

Türkçe Soru Cevaplamada Açık Kaynak BDM'lerinin İnce Ayarı: Nutuk Örneğinde Veri Kümesi Boyutu ve Parametre Seçiminin Etkisi

Fine-Tuning Open-Source LLMs For Turkish Question Answering: The Impact of Data Set Size and Parameter Selection in the Nutuk Example

İlknur Gür Nalçacı

Sobrantech Dijital Dönüşüm ve Yapay Zekâ A.Ş.

İstanbul, Türkiye

ilknur.gur@sobrantech.com.tr

Ferhat Şeker

Sobrantech Dijital Dönüşüm ve Yapay Zekâ A.Ş.

İstanbul, Türkiye

ferhat.seker@sobrantech.com.tr

Kasım Murat Karakaya

TED Üniversitesi

Yazılım Mühendisliği

Ankara, Türkiye

murat.karakaya@tedu.edu.tr

Öz—Çalışmamızda açık kaynak büyük dil modellerinin (BDM) Türkçe metin derlemeleri üzerinde gerçekleştirilmiş ince ayar (fine-tuning) süreçleri ele alınarak, bu süreçlerde model başarımını ve eğitim verimliliğini etkileyen temel etkenleri incelenmiştir. Bu kapsamda Mustafa Kemal Atatürk'ün "Nutuk" eseri üzerinden sentetik olarak farklı veri boyutlarında soru cevap çiftleri üretilmiş ve bu veri kümeleri üzerinden, farklı açık kaynak dil modelleri parametre verimli ince ayar yöntemi: LoRA kullanılarak eğitilmiştir. Kıyaslama için üretilen veri kümesi üzerinden ana modellerin ve 12 adet ince ayar (fine tune) yapılmış modellerin başarımları karşılaştırılmıştır. Nutuk bağlamında oluşturulan özgün soru-cevap çiftleri üzerindeki yanıt doğruluğu ve anlamsal tutarlılık değerlendirmesinde elde edilen bulgular, ince ayar çalışmalarının Türkçe belgelerden oluşturulacak soru cevap ince ayar eğitimlerinde başarımları artırabileceğini göstermekte; özellikle kaynak kısıtlı ve alan özgü araştırmalara yönelik senaryolarda yol gösterici nitelik taşımaktadır.

Anahtar Sözcükler — Açık Kaynaklı Büyük Dil Modelleri; Soru Cevaplama; Türkçe Doğal Dil İşleme; LoRA; Parametre Verimli İnce Ayar, Veri Kümesi Ölçeklendirme; Başarım Değerlendirmesi; Sentetik Veri Üretimi; Nutuk.

Abstract—In this study, fine-tuning processes of open source large language models (LLM) performed on Turkish text corpora were examined and the main factors affecting model performance and training efficiency in these processes were examined. In this context, question-answer (QA) pairs were synthetically generated in different data sets from Mustafa Kemal Atatürk's work "Nutuk" and different open source language models were trained on these data sets using the parameter-efficient fine-tuning method: LoRA. The performances of the base models and 12 fine-tuned models were compared using the benchmark dataset that was generated for evaluation purposes. The findings obtained from the evaluation of answer accuracy and semantic consistency on the original question-answer pairs created in the context of Nutuk indicate that fine-tuning can enhance the performance of Turkish question-answering tasks. These results are particularly informative for resource-constrained and domain-specific research scenarios.

Keywords — Open Source Large Language Models; Question Answering; Turkish Natural Language Processing; LoRA;

parameter-efficient fine-tuning; Dataset Scaling; Performance Evaluation; Synthetic Data Generation; Nutuk.

1. GİRİŞ

Büyük Dil Modelleri (BDM- LLM'ler), doğal dil işleme alanında önemli bir paradigma kayması yaratarak, metin anlama ve üretme konusunda önemli yetenekler sergilemiştir. Bu modellerin SC sistemlerindeki potansiyeli, karmaşık sorgulara anlamlı ve doğru yanıtlar üretebilme kapasiteleri nedeniyle özellikle dikkat çekicidir. Bu bağlamda, modellerin hedeflenen görev ve alanlara uyarlanması için ince ayar süreçleri kritik bir rol oynamaktadır.

