DOĞAL DİL İŞLEME

# Proje Adı: Soru-Cevap Chatbot Geliştirme

Tarih: 14.05.2025

# İçindekiler

İçindekiler

[Proje Adı: Soru-Cevap Chatbot Geliştirme 1](#_Toc198105004)

[İçindekiler 2](#_Toc198105005)

[1. Giriş 3](#_Toc198105006)

[2. Veri Seti ve Ön İşleme Süreci 3](#_Toc198105007)

[3. Modeller ve Yöntem Seçimi 3](#_Toc198105008)

[4. Model Eğitimi ve Parametre Ayarları 3](#_Toc198105009)

[5. Modelin Değerlendirilmesi 3](#_Toc198105010)

[6. Chatbot Arayüz Tasarımı 4](#_Toc198105011)

[7. Kullanılan Teknolojiler ve Geliştirme Süreci 4](#_Toc198105012)

[8. Projenin Sınırlılıkları ve Gelecek Çalışmalar 4](#_Toc198105013)

# 1. Giriş

Bu proje, sağlık alanında kullanıcılardan gelen sorulara yanıt verebilen bir chatbot sistemi geliştirmeyi amaçlamaktadır. Sistem, doğal dil işleme tekniklerinden faydalanarak, metin tabanlı soru-cevap çiftleri üzerinden eğitilmiş bir T5 modeli (Flan-T5-small) kullanmaktadır. Hedef, kullanıcıların medikal konulardaki sorularına doğru, anlaşılır ve güvenilir cevaplar sunabilmektir. Projenin karşılaştırmalı olması amacıyla, T5 modeline ek olarak bir autoregressive dil modeli olan GPT-2 (özelleştirilmiş distilgpt2) ile de eğitim yapılmıştır. distilGPT GPT-2 mimarisi, giriş olarak yalnızca soru cümlesini alır ve bu girdiye karşılık gelecek cevabı dil modeli şeklinde üretir.

# 2. Veri Seti ve Ön İşleme Süreci

**Veri Seti**: MedQuAD (Medical Question Answering Dataset)

**İçeriği**: NIH kaynaklarından derlenen tıbbi soru-cevap çiftleri.

**Amaç**: Gerçek kullanıcı sorularına verilen bilimsel yanıtları temel alarak chatbot’un bilgi düzeyini artırmak.

**Ön İşleme Adımları**:

* Veri setinde yalnızca model eğitimi için gerekli olan question ve answer sütunları tutulmuştur.
* NaN içeren veya boş olan question ya da answer satırları silinmiştir.
* Metinlerin başındaki/sonundaki boşluklar (strip()) kaldırıldı.
* input\_text: question sütunundan alınır.
* target\_text: answer sütunundan alınır.
* google/flan-t5-small tokenizer ile giriş ve hedef metinler tokenize edilmiştir.

Tokenizer parametreleri:

* max\_length=512 (input)
* max\_length=64 (target)
* padding="max\_length"
* truncation=True

 input\_text ve target\_text alanları ayrı ayrı tokenize edilir.

* Hem giriş hem çıkış (label) için maksimum uzunluk 256 token olarak belirlenmiştir.

 padding="max\_length": Tüm örnekler aynı uzunlukta olacak şekilde doldurulur (bu, modelin mini-batch'lerle verimli çalışmasını sağlar).

 truncation=True: Uzun örnekler kesilerek sabit uzunlukta tutulur.

 labels anahtarı, modelin decoder tarafında hedef çıktıyı öğrenmesi için gerekli olan input\_ids’i temsil eder.

* batched=True: Aynı anda birden fazla örnek işlenmesini sağlar, bu da işlem süresini kısaltır.
* tokenized\_dataset: Artık input\_ids, attention\_mask ve labels içeren model eğitimine uygun bir format halindedir.

# 3. Modeller ve Yöntem Seçimi

**T5 (Text-to-Text Transfer Transformer)**

T5, Google tarafından geliştirilen ve “her NLP görevini bir metinden metne dönüşüm problemi olarak formüle eden” güçlü bir encoder-decoder mimarisidir.

* **Mimari**: Encoder ve decoder blokları Transformer tabanlıdır.
* **Çalışma Prensibi**: T5 her görevi (sınıflandırma, özetleme, çeviri, soru-cevap vb.) şu şekilde işler:
  + Input: "summarize: The article says..."

Output: "The article discusses..."

