

Türkçe gibi Az Kaynaklı Diller için Büyük Dil Modeli Tutarlılığının Min-P ve Top-P Örneklem Parametreleri ile Analizi

Analysis of Consistency of Large Language Models for Low-Resource Languages like Turkish with Min-P and Top-P Sampling Parameters

Taha Üzümcü
VeriUs Teknoloji
Marmara Üniversitesi
taha.uzumcu@verius.com.tr

Murat Can Ganiz
VeriUs Teknoloji
Marmara Üniversitesi
murat.ganiz@verius.com.tr

Özetçe —Büyük Dil Modelleri (BDM), yüksek sıcaklık parametrelerinde örneklem yaparken Türkçe gibi az kaynaklı dillerde dil tutarlılığını korumakta zorlanmaktadır. Bu çalışmada, yakın zamanda tanıtılan ve düşük olasılıklı kelimeleri (token) filtreleyen min-p ve top-p parametre değerlerinin, ağırlıklı olarak İngilizce eğitimi almış açık kaynak BDM’lerde Türkçe metin üretimi üzerindeki etkisi incelenmiştir. Min-p’nin Türkçe tutarlılığını farklı sıcaklık ve top-p ayarlarında koruma etkinliği, Yüksek Mahkeme karar özetleri kullanılarak değerlendirilmiştir. Ayrıntılı deneyler sonucunda, min-p örnekleme yönteminin yüksek sıcaklıklarda dilsel tutarlılığı önemli ölçüde artırdığı ve tutarlılıktan ödün vermeden daha fazla yaratıcılığa olanak tanıdığı gözlemlenmiştir.

Anahtar Kelimeler—Büyük Dil Modelleri, Az Kaynaklı Diller, Türkçe Metin Üretimi, Sıcaklık Örneklemesi, Min-P, Top-P, Örneklem Stratejileri

Abstract—Large Language Models (LLMs) struggle to maintain language consistency in low-resource languages like Turkish when sampling at high temperature parameters. This study investigates the effects of recently introduced min-p and top-p parameter values, which filter low-probability tokens, on Turkish text generation in open-source LLMs trained predominantly on English. The effectiveness of min-p in maintaining Turkish consistency across different temperature and top-p settings is evaluated using Supreme Court decision summaries. Detailed experiments demonstrate that min-p sampling significantly increases linguistic consistency at high temperatures and allows for greater creativity without compromising consistency.

Keywords—Large Language Models, Low-Resource Languages, Turkish Text Generation, Temperature Sampling, Min-P, Top-P, Sampling Strategies

I. GİRİŞ

Büyük dil modelleri (BDM) metin üretme konusunda olağanüstü performans göstermektedir ancak bu modellerin az

kaynaklı dillerdeki performanslarında çeşitli sorunlar bulunmaktadır [1]. Bu sorunlardan biri örneklem stratejileri ile ilgilidir. Sıcaklık, BDM çıktılarındaki rastgelelik veya "yaratıcılık" seviyesini kontrol eden bir örneklem parametresidir [2]. Bu parametre, metin üretiminde kelime (token) seçimi sırasında olasılık dağılımını etkiler. Sıcaklık değerinin artması, modelin daha çeşitli ve zengin içerik üretmesini sağlarken, aynı zamanda hatalara da yol açabilir [3]. Model daha az olası kelimeleri seçtikçe dil bilgisi hataları, tuhaf ifadeler ve söz dizimsel hataların olasılığı artar [2].

Bu durum, ana amacı İngilizce metin üretme olarak eğitilmiş ancak eğitim kümesinde daha az miktarlarda başka dillerden de metinler içeren LLama, Gemma gibi açık kaynak BDM’lerde İngilizce dışı metin üretmede farklı sorunlara yol açmaktadır. Daha ayrıntılı olarak, yüksek sıcaklık değerleri, bu tarz açık kaynak BDM’lerin Türkçe gibi az kaynaklı dillerde metin üretirken İngilizce veya diğer dillerden kelimeler yerleştirme eğilimini artırmaktadır [4].