Geleneksel tam model ince ayar yöntemlerinin yüksek hesaplama maliyetleri göz önüne alındığında, LoRA (Low-Rank Adaptation) gibi Parametre Verimli İnce Ayar (PEFT) teknikleri [1], daha az kaynakla etkili bir uyarılma sağlamak için umut verici bir alternatif olarak ortaya çıkmıştır. Bu çalışmada, dört farklı açık kaynak BDM ailesine ait olan Gemma-2-9B [2], Meta-Llama-3.1-8B [3], Mistral-7B-Instruct-v0.3 [4] ve Qwen3-8B [5] modellerinin, yönerge temelli ve LoRA yöntemiyle ince ayar uygulanmış sürümleri, Nutuk [6] metni üzerinden Türkçe SC başarımı açısından çözümlenmiştir. Çalışmanın amacı farklı açık kaynak BDM mimarilerinin, Nutuk gibi özgün bir Türkçe tarihî metin üzerinde yapılan ince ayarlara nasıl tepki verdiğini çözümlenmek; veri miktarındaki artışın model başarımı üzerindeki etkisini (ölçeklenebilirlik) değerlendirmek; sabit hiperparametrelerle uygulanan LoRA tekniğinin genel etkinliğini ve her modelde sağladığı başarımların kazanımını ölçmek yer almaktadır. Ayrıca, genel amaçlı modellerin bu özel alana ne ölçüde başarıyla adapte olabildiğini değerlendirmek ve geliştirilen sentetik veri üretim stratejisinin, modelleri hedeflenen doğrultuda eğitime kapasitesini dolaylı olarak gözlemlemek de çalışmanın amaçları arasındadır.

II. İLGİLİ ÇALIŞMALAR

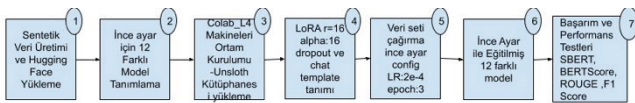
BDM'lerin görev odaklı veya alan özgü veri kümeleriyle ince ayar sürecinden geçirilmesi, özellikle Türkçe gibi kaynakları sınırlı dillerde bağlamsal uyumu ve yanıt doğruluğunu artırarak model başarımını anlamlı ölçüde geliştirmektedir; bu süreç, temel modellerin özel ya da kültürel bağlamlarda genelleme kapasitesinin yetersiz kaldığı durumlarda kritik bir gereklilik halini almaktadır [7],[8].

Geleneksel ince ayar yöntemleri ilgili modelin tüm parametrelerinin güncellenmesini gerektirmekte bu da çok büyük hesaplama gücü (GPU), bellek ve zaman gerektirmektedir. Bu soruna çözüm olarak geliştirilen PEFT yöntemlerinden biri olan LoRA büyük dil modellerinin farklı görevlere düşük maliyetle, hızlı ve verimli şekilde uyarlanmasını sağlayan bir tekniktir. Bu sayede hem eğitim süreci hafifletilir hem kısıtlı bilgi bulunan alanlardaki veri kümeleriyle etkili sonuçlar elde edilebildiği görülmüştür [1]. Büyük Dil Modellerinin ince ayar süreçleri ve bu süreçlerin başarımını etkileyen etkenler, literatürde kapsamlı şekilde ele alınmıştır. BDM ince ayarında model boyutu, veri boyutu gibi ölçeklendirme etkenlerinin etkisi sistematik olarak araştırılmış; ince ayar süreçlerinde optimal başarımın genellikle önemli miktarda veri kümesi gerektirdiği ve daha büyük modellerin potansiyelini gösterebilmesi için yeterli ve ilgili eğitim verisine ihtiyaç duyulduğu belirtilmektedir [8]. Eğitim veri kümesi oluşturulmasında; sentetik veri üretimi, özellikle alan özgü çalışmalarda veya belirli soru türleri için hedeflenmiş veri kümeleri oluşturmada kritik bir strateji olarak öne çıkmaktadır [9]. Sentetik olarak SC veri kümesi üretilmesinde Gemini modelinin GPT-4, Claude modellerine göre hem doğruluk hem de bütünlük bakımından veri kümesi üretiminde en güvenilir LLM olarak öne çıkmıştır [10].