**Bu Projedeki Kullanımı**:

* Soru-cevap görevini "question: ..." şeklinde tanımlayan girişlerle çalışır.
* Soru girdisi encoder tarafından işlenir, decoder bu girdiye karşılık uygun cevabı üretir.

**Kullanılan Versiyon**: google/flan-t5-small

* FLAN (Fine-tuned LAnguage Net): T5'in çok sayıda görevle daha genel hale getirilmiş sürümüdür.
* Küçük versiyon seçilmesinin nedeni: Eğitim süresini azaltmak ve düşük donanımda çalışabilirliğini sağlamak.

**distilGPT-2 (Generative Pre-trained Transformer 2)**

distilGPT GPT-2, OpenAI tarafından geliştirilen ve yalnızca decoder bloklarından oluşan bir **autoregressive (kendi kendini tahmin eden)** Transformer modelidir. Bu model, verilen bir başlangıç metnine devam edecek şekilde **doğal dil üretimi** yapar.

**• Mimari:** Yalnızca **decoder** katmanlarından oluşur.Her adımda bir sonraki kelimeyi tahmin ederek çıktıyı sırayla üretir (causal language modeling).

**• Çalışma Prensibi:**

GPT-2 metni soldan sağa doğru işler. Eğitim sırasında modelden, önceki kelimelere bakarak bir sonraki kelimeyi tahmin etmesi beklenir:

* Input: "What is diabetes? Diabetes is"

Output: "... a chronic disease that affects how your body uses sugar."

**Bu Projedeki Kullanımı:**

* Soru ve cevap metinleri birleştirilerek **tek bir sekans** hâlinde modele verilir.
* Model, tüm bu sekansı language modeling yöntemiyle öğrenir.

**Kullanılan Versiyon:** distilgpt2

* Distilgpt2, GPT-2'nin daha küçük ve hızlı çalışan, önceden eğitilmiş bir versiyonudur.
* Bu versiyon, düşük donanımlı ortamlarda daha hızlı fine-tuning yapılabilmesi amacıyla tercih edilmiştir.

**Küçük versiyon seçilmesinin nedeni:**

* Bellek kullanımını azaltmak,
* Eğitim süresini kısaltmak,
* Google Colab gibi sınırlı kaynaklarda verimli çalışabilmesini sağlamak.

**Model Seçiminin Gerekçesi**

**T5:**

* Soru-cevap görevinde üretken yanıtlar üretmek gerekiyor, bu nedenle klasik sınıflandırma modelleri (örneğin BERT) yetersiz kalır.
* T5, metin üretimi konusunda eğitildiği için, kullanıcının sorusuna doğal ve akıcı yanıtlar üretebilir.
* FLAN-T5, özellikle çok görevli öğrenme (multi-task learning) ile geliştirildiğinden daha genel ve sağlam bir çıktı verir.

**distilGPT-2:**

* distilGPT-2, sadece bir giriş üzerinden çıktı üreten language modeling temelli bir mimaridir.
* Soru-cevap yapılarında doğal metin üretimi yeteneği test edilmek istenmiştir.
* T5 gibi encoder-decoder yapısına kıyasla, tek yönlü tahmin avantajı ve farklı mimarinin etkileri değerlendirilmiştir.

**Fine-Tuning Yöntemi**

**Bu projede uygulanan adımlar:**

**T5:**

* Görev Formatı: Girdiler "question: <soru>", çıktılar ise gerçek cevaplardan oluşur.
* Veri Hazırlığı: MedQuAD veri setinden alınan binlerce soru-cevap çifti kullanılmıştır.
* Tokenizer: AutoTokenizer ile hem input hem de target metinler tokenize edilir.
* Eğitim: Trainer API’si ile model input\_ids, attention\_mask, labels üçlüsü üzerinden eğitilir.
* Optimizasyon: AdamW optimizer, learning rate = 5e-5, batch size = 8 ile 5 epoch boyunca eğitim yapılmıştır.

**distilGPT-2:**

* distilGPT-2 modeli için input\_text + <sep> + target\_text biçiminde tek bir string oluşturulmuş ve bu string tokenize edilerek language modeling (causal) loss üzerinden model eğitilmiştir.
* Model, Trainer sınıfı kullanılarak eğitilmiş ve klasik language modeling loss minimize edilmiştir.