BDM’ler genellikle İngilizce ağırlıklı veri kümeleriyle eğitilir. Bu nedenle, Türkçe gibi dillerin token dağılımları, modelin dahili temsilde daha düşük olasılıklara sahiptir. Yüksek sıcaklık, daha az yaygın tokenların seçilmesini teşvik etse de, İngilizce’nin veri fazlalığı bu etkiyi bastırabilir. Yüksek sıcaklık ayarlarında, modeller Türkçe gibi az kaynaklı dillerin düşük olasılıklı kelimelerini üretmektense, araya baskın dilin düşük olasılıklı kelimelerini koyabildiği gözlemlenmiştir [4]. Sıcaklık arttıkça, model dil sınırlarını koruma yeteneğini kaybeder. Öte yandan diller arasındaki sınırlar alana göre muğlak olabilir. Örneğin, özellikle bilgisayar bilimleri gibi teknik alanlardaki Türkçe metinlerde birçok İngilizce kelime geçmektedir. Bu tür metinler eğitim kümelerinde yer aldıklarında modelin Türkçe metin üretimi içinde İngilizce kelime üretme davranışını öğrenmesine yol açabilir. Bu sorunu çözmek için çeşitli yöntemler önerilebilir. Model, Türkçe metin üretiminde düşük sıcaklık değerleri ile kullanılabilir ancak bu, özellikle yaratıcı yazma gerektiren uygulamalarda modelin farklı metinler üretmesini

kısıtlayacaktır. Bir başka çözüm yöntemi olarak modeller sadece Türkçe metinlerle ince ayarlama eğitimine tabi tutulabilir. Bu durumda ise özellikle on milyarlarca hatta yüz milyarlarca parametreye sahip BDM'lerde modelin yüksek sıcaklıklardaki davranışını değiştirebilmek için çok büyük miktarlarda veri gerekliliği ve bununla bağlantılı çok pahalı hesaplama kaynakları ihtiyacı önemli bir sorun oluşturmaktadır.

Bu çalışmada Türkçe bağlamında, daha özel olarak Türkçe metin özetleme için metin üretimi bağlamında dil dışı kelime hatalarını analiz etmek ve azaltmak için min-p ve top-p örnekleme parametreleri incelenmiştir. Top-k ve top-p gibi geleneksel örnekleme yaklaşımlarının aksine, min-p düşük olasılıklı kelimeleri dinamik bir şekilde elimine eder. Bu sayede çeşitliliği ve yaratıcılığı korurken, dil dışı kelimelerin örneklenmesini engeller [3]. Çalışmada min-p parametresinin etkisinin değerlendirilmesi için Türkçe Yargıtay kararlarından ve özetlerinden oluşan metinler kullanılmıştır. Bu veri kümesi Türkçe'nin zorluklarına ek olarak, içerisinde hukuk alanına özel, günlük Türkçede çok nadir kullanılan kelimeleri barındırdığından ve özetler metin uzunluğuna kıyasla çok kısa olduğundan metin üretiminin öğrenilmesi açısından önemli zorluklar içermektedir.

Ana katkılar şu şekilde özetlenebilir:

- Sıcaklık, top-p ve min-p değerlerinin Türkçe metin üretiminde dil dışı kelime üretimi sorunu bağlamında sistematik bir şekilde incelenmesi yapılmıştır.
- Min-p yaklaşımının Türkçe metin üretiminde kararlılığı artırma etkisi analiz edilmiştir.
- Farklı örnekleme ayarları ile modelin yapmış olduğu hata tiplerini ortaya konmuştur.

II. ALAKALI ÇALIŞMALAR

Meta firması tarafından geliştirilen LLaMa 3 açık kaynak BDM ailesi özellikle 405 milyar parametreye kadar çıkan modelleri ile kendisini diğer model ailelerinden ayırır. 128 bin kelimeye (token) kadar bağlam penceresi vardır. Bu çalışmada sırasıyla 3 milyar ve 8 milyar parametrelili LLaMa-3.2-3B ve LLaMa-3.1-8B modelleri kullanılmıştır. Önemli bir not olarak LLaMa dil modeli ailesinin İngilizce dışında 7 dili desteklediği belirtilmiştir. Fakat bu diller arasında Türkçe yoktur [5].