İngilizce dışındaki dillerde ve tarihi metinler gibi özel alanlarda LLM uygulamaları üzerine yapılan araştırmalar da giderek önem kazanmaktadır. Örneğin, Mısır tarihi üzerine Arapça Soru-Cevap sistemlerinde LoRA ile ince ayarlanmış BDM'lerin etkinliğini inceleyen çalışmalar, bu alandaki potansiyeli göstermektedir [11]. Bu tür alan ve dil özelleşmesi odaklı araştırmalar, BDM'lerin farklı kültürel ve dilsel bağlamlardaki uyarlama yeteneklerini ve sınırlarını anlamak açısından büyük değer taşımaktadır.

III. YÖNTEM

Nutuk metni üzerinden sentetik veri üretim aşamaları, ana modeller üzerinden ince ayar uygulanmış yeni modellerin oluşturulma ve eğitim ortamı kurulumlarına yönelik çalışmalar bu bölümde ele alınmıştır. Çalışmanın akış şeması ve çalışmada kullanılan modeller, veri üretim süreci, ince ayar metodolojisi ile değerlendirme metrikleri Şekil 1'de gösterilmiştir.



Şekil 1. Modellere İnce Ayar Uygulanması Akış Şeması

A. Kullanılan Modeller ve Temel Kurulum

Çalışmamızda, dört farklı BDM ailesinin, genel amaçlı ön eğitimlerinin ardından yönerge tabanlı model (instruction-

tuned) versiyonları kullanılmıştır: Gemma-2-9B-IT, Meta-Llama-3.1-8B-Instruct, Mistral-7B-Instruct-v0.3 ve Qwen3-8B. Tüm modellerin, Unsloth [12] kütüphanesi aracılığıyla erişilen ve 4-bit niceleme uygulanmış, Q8_0 GGUF [13] formatına dönüştürülebilen versiyonları tercih edilmiştir. Böylelikle verimliliğin artırılması ve daha erişilebilir donanımlarda çalışılması mümkün kılınmıştır. Eğitim ortamı olarak Colab [14] L4 makineleri (22, GPU RAM, 53GB Sistem RAM, 235,7 Disk) kullanılmış ve bu ortam üzerinden başarımlar ölçümleri alınmıştır.

B. Sentetik Veri Üretim Stratejisi

Bu çalışmada kullanılan tüm veri kümeleri, Nutuk adlı eserin tam metninden, sentetik Soru-Cevap (SC- QA) çiftleri üretmek amacıyla geliştirdiğimiz özel bir Yapay Zekâ işgöreni (AI Agent) aracılığıyla oluşturulmuştur. Bu Yapay Zekâ işgöreni, Google Gemini 1.5 Flash API [15]'si üzerinden belgenin her sayfasından, önceden tanımlanmış sekiz farklı soru tipini (Gerçek, Kavramsal, Bağlamsal, Nedensel, Analitik, Yansıtıcı, Listeleme, Özetleme) hedeflemiş ve üretilen çıktıların format, yapı ve hedeflenen sayıya uygunluğunu kontrol ederek nutuk_min_dataset, nutuk_mid_dataset ve nutuk_max_dataset olarak her birisi sırasıyla 1290, 2580 ve 4920Türkçe SC çiftleri meydana getirmiştir. Kıyaslama (Benchmark) veri kümesinin üretiminde ise Gemini 2.0 Flash API'si ile ince ayar verisinden niteliksel olarak ayrılan ve modelleri daha zorlayıcı sorgularla sınaması hedeflenen 500 özgün SC çifti elde edilmiştir.