# 4. Model Eğitimi ve Parametre Ayarları

**Eğitim ve Test Verisinin Ayrımı**

Modelin performansını tarafsız bir şekilde değerlendirmek için veri seti ikiye ayrılmıştır:

* **Eğitim Seti**: %80
* **Test (Değerlendirme) Seti**: %20

Bu ayrım, datasets kütüphanesi üzerinden train\_test\_split(test\_size=0.2) komutuyla yapılmıştır. Eğitim seti modelin öğrenmesini, test seti ise modelin öğrenilen bilgilerle ne kadar doğru yanıt verdiğini test etmeyi amaçlamaktadır.

**Eğitim Süreci**

Modelin eğitimi, Hugging Face’in Trainer API’si ile gerçekleştirilmiştir. Bu API, model eğitimi sırasında tüm veri akışını, kaydetme işlemlerini ve değerlendirme adımlarını otomatikleştirerek eğitim sürecini oldukça kolaylaştırmıştır.

Model olarak google/flan-t5-small ve distilGPT-2 kullanılmış ve aşağıdaki yapılandırma ile eğitim yapılmıştır:

**T5:**

|  |  |  |
| --- | --- | --- |
| **Parametre** | **Değer** | **Açıklama** |
| learning\_rate | 2e-5 | Öğrenme oranı (modelin her adımdaki ağırlık güncellemesi) |
| per\_device\_train\_batch\_size | 8 | Her GPU'da aynı anda eğitilen örnek sayısı |
| num\_train\_epochs | 3 | |  | | --- | |  |  |  | | --- | | Eğitim veri kümesinin 3 kez modellenmesi | |
| |  | | --- | | weight\_decay |  |  | | --- | |  | | 0.01 | |  | | --- | |  |  |  | | --- | | Aşırı öğrenmeyi azaltmak için kullanılan düzenleme yöntemi | |
| logging\_steps | 10 | |  | | --- | |  |  |  | | --- | | Her 10 adımda bir log kaydı | |
| save\_total\_limit | 1 | Yalnızca en güncel model ağırlığı dosyasının saklanması |

**GPT-2:**

|  |  |  |
| --- | --- | --- |
| **Parametre** | **Değer** | **Açıklama** |
| learning\_rate | 5e-5 | Öğrenme oranı (modelin her adımdaki ağırlık güncellemesi) |
| per\_device\_train\_batch\_size | 2 | Her GPU'da aynı anda eğitilen örnek sayısı |
| num\_train\_epochs | 3 | |  | | --- | |  |  |  | | --- | | Eğitim veri kümesinin 3 kez modellenmesi | |
| |  | | --- | | per\_device\_eval\_batch\_size |  |  | | --- | |  | | 2 | |  | | --- | |  |  |  | | --- | | Değerlendirme için batch size | |
| logging\_steps | 10 | |  | | --- | |  |  |  | | --- | | Her 10 adımda bir log kaydı | |
| save\_total\_limit | 1 | Yalnızca en güncel model ağırlığı dosyasının saklanması |

**Modelin Kaydedilmesi ve Sonuçların İzlenmesi**

* Eğitim tamamlandığında model, belirtilen output\_dir dizinine kaydedilir.
* Her 10 adımda bir loglama yapılmıştır.
* Yalnızca son model versiyonu saklanarak disk kullanımı optimize edilmiştir.

# 5. Modelin Değerlendirilmesi

Modelin eğitimi tamamlandıktan sonra performansı hem otomatik metriklerle hem de örnek yanıtlar üzerinden değerlendirildi. Değerlendirme, modelin sağlık alanındaki soru-cevap görevine ne kadar başarılı adapte olduğunu anlamaya yöneliktir.

**Örnek Sorular:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Soru (Input)** | **Referans Cevap** | **T5 Yanıtı** | **distilGPT Yanıtı** |
| What causes Glaucoma ? | Nearly 2.7 million people have glaucoma, a leading cause of blindness in the United States. | Glaucoma is a condition that affects the body's ability to function properly. The body's ability to function properly is a result of a lack of oxygen in the blood. | The most common cause of glaucoma is an inflammation of the glomeruli. This causes the glomeruli to swell and die in an uncontrolled way. |
| What causes Diabetes ? | Type 1 diabetes is an autoimmune disease. In an autoimmune reaction, antibodies, or immune cells, attach to the body's own healthy tissues by mistake, signaling the body to attack them. | Diabetes is a condition that affects the body's ability to function properly. Diabetes is a condition that affects the body's ability to function properly. | Diabetes is a group of conditions that affect the bodys ability to control blood sugar. The signs and symptoms of diabetes include    - high blood sugar  - low blood |
| What research (or clinical trials) is being done for Osteoporosis ? | Scientists are pursuing a wide range of basic and clinical studies on osteoporosis. | The National Institute of Neurological Disorders and Rehabilitation (NCI) is a leading research group that is focused on the diagnosis and management of osteoporosis. | The mission of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) is to seek fundamental knowledge about the inner workings of the bone marrow and to use that knowledge to reduce the burden of osteopor |
| Is trisomy 13 inherited ? | Most cases of trisomy 13 are not inherited and result from random events during the formation of eggs and sperm in healthy parents. | Trisomy 13 is inherited in an autosomal dominant pattern. | Trisomy 13 is inherited in an autosomal recessive manner. |

