potansiyeline sahiptir.