C. İnce Ayar Süreci

Bu çalışmada, dört farklı LLM'in her biri için, yukarıda açıklanan üç farklı veri kümesi (Min, Mid, Max) üzerinde, Unsloth kütüphanesi aracılığıyla LoRA tekniği kullanılarak Gemma-2-9B, Meta-Llama-3.1-8B, Mistral 7B Instruct v0.3 ve Qwen3-8B modelleri üzerinden birbirinden bağımsız bir ince ayar süreci yürütülmüştür. Bu kapsamda Şekil 1'deki akışta belirtilen hiperparametreler her modelde sabit olarak kullanılmıştır. İnce ayar sonrasında, eğitilen LoRA adaptörlerinin temel modelle birleştirilmesiyle elde edilen modeller standart 16-bit formatında kaydedilmiş ve ardından Q8_0 niceleme metodu ile GGUF formatına dönüştürülmüştür.

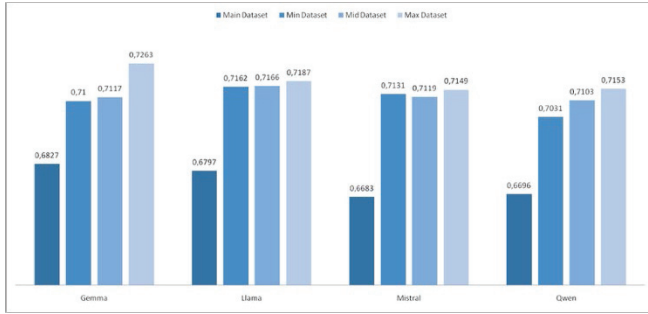
Ç. Kıyaslama ve Değerlendirme Ölçütleri

Bu çalışmada, toplam 16 modelin başarımını değerlendirmeye alınmıştır. Bu modeller; dört farklı BDM ailesine ait yönerge tabanlı versiyonları ve bu versiyonların her birinin üç farklı veri kümesiyle LoRA tekniği kullanılarak ince ayar uygulanmış üçer varyantı olmak üzere toplam 12 farklı modelden oluşmaktadır. Tüm modeller, daha önce detayları verilen ve 500 özgün SC çiftinden oluşan kıyaslama veri kümesi üzerinde test edilmiştir. Karşılaştırmaların adil ve tutarlı olması amacıyla, tüm kıyaslama testleri aynı L4 GPU donanımında ve önceden belirlenmiş sabit çıkarım parametreleri kullanılarak yürütülmüştür. Modellerin soru cevaplama başarımının farklı açılardan gözlenmesi amacıyla SBERT Similarity [16], Bert Score F1 [17], ROUGE [18], F1 Skor [19] ölçümleri alınarak değerlendirilmiştir. Böylelikle modellerin karşılaştırmalı olarak hangi yönlerinin güçlü olduğu gözlemlenebilecektir.

IV.ÇÖZÜMLEME SONUÇLARI

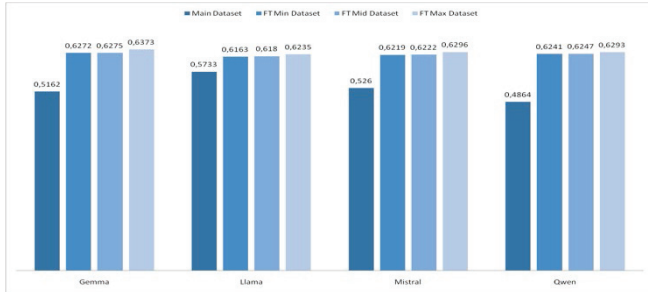
Bu bölümde, çalışmada değerlendirilen toplam 4 yönerge tabanlı ana model ve 12 ince ayarlanmış toplam 16 modelin, kıyaslama veri kümesi üzerindeki çözümleme bulguları değerlendirilmiştir. LoRA tekniği ile yapılan ince ayar işleminin ve kullanılan veri kümesi boyutunun, tüm değerlendirilen BDM ailelerinin Türkçe SC başarımı üzerinde belirgin ve pozitif bir etkisi olduğu gözlemlenmiştir. Bu durum, sınırlı miktarda alan özgü veriyle bile olsa, PEFT yöntemleri aracılığıyla yapılan ince ayarın model uyarılmasında temel bir fayda sunduğunu göstermektedir.