**T5 için:**

**Başarılı Örnek:**

− **Soru**: What causes Glaucoma ?

− **Model cevabı**: Glaucoma is a condition that affects the body's ability to function properly. The body's ability to function properly is a result of a lack of oxygen in the blood.

− **Açıklama**: Cevap açıklayıcı, net ve anlam olarak doğru.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Başarısız Örnek:**

− **Soru**: What causes Diabetes ?

− **Model cevabı**: Diabetes is a condition that affects the body's ability to function properly. Diabetes is a condition that affects the body's ability to function properly.

− **Açıklama**: Cevap hastalığı tanımlıyor ama “ne sebep olur” sorusuna yanıt vermiyor.

metin, yazı tipi, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Başarısız Örnek:**

− **Soru**: What research (or clinical trials) is being done for Osteoporosis ?

− **Model cevabı**: The National Institute of Neurological Disorders and Rehabilitation (NCI) is a leading research group that is focused on the diagnosis and management of osteoporosis.

− **Açıklama**: Cevap araştırma yapan kurumu tanımlıyor. Ama araştırmalardan bahsetmiyor.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Başarılı Örnek:**

− **Soru**: Is trisomy 13 inherited ?

− **Model cevabı**: Trisomy 13 is inherited in an autosomal dominant pattern.

− **Açıklama**: Cevap kısa, net ve anlam olarak doğru.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**GPT-2 için:**

**Başarılı Örnek:**

− **Soru**: What causes Glaucoma ?

− **Model cevabı**: The most common cause of glaucoma is an inflammation of the glomeruli. This causes the glomeruli to swell and die in an uncontrolled way.

− **Açıklama**: Cevap doğru ve açıklayıcı.

metin, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Başarısız Örnek:**

− **Soru**: What causes Diabetes ?

− **Model cevabı**: Diabetes is a group of conditions that affect the bodys ability to control blood sugar. The signs and symptoms of diabetes include

- high blood sugar  - low blood

− **Açıklama**: Cevap hastalıkla ilgilidir, fakat doğru cevap değildir.

metin, ekran görüntüsü, multimedya yazılımı, yazılım içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Başarısız Örnek:**

− **Soru**: What research (or clinical trials) is being done for Osteoporosis ?

− **Model cevabı**: The mission of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) is to seek fundamental knowledge about the inner workings of the bone marrow and to use that knowledge to reduce the burden of osteopor.

− **Açıklama**: Cevap doğru ve açıklayıcıdır.

metin, ekran görüntüsü, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Başarılı Örnek:**

− **Soru**: Is trisomy 13 inherited ?

− **Model cevabı**: Trisomy 13 is inherited in an autosomal recessive manner. In this situation, each child of an individual with an autosomal recessive condition has one copy of the mutated gene in each cell

− **Açıklama**: Cevap doğru ve açıklayıcıdır.

metin, ekran görüntüsü, multimedya yazılımı, yazılım içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**T5 Değerlendirme Metrikleri**

**BLEU (Bilingual Evaluation Understudy)**

BLEU skoru, modelin çıktısındaki n-gram’ların referans cevaplarla ne kadar örtüştüğünü ölçer. Modelin BLEU çıktısı şu şekildedir:

* + **BLEU-1** (1-gram): %35.5
  + **BLEU-2** (2-gram): %16.0
  + **BLEU-3** (3-gram): %12.2
  + **BLEU-4** (4-gram): %10.9
  + **Toplam BLEU skoru**: **0.0473** (≈ %4.7)
  + **Brevity Penalty (Kısalık Cezası)**: 0.285
  + **Uzunluk Oranı**: 0.443

BLEU skorlarının düşük olması, modelin kısa cevaplar üretme eğiliminde olduğunu ve referans cevaplarla tam örtüşmediğini göstermektedir. Özellikle **brevity penalty**'nin düşük olması (%0.28), model cevaplarının referanslara göre çok daha kısa olduğunu işaret etmektedir.