Microsoft Phi-4, ise 14 milyar parametrelili bir dil modelidir. 16 bin kelimeye (token) kadar bağlam penceresi vardır. Eğitim kümesi içerisinde ki dillerin değerlendirilmesinde fastText [6] kullanılmış ve 176 adet dil tespit edilmiştir. Benzer şekilde Türkçe resmi olarak listelenen dillerin arasında değildir [7].

Sıcaklık parametresi örnekleme alanında çeşitliliği arttırmak için sıkça kullanılan yöntemlerden bir tanesidir. Düşük sıcaklık değerlerinde model daha kesin cevaplar verme eğilimindeyken, yüksek sıcaklıklarda kelime (token) çeşitliliği artmaktadır. Ancak bu çeşitlik modelin kararlılığını düşürmektedir. [3], [8].

Top-k örnekleme yöntemi BDM modelinin seçimlerini olasılığı en yüksek olan k seçenek ile sınırlandırmakta, ve düşük olasılıkla çıktılarını elemektedir. Nucleus (top-p) yöntemi ise, dinamik olarak kümülatif şekilde en yüksekten en düşük olasılığa doğru toplar ve belli bir eşğin altında kalan tokenları elimine eder. Bu iki yöntemde metnin akıcılığı iyi yönde

etkilese de, dil dışı tokenların örneklenmesini tam anlamıyla önlemez [8].

Min-p örnekleme, düşük olasılıklı tokenları en yüksek olasılıklı tokenın değerine göre ölçekleyip ayıklar. Yapılan araştırmalarda Min-p yönteminin dil içi kararlılığını etkisi incelenmiştir [3]. Bu çalışmada ise bu üst parametrelerin Türkçe metin üretimindeki tutarlılığa olan etkisi incelenmektedir.

Çok dilli modeller genellikle az kaynaklı dillerde veri kıtlığı ve dengesiz eğitimden kaynaklı zorluk çekmektedirler [1]. Bu modeller özellikle yüksek sıcaklıklarda, kararsız kaldıklarında yüksek-kaynaklı dillerin kelimelerinden örnekleme yapmaktadırlar. Var olan araştırmalar bu problemin ince-ayarlama ve pekiştirmeli-öğrenme yöntemleri ile performans artışı incelemiştirler [9], [10].

Az kaynak modelleri inceleyen bir çalışmada, tek dilli soru cevaplarda, en gelişmiş BDM'lerin bile aynı dilde metin üretmekte zorlandığı, özellikle LLaMa ve Mistral modellerinin bu konuda zayıf olduğu belirtilmiştir [4]. LLaMa3 modelleri düşük parametrelerde ortalama %18-%33 kararlılık sağlarken, Mistral Modelleri %70-%90 arası kararlılık sağlamaktadır .

Dil karışıklığını niceliksel olarak ölçmek için Dil Karışıklık Kıyası (Language Confusion Benchmark) geliştirilmiştir [4]. Bu kıyas, model cevaplarının hem satır (LPR) hem de kelime düzeyinde (WPR) ölçülmesini hedefler. Bu metrikleri kullanarak BDM çıktılarının istenilen dilde cevap verip vermediği niceliksel olarak ölçülebilmektedir. Testler dışında yazarlar aynı zamanda bu problemin önlenmesini de araştırmaktalar. Tek dil testinde birkaç az örnek (few-shot) yönteminin büyük oranda sorunu azalttığı, aynı zamanda hedef dil veri kümeleri ile ince-ayarlama ve tercih-ayarlama eğitimi yapmanın modeli istenilen dilde daha iyi hizalandığını gözlemlemiştirler. Bu yöntemlerin problemi kısmen çözdiği ve en güçlü GPT-4 gibi modellerin bile ara sıra hata yaptığını ve bu problemin hala geliştirmeye açık olduğu not edilmiştir [4].

III. YÖNTEM

Bu çalışmada Yargıtay içtihatları ve bunların kısa özetlerini içeren bir veri kümesi kullanılmıştır. Bu veri kümesi 290 içtihat ve özet ikilisinden oluşmaktadır. Veri kümesi hukuki terminoloji ve nadir kelime kullanımı nedeniyle BDM'ler için zorluklar sunmaktadır. İchtihat özetleri içtihat metnine kıyasla oldukça kısadır ve gerek dil kullanımı gerekse kelimeler itibarıyla genel içtihat metninden farklılık göstermektedir. Bu açıdan bu çalışmamızda odaklandığımız sorunlar açısından ideal bir veri kümesi olarak değerlendirilmektedir çünkü metin içinden birebir cümleler ve kelimeler ile bu özetin üretilmesi oldukça zordur.