Şekil 2’de görüldüğü üzere ince ayar yapılmamış ana modellerden en iyi başarıma 0,6827’lik skorla Gemma, en düşük başarıma 0,6683 skoruyla Mistral modeli sahiptir. İnce ayar ve artan veri kümesi boyutunun SBert Benzerlik skorları üzerindeki pozitif etkileri Şekil 2’de açıkça gösterilmiştir. nutuk_max_dataset ile yapılan ince ayar sonucunda en yüksek SBert Benzerlik skoruna (0.7263) ulaşan Gemma-2-9B modeli, ana modeline kıyasla yaklaşık %6.39’luk bir gelişim sergilemiştir. Benzer şekilde Meta-Llama 3.1 8B modeli nutuk_max_dataset ile SBERT Benzerlik skorunda ana modele göre yaklaşık %5.74’lük bir artış göstermiştir. Mistral-7B, yaklaşık %6.97 ve Qwen3-8B modeli de yaklaşık %6.83 ile SBERT Benzerlik metriğinde başarımları artışı göstermişlerdir.



Şekil 2. BDM Ana Modellerin ve İnce Ayarlanmış Modellerinin SBert Benzerlik Skoru Karşılaştırması

Şekil 3’te görüldüğü üzere BERTScore F1 metriklerinde de ince ayar ve artan veri kümesi boyutunun olumlu etkileri incelendiğinde tüm modeller en yüksek skorlara nutuk_max_dataset ile ulaşmıştır.



Şekil 3. BDM Ana Modellerin ve İnce Ayarlanmış Modellerin BERTSkor F1 Karşılaştırması

Her bir modelin, ana modellerine oranla iyileşme oranlarını ele alırsak bu ölçütlerden en yüksek BertSkor F1 değerine (0.6373) ulaşan Gemma-2-9B modeli %23.46’lık, en düşük başlangıç BertSkor F1 değerine sahip Qwen3-8B

modeli %29.38’lık, Mistral-7B modeli %19.70’lik ve Meta-Llama-3.1-8B modeli ana modelde en yüksek skora ulaşmasına rağmen ince ayarlama işlemi sonrasında ise %8.76’lık bir başarımları artışı sergilemiş, nihai skor değeri diğerlerinin çok az gerisinde kalmıştır. Tüm modeller, artan veri miktarıyla Bert Score F1 başarımlarını yükseltmiş, ince ayar sonrası modellerin BertSkor F1 değerleri birbirine yakınsamıştır.

Bert Score F1 metriğinde, Qwen3-8B ve Mistral-7B modelleri de Gemma’ya oldukça yakın ve güçlü birer başarımları sergilemiştir. Llama-3.1-8B ise, ana (instruct) model olarak rekabetçi bir başlangıç yapsa da ince ayar sonrası bu iki ana metrikteki nihai skorlar açısından diğerlerine kıyasla bir miktar geride kalmıştır.

Başarım Çözümlemesi ve Süre Değerlendirmesi için ROUGE-L ve genel F1 skorlarında kaydettiği yüzdesel gelişimler de Çizelge I’de incelenmiş her modelin kendi ana versiyonuna kıyasla ince ayarlanmış modellerin elde ettiği oransal artışları özetlenmiştir.