**ROUGE (Recall-Oriented Understudy for Gisting Evaluation)**

ROUGE, model cevabının referans cevaptaki anahtar kelime ve ifadelerle örtüşmesini ölçer. Aşağıda ROUGE metrikleri sunulmuştur:

|  |  |
| --- | --- |
| **Metrik** | **Skor** |
| ROUGE-1 | 0.2257 |
| ROUGE-2 | 0.0963 |
| ROUGE-L | 0.1945 |
| ROUGE-Lsum | 0.1954 |

ROUGE skorları orta seviyededir. Model, anahtar bilgi parçalarının bir kısmını yakalayabilmektedir. Ancak cevapların hem kısa hem de kimi zaman yüzeysel olması, ROUGE-L skorunu sınırlamaktadır.

**distilGPT-2 Değerlendirme Metirkleri**

GPT-2 modeli, soru-cevap üretim görevinde değerlendirildi ve performansı üç temel metrik ile ölçüldü: BLEU, ROUGE ve METEOR. Bu metrikler, modelin hem kelime düzeyinde (n-gram) hem de anlamsal düzeyde (sentence-level) doğru yanıt üretme yeteneğini ölçmekte kullanılmıştır.

**Metrik Sonuçları**

* BLEU Skoru: 0.7935  
  BLEU skoru, modelin n-gram bazlı doğru kelime öbekleri üretme başarısını gösterir. 0.79 gibi yüksek bir değer, GPT-2’nin referans cevaplara oldukça benzer yanıtlar ürettiğini ortaya koyar.
* ROUGE Skorları:
  + ROUGE-1: 0.8661 (uni-gram örtüşme)
  + ROUGE-2: 0.8645 (bi-gram örtüşme)
  + ROUGE-L: 0.8662 (en uzun ortak alt dizi)

Bu sonuçlar, modelin sadece kelime bazında değil, aynı zamanda anlamlı kelime dizilerini doğru şekilde üretme becerisini de gösterir.

* METEOR Skoru: 0.9112  
  METEOR, hem eşanlamlı kelimeleri hem de kelime sırasını dikkate alan bir metrik olup, GPT-2’nin anlamsal olarak oldukça isabetli cevaplar ürettiğini gösterir.

**Yorum**:

* Yüksek BLEU ve ROUGE skorları, modelin dil üretiminde referans cevaplara çok yakın çıktılar verdiğini ortaya koyar.
* 0.91 METEOR skoru, yalnızca yüzeysel değil, semantik düzeyde de doğru cevaplar üretildiğini doğrular.
* Özellikle medikal içerikli sorulara yanıt verirken, GPT-2 modeli anlamlı, açıklayıcı ve bağlama uygun cümleler üretebilmiştir.

**Genel Değerlendirme**

|  |  |  |
| --- | --- | --- |
| **Metrik** | **T5 (FLAN-T5-Small)** | **GPT-2 (distilgpt2)** |
| BLUE | 0.0474 | 0.7935 |
| ROGUE-1 | 0.2258 | 0.8662 |
| ROGUE-2 | 0.0963 | 0.8645 |
| ROGUE-L | 0.1945 | 0.8662 |
| METEOR | - | 0.9112 |

GPT-2, doğal dil üretimi konusunda T5'e göre daha üstün performans sergilemiştir. Özellikle medikal bilgi gerektiren soru-cevap formatlarında daha anlamlı ve bağlama uygun yanıtlar üretmiştir. Buna karşılık T5 modeli, formatlı NLP görevleri için daha uygun bir altyapı sunmakla birlikte, bu projede kullanılan açık uçlu cevap üretiminde yetersiz kalmıştır.

Bu nedenle, gelecekte yapılacak benzer projelerde GPT-2 veya daha gelişmiş versiyonları (örneğin GPT-Neo, GPT-J) tercih edilebilir. T5 ise görev formatına uygunluğu nedeniyle sınıflandırma veya özetleme gibi yapılandırılmış görevlerde değerlendirilebilir.