Bu çalışma için Türkiye'de gerek endüstri gerekse akademia yaygın olarak kullanılan, resmi olarak Türkçe desteklemeyen ama Türkçe metin üretimi kabiliyeti olan LLaMa ve Phi ailesi açık kaynak modeller seçilmiştir. Bu modeller kullanılarak toplamda 400'den fazla deney yapılmıştır. LLaMa 3 milyar parametrelili model Türkçe metin üretimi açısından son derece zayıf olduğu, Phi 14 milyar parametrelili model ise deney sürelerini uzattığı ve bağlam uzunluğu göreceli olarak az olduğu için sadece LLaMa 8 milyarlık model üzerinden ilerlenmiştir. Temel olarak sıcaklık, top-p ve min-p olmak üzere üç farklı örnekleme üst parametresi incelenmiştir (TABLO II: Örnekleme Üst Parametreleri ve Aralıkları).

Model	Boyut	Bağlam	Firma	Çıkış
Microsoft Phi-4	14B	16K	Microsoft	Aralık 2024
Meta LLaMA-3.2-3B	3B	128K	Meta	Eylül 2024
Meta LLaMA-3.1-8B	8B	128K	Meta	Temmuz 2024

TABLO I: Modeller

Parametre	Min	Max	Açıklama
Sıcaklık	0.3	4.0	Olasılık dağılımını kontrol eder
Top-p	0.4	1.0	Kümülatif olarak olasılıkları sınırlar
Min-p	0.0	0.4	Düşük olasılıklı tokenları elimine eder

TABLO II: Örnekleme Parametreleri ve Aralıkları

Modeller (TABLO I: Modeller) vLLM [11], kütüphanesi ile çalıştırılmıştır. Değerlendirme için özetleme literatüründe sık kullanılan ROUGE-1, ROUGE-2, ve ROUGE-L puanları kullanılmıştır. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) puanı, kelime sekanslarında n parametresi ile farklı büyüklüklerde n-gram örtüşmelerini ölçen, ve L parametresi ile en uzun örtüşen alt dizinleri hesaplayan bir ölçüttür [12].

ROUGE puanı dışında aynı zamanda Dil Karmaşıklık Kıyası (Language Confusion Benchmark - LCB) incelenmiştir [4]. Bu kıyasta Line-level Pass Rate (LPR) satır bazında dilin hedef dil olup olmadığını kontrol ederken, Word-level Pass Rate (WPR) satırın içerisindeki kelimelerin Türkçe olup olmadığını kontrol etmekte ve hata oranlarını hesaplamaktadır. LPR puanı Türkçe bir cümle içerisindeki az sayıda farklı dilden kelime olsa dahi çoğunluk Türkçe kelimelerden oluşuyorsa olumlu sonuç vermektedir. WPR puanı ise kelime bazında olup Türkçe desteklememektedir. WPR puanı hesaplamaya Türkçe desteği eklemek için temel bir Türkçe kelime listesine ilave olarak içtihatlarından çıkartılan kelimeler, özel isimler ayıklandıktan sonra FPS5 [13], [14] kök bulma fonksiyonundan geçirilerek elde edilen kelime listesi kullanılmıştır.

IV. BULGULAR VE TARTIŞMA

Yapılan Deneyler modellerin üç ana şekilde bozulduğunu ortaya koymaktadır.

- 1) **Küçük Yazım Hataları** – Küçük yazım hataları veya dil bilgisi tutarsızlıkları (Bkz. Şekil 1).
- 2) **Dil Dışı Token Ekleme** – Bazı kelimeler farklı dilde geçmektedir ancak çevrildiğinde doğru anlamını korumaktadır (Bkz. Şekil 2).
- 3) **Anlamsız Token Çıktısı** – Anlamsız karakterler vermektedir (Bkz. Şekil 3).

Girdi:Aşağıdaki metni özetle: YARGITAY İLAMI Yerel mahkemece verilen hüküm temyiz edilmekle; başvurunun nitelik, ceza türü, süresi ve suç tarihine göre dosya okunduktan sonra Türk Milleti adına gereği görüşülüp düşünüldü...