ÇİZELGE I. ROUGE-L VE GENEL F1 SKORLARININ YÜZDESEL DEĞİŞİMLERİ

Model Ailesi	ROUGE-L (Max Dataset FT vs Ana (Instruct) Model) Artış %	F1 Score (Max Dataset FT vs Ana (Instruct) Model) Artış %
Gemma-2-9B	%83,19	%81,54
Llama-3.1-8B	%24.02	%14.05
Mistral-7B	%63,52	%34,49
Qwen3-8B	%122,35	%104,78

Modellerin cevap süresi değerlendirildiğinde Çizelge II ‘de görüldüğü üzere *max_nutuk_dataset* ile yapılan ince ayar sonrasında, tüm modellerdeki belirgin başarımları artışının yanı sıra modellerin çıkarım sürelerinde de yaklaşık %55 ile %80 arasında belirgin bir hızlanma gözlemlenmiştir, özellikle Qwen3-8B modeli yaklaşık %79,44 oranında hızlanma ile en yüksek oranda iyileşme sergilemiştir.

ÇİZELGE II. MODELLERİN SORU BAŞINA ORTALAMA YANITLAMA SÜRESİ (s)

Model Adı	Ana Model	İnce Ayarlanmış Model(max_dataset)	İyileşme oranı
Llama 3.1-8B	33,84	15,24	%54,96
Gemma-2-9B	76,20	16,80	%77,95
Qwen-3-8B	87,00	17,88	%79,44
Mistral_v0.3-7B	64,80	21,24	%67,22

Bu bulgular, yapılan ince ayarın ve sonraki Q8_0 GGUF dönüşümünün, modellerin yalnızca yanıt kalitesini değil, aynı zamanda çıkarım süre verimliliğini de önemli ölçüde artırabileceğini ortaya koymaktadır. Bununla birlikte, her

modelin bu iyileşmeden faydalanma derecesi ve nihai başarımlar-süre dengesi farklılık göstermektedir.

V. SONUÇ

Bu çalışma, dört farklı açık kaynak BDM'nin Nutuk metni bağlamında Türkçe soru cevaplama başarımını, farklı boyutlardaki veri kümeleri ve LoRA ile yapılan ince ayar uyarlamaları özelinde kapsamlı bir şekilde çözümlemiştir. Elde edilen temel bulgular, LoRA tabanlı ince ayarın ve özellikle artan veri kümesi boyutunun, değerlendirilen tüm modellerde SBert Benzerlik ve Bert Skor F1 gibi anahtar ölçütler üzerinde anlamlı başarımlar iyileştirmeleri sağladığını ortaya koymuştur. Bu sürecin sonunda, Gemma-2-9B modeli, nutuk_max_dataset ile yapılan ince ayar neticesinde farklı ölçüm kriterlerinde de en yüksek başarımları sergileyerek bu spesifik görev için öne çıkmıştır. Bununla birlikte, Qwen3-8B ve Mistral-7B gibi diğer modellerin de, özellikle en kapsamlı veri kümesiyle, Gemma-2-9B'ye yakın ve rekabetçi sonuçlar elde etmesi, uygun ince ayar stratejileri ve yeterli veri ile farklı mimarilerinin de birbirine yakın başarımlar potansiyeli taşıdığını göstermektedir.