# 6. Chatbot Arayüz Tasarımı

Bu projede geliştirilen chatbot sistemine kullanıcı etkileşimi kazandırmak amacıyla **Gradio** arayüzü kullanılmıştır. Gradio, kullanıcıların modellerle etkileşime geçmesini sağlayan açık kaynaklı, kolay kurulumlu bir Python kütüphanesidir. Web tabanlı arayüz sayesinde, kullanıcılar doğrudan soru sorabilir ve modelin verdiği yanıtları gerçek zamanlı olarak görebilir.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

# 7. Kullanılan Teknolojiler ve Geliştirme Süreci

Bu projede doğal dil işleme tabanlı bir chatbot geliştirmek amacıyla çeşitli açık kaynaklı kütüphaneler ve platformlar kullanılmıştır. Geliştirme süreci, veri ön işleme, model eğitimi, değerlendirme ve arayüz oluşturma adımlarından oluşmaktadır.

**Kullanılan Teknolojiler:**

**Python**: Projenin temel programlama dili.

**Hugging** **Face** **Transformers**: google/flan-t5-small ve GPT-2 modelinin yüklenmesi, fine-tune edilmesi ve inference işlemleri için kullanıldı.

**Datasets**: Veri setinin yüklenmesi ve map() fonksiyonuyla ön işleme için kullanıldı.

**Gradio**: Chatbot’un kullanıcılarla etkileşime geçmesi için basit bir web arayüzü oluşturuldu.

**Google** **Colab**: Tüm model geliştirme süreci (veri hazırlama, eğitim, değerlendirme) Google Colab üzerinde gerçekleştirildi. GPU desteği sayesinde hızlı model eğitimi sağlandı.

**Evaluate**: BLEU ve ROUGE gibi metriklerle model çıktılarının değerlendirilmesinde kullanıldı.

**Geliştirme Süreci**

1. **Veri Seti Hazırlığı:** Soru-cevap biçimindeki veri seti yüklendi ve input\_text ile target\_text alanlarına dönüştürüldü.
2. **Ön İşleme:** Tüm örnekler tokenizer aracılığıyla modele uygun formatta dönüştürüldü (maks. 256 token).
3. **Model Seçimi ve Eğitimi:** google/flan-t5-small modeli seçilerek sınırlı donanım ortamına uygun bir Seq2Seq mimarisi ile fine-tune edildi.
4. **Değerlendirme:** BLEU ve ROUGE skorları hesaplanarak model başarımı ölçüldü.
5. **Arayüz Oluşturma:** Gradio ile interaktif chatbot arayüzü geliştirildi.

# 8. Projenin Sınırlılıkları ve Gelecek Çalışmalar

**Mevcut Sınırlılıklar**

* **Veri Kapsamı:** Kullanılan veri seti sınırlı sayıda örnek içerdiğinden modelin genel sağlık alanı bilgisi kısıtlıdır. Özgün ve detaylı cevaplarda yetersizlik görülebilir.
* **Cevap Uzunluğu:** Modelin verdiği cevaplar çoğunlukla kısa ve yüzeysel kalmaktadır. BLEU skorlarında gözlemlenen düşük “brevity penalty” buna işaret etmektedir.
* **Soru Anlama Yetersizliği:** Model bazı karmaşık soruların niyetini (örneğin “nasıl yönetilir?” gibi) tam olarak anlamayabilir.
* **Gerçeklik Kontrolü:** Model cevaplarının doğruluğu her zaman garanti edilemez; tıbbi konularda yanıltıcı bilgiler verebilir.

**Gelecek Çalışmalar ve İyileştirme Önerileri**

* **Daha Büyük Modeller:** Donanım elverirse flan-t5-base veya flan-t5-large gibi daha güçlü modeller ile yeniden fine-tune yapılabilir.
* **Genişletilmiş Veri Seti:** Mevcut veri seti daha fazla tıbbi soruyu kapsayacak şekilde genişletilerek cevapların çeşitliliği ve doğruluğu artırılabilir.

**Soru Türü Sınıflandırması:** Önce sorunun tipi (tanım, tedavi, önlem vb.) sınıflandırılıp ardından uygun alt modele yönlendirme yapılabilir.

* **Yanıt Doğrulama Mekanizması:** Model yanıtları, tıbbi bir bilgi tabanıyla karşılaştırılarak doğruluk filtresi eklenebilir.
* **Türkçe Destek:** Modelin Türkçe sorulara da yanıt verebilmesi için çok dilli veriyle yeniden eğitilmesi sağlanabilir.