Çıktı: ...beraatlerine hüküm verilmeside sanıglara çok yanlış...

Şekil 1: Küçük Yazım Hataları

Model kararlı bir şekilde cevap verdiğinden, III numaralı tabloda görüldüğü gibi, top-p ve min-p değerleri modelin puanını anlamlı bir şekilde etkilememiştir.

Girdi:Aşağıdaki metni özetle: Yerel Mahkemece verilen hükümler temyiz edilmekle, başvurunun süresi ve kararın niteliği ile suç tarihine göre dosya görüşüldü: Temyiz isteğinin reddi nedenleri bulunmadığından işin esasına geçildi. Vicdani kanının oluştuğu duruşma sürecini yansıtan tutanaklar...

Çıktı: Temyiz kararı değerlendirildi. Açıklanan durumla ilgili karar如下 - Başvuru tarihi ve suça özgürlük tarihine bağlı olarak temyiz değerlendirilmesi yapıldı...

Şekil 2: Dil Dışı Token Ekleme

Girdi:Aşağıdaki metni özetle: Bölge Adliye Mahkemesi Ceza Dairesince verilen hükümler temyiz edilmekle; dosya incelendi, gereği görüşülüp düşünüldü:

Çıktı: getMessage_amccion.me covereer deallocorow initializing(reference Hearts quatizibette mine GitHubFebruary Nancyivistprogramming)

Şekil 3: Anlamsız Token Çıktısı

IV numaralı tabloda, 0.7 Sıcaklık değeri itibarıyla üretilen metnin Türkçesinin bozulmaya başladığı görülmektedir. Buradaki bozulmalar, genellikle küçük yazım hataları ve nadir dil dışı token ekleme şeklinde olmaktadır. III numaralı tablonun aksine, burada ki ufak top-p ve min-p değerleri modelin performansını düzeltmektedir.

Sıcaklık 1'e geldiğinde (Tablo V), model anlamsız metin üretmeye başlamıştır. Top-p ve min-p değerleri hala bu bozulmanın önüne geçebildiği gözlemlenmiştir.

Sıcaklık 1.5'e çıktığında (Tablo VI), top-p ve min-p değerleri tek başlarına kullanıldığında modelin anlamsız token çıktısı vermesini önlemekte, ancak küçük yazım hataları ve dil dışı token ekleme devam etmektedir. Model performansının düzelmesi için top-p ve min-p'nin birlikte kullanılması gerekmektedir.

Sıcaklık 1.75'e çıktığında (Tablo VII), top-p ve min-p birlikte kullanılmasına rağmen model performansının düzelmediği gözlemlenmiştir.

Sıcaklık 1.75'ten (Tablo VII) sonra model başarımı düşmeye devam etmektedir. Sıcaklık 3'e (Tablo VIII) geldiğinde ise örnekleme parametrelerinden bağımsız model anlamsız token çıktısı vermektedir.

V. SONUÇ

Türkçe metin üretiminde, ince ayarlama gibi maliyetlere katlanmadan, açık kaynak BDM'lerin yüksek sıcaklık değerlerinde, amaçlanan dil dışı token üretimi sorununun min-p ve top-p parametre değerinin ayarlanarak 1.5 sıcaklığına kadar tamamen, 2.5 sıcaklığına kadar kısmen kontrol altına alınabildiğini görülmektedir.

WPR ağırlıklı olarak kaynak derleme (örn. Hukuk) bağlı olduğundan, WPR bozuk kelimeler dışında aynı zamanda alan dışı kelimelere hassas hale gelmiştir. Dolayısıyla düşük WPR değerleri modelin tamamen bozulmaya uğradığına işaret

Top-P	Min-P	WPR	LPR	R-1	R-2	R-L
N/A	N/A	0.67	1.00	0.32	0.17	0.23
0.8	N/A	0.66	1.00	0.32	0.17	0.23
N/A	0.1	0.68	1.00	0.32	0.17	0.23
0.8	0.1	0.67	1.00	0.33	0.18	0.23

TABLO III: Sıcaklık 0.3 için top-p, min-p değerlerine karşılık gelen değerlendirme ölçütleri