Nutuk metninden 8 farklı soru tipini hedefleyen sentetik veri üretim yönteminin etkinliği, modellerin ince ayar sonrası gösterdiği belirgin başarımlar artışıyla dolaylı olarak doğrulanmaktadır. Hedefe yönelik sentetik verinin BDM ince ayar uyarlanmasında ve kıyaslama sürecinde değerli bir araç olabileceği görüşü çalışmamızın sonuçları ile paraleldir. Bu sonuçlar, Türkçe doğal dil işleme ve tarihi metin çözümlemesi gibi özelleşmiş alanlarda BDM'lerin ince ayar uyarlanmasının başarımlar potansiyelinin yüksek olduğunu ve parametre verimli ince ayar yöntemlerinin kaynak kısıtlı senaryolarda dahi değerli sonuçlar üretebileceğini göstermektedir. Çalışmamız, aynı zamanda, hedefe yönelik sentetik veri üretiminin ve veri kümesi boyutunun ölçeklendirilmesinin, bu uyarlanma sürecindeki kritik rolünü teyit etmiştir. Mevcut çalışmada başarılı sonuçlar veren sentetik veri üretim stratejisinin, üretilen soruların bilişsel derinliğini ve çeşitliliğini daha da artıracak yeni prompt mühendisliği teknikleri veya farklı üretici model konfigürasyonları ile zenginleştirilmesi potansiyel bir adımdır. Ayrıca, veri kümesi ölçeklendirmesinin bu çalışmada gözlemlenen pozitif etkilerinin optimal sınırlarının ve çok daha büyük alan özgü veri kümeleri eklendiğinde elde edilecek ek başarımlar düzeyinin araştırılması da değerli olacaktır. Bütün bunlardan yola çıkarak BDM'lerin sınırlı veriye sahip farklı alanlarda ince ayar uygulamaları sayesinde başarımlarının artırılabilirliği görülmüştür. Son olarak, farklı mimarilere sahip yeni nesil açık kaynak BDM'lerin benzer bir metodolojiyle Türkçe uyarlanma potansiyellerinin ve başarımlar-süre dengelerinin karşılaştırmalı çözümlemeleri, bu alandaki pratik uygulamalar için de öngörü niteliğinde olacaktır.

BİLGİLENDİRME

Bu çalışma; firmamızın "Dijital Sözleşme Süreç Yönetimi ve Yapay Zekâ Destekli Çözümler" konulu TÜBİTAK TEYDEB projesi için yapılabirlik çalışmaları niteliği taşımakta; sözleşmelerin yapay zekâ destekli soru cevap bileşeni için alan özgü kısıtlı veri kaynağına sahip alanlarda ince ayar uygulamalarının çözümleme ve

raporlama tarafında etkin bir şekilde rol alabileceğini göstermektedir.

Yapılabilirlik çalışmalarımız kapsamında modellerin başarımlarını ölçmek üzere oluşturduğumuz kıyaslama veri kümesi akademik çalışmalara katkı sağlamak amacıyla https://huggingface.co/datasets/dijisoz23/nutuk_final_benchmark_data bağlantısında açık olarak erişime sunulmuştur.

KAYNAKÇA

- [1] Z. Hu, L. Wang, Y. Lan, W. Xu, E.-P. Lim, L. Bing, X. Xu, S. Poria, and R. K.-W. Lee, "LLM-Adapters: An adapter family for parameter-efficient fine-tuning of large language models," in Proc. 2023 Conf. on Empirical Methods in Natural Language Processing (EMNLP), Singapore, Dec. 2023, pp. 5254-5276. [Çevrimiçi]. Erişim Adresi: <https://doi.org/10.48550/arXiv.2304.01933>
- [2] Google, "gemma-2-9b-it," Hugging Face – Model Card. [Çevrimiçi]. Erişim Adresi: <https://huggingface.co/google/gemma-2-9b-it>. [Erişim Tarihi: Apr. 29, 2025].
- [3] Meta AI, "Meta-Llama-3.1-8B-Instruct," Hugging Face – Model Card. [Çevrimiçi]. Erişim Adresi: <https://huggingface.co/meta-llama/Meta-Llama-3.1-8B-Instruct>. [Erişim Tarihi: Apr. 29, 2025].
- [4] Mistral AI, "Mistral-7B-Instruct-v0.3," Hugging Face – Model Card. [Çevrimiçi]. Erişim Adresi: <https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3>. [Erişim Tarihi: Apr. 29, 2025].
- [5] Qwen Team, Alibaba Cloud, "Qwen/Qwen3-8B," Hugging Face – Model Card. [Çevrimiçi]. Erişim Adresi: <https://huggingface.co/Qwen/Qwen3-8B>. [Erişim Tarihi: Apr. 29, 2025].
- [6] M. K. Atatürk, *Nutuk – Söylev, Cilt I: 1919–1920*, 10. Baskı, Yay. Haz. İ. Arar, U. İğdemir ve S. N. Özerdim. Ankara: Türk Tarih Kurumu, 2019. [Çevrimiçi]. Erişim Adresi: <https://ekitap.ktb.gov.tr/Eklenti/76279,i-ciltpdf.pdf?0>. [Erişim Tarihi: Apr. 28, 2025].
- [7] E. Budur, R. Özçelik, D. Soylu, O. Khatat, T. Güngör, & C. Potts (2024). Building efficient and effective OpenQA systems for low-resource languages. Knowledge-Based Systems, 302, Article 112243. <https://doi.org/10.1016/j.knsys.2024.112243>.
- [8] B. Zhang, Z. Liu, C. Cherry, and O. Firat, "When scaling meets LLM finetuning: The effect of data, model and finetuning method," presented at the Int. Conf. on Learning Representations (ICLR), Vienna, Austria, 2024. [Çevrimiçi]. Erişim Adresi: <https://doi.org/10.48550/arXiv.2402.17193>
- [9] T. Liu, Y. Chen, Y. Chen, D. Yan, Z. Lin and J. Luo, "Comparison of LLMs in Extracting Synthesis Conditions and Generating Q&A Datasets for Metal–Organic Frameworks," ChemRxiv, Preprint, May 2024. [Çevrimiçi]. Erişim Adresi: <https://doi.org/10.26434/chemrxiv-2024-nt7vw>
- [10] I. Wu, S. Jayanthi, V. Viswanathan, S. Rosenberg, S. Pakazad, T. Wu, and G. Neubig, "Synthetic multimodal question generation," in Findings of the Association for Computational Linguistics: EMNLP 2024, 2024, pp. 12960-12993. [Çevrimiçi]. Erişim Adresi: <https://doi.org/10.48550/arXiv.2407.02233>
- [11] S. Maged, A. ElMaghraby, A. Marzbani, M. Essawey, A. Ahmed, E. Negm, and W. H. Gomaa, "HistoryQuest: Arabic question answering in Egyptian history with LLM fine-tuning and transformer models," in Proc. 2024 Intelligent Methods, Systems, and Applications (IMSA), July 2024, pp. 135-140. [Çevrimiçi]. Erişim Adresi: <https://doi.org/10.1109/IMSA61967.2024.10652824>
- [12] UnslothAI, "unsloth," GitHub Repository. [Çevrimiçi]. Erişim Adresi: <https://github.com/unslothai/unsloth>. [Erişim Tarihi: Apr. 30, 2025].
- [13] apxml.com. (t.y.). Model formats - GGUF. Getting Started with Local LLMs. [Çevrimiçi]. Erişim Adresi: <https://apxml.com/courses/getting-starting-local-llms/chapter-3-finding-selecting-local-llms/model-formats-gguf> Erişim tarihi: Erişim: 20 Haziran 2025
- [14] Google, "Google Colaboratory." [Çevrimiçi]. Erişim Adresi: <https://colab.research.google.com/>. [Erişim Tarihi: 30 Nisan 2025].

- [15] Google, "Gemini API Documentation - Models," Google AI for Developers. [Online]. Available: <https://ai.google.dev/gemini-api/docs/models?hl=tr>. [Eriřim Tarihi: 30 Nisan 2025].
- [16] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP), 2019, pp. 3982-3992. [Çevrimiçi]. Eriřim Adresi: <https://arxiv.org/abs/1908.10084>
- [17] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," presented at the Int. Conf. on Learning Representations (ICLR), 2020. [Çevrimiçi]. Eriřim Adresi: <https://arxiv.org/abs/1904.09675>
- [18] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," in Proc. Workshop on Text Summarization Branches Out (ACL), Barcelona, Spain, 2004, pp. 74-81.
- [19] Y. Liu, Y. Yuan, C. Zhang, Y. Wang, Y. Fu, and C. Xiong, "Beyond Accuracy: A Practical Guide to G-Eval and F1-Score for Evaluating LLMs," arXiv preprint arXiv:2404.09135, 2024. [Çevrimiçi]. Eriřim Adresi: <https://arxiv.org/abs/2404.09135>