Top-P	Min-P	WPR	LPR	R-1	R-2	R-L
N/A	N/A	0.53	1.00	0.31	0.14	0.20
0.8	N/A	0.66	1.00	0.32	0.16	0.22
N/A	0.1	0.68	0.99	0.15	0.21	0.11
0.8	0.1	0.70	1.00	0.32	0.17	0.22

TABLO IV: Sıcaklık 0.7 top-p, min-p değerlerine karşılık gelen değerlendirme ölçütleri

Top-P	Min-P	WPR	LPR	R-1	R-2	R-L
N/A	N/A	0.01	0.34	0.14	0.04	0.08
0.6	N/A	0.66	1.00	0.31	0.15	0.21
0.4	N/A	0.69	1.00	0.32	0.17	0.22
N/A	0.1	0.60	1.00	0.30	0.14	0.20
N/A	0.2	0.68	1.00	0.32	0.16	0.22
0.6	0.2	0.64	1.00	0.32	0.17	0.22

TABLO V: Sıcaklık 1.0 top-p, min-p değerlerine karşılık gelen değerlendirme ölçütleri

Top-P	Min-P	WPR	LPR	R-1	R-2	R-L
N/A	N/A	0.00	0.00	0.03	0.00	0.02
0.6	N/A	0.00	0.00	0.02	0.00	0.01
0.4	N/A	0.00	0.08	0.08	0.01	0.04
N/A	0.1	0.33	1.00	0.28	0.11	0.17
N/A	0.2	0.54	1.00	0.31	0.14	0.20
0.6	0.2	0.39	0.99	0.28	0.11	0.17
0.4	0.2	0.56	0.99	0.31	0.14	0.20

TABLO VI: Sıcaklık 1.50

Top-P	Min-P	WPR	LPR	R-1	R-2	R-L
N/A	N/A	0.00	0.00	0.03	0.00	0.02
0.4	N/A	0.00	0.00	0.04	0.00	0.02
N/A	0.1	0.21	1.00	0.26	0.08	0.15
N/A	0.2	0.44	1.00	0.29	0.12	0.18
0.6	0.2	0.47	1.00	0.29	0.12	0.18
0.4	0.2	0.47	1.00	0.28	0.11	0.18

TABLO VII: Sıcaklık 1.75

Top-P	Min-P	WPR	LPR	R-1	R-2	R-L
N/A	0.1	0.00	0.05	0.06	0.00	0.03
N/A	0.2	0.02	0.97	0.19	0.03	0.10
0.4	0.2	0.02	0.96	0.20	0.03	0.11

TABLO VIII: Sıcaklık 3

etmeyebilir. Örneğin, 290 karar özeti için yaklaşık 28 bin kelime üreten bir modelin (Sıcaklık 2.5, Top-p 0.4, Min-p 0.2) almış olduğu 0.1 WPR puanı incelendiğinde, gerçekten hatalı olarak bulunan 26 kelime şu şekildedir: processindeki, argumentü, yukarıdaki, mikroforu, however, feshte, someone, niteliikli, aboniye, inceledi, absence, feraat, muvaza, avukatının, dosalarla, etti"nin, reasonlar, case, tebligname, management, tesbelliinin, kanunu"un, iddiastan, davinin, companies, argumentleri.

TEŞEKKÜR

Bu çalışma TÜBİTAK TEYDEB 3231058 numaralı proje ile kısmi olarak desteklenmiştir.

KAYNAKLAR

- [1] S. Cahyawijaya, "LLM for Everyone: Representing the Underrepresented in Large Language Models", arXiv preprint arXiv:2409.13897, 2024.
- [2] M. Peeperkorn, T. Kouwenhoven, D. Brown, and A. Jordanous, "Is Temperature the Creativity Parameter of Large Language Models?", *arXiv preprint arXiv:2405.00492*, 2024. [Online]. Available: <https://arxiv.org/abs/2405.00492>
- [3] M. Nguyen, A. Baker, C. Neo, A. Roush, A. Kirsch, R. Shwartz-Ziv, "Turning Up the Heat: Min-p Sampling for Creative and Coherent LLM Outputs", arXiv preprint arXiv:2407.01082, 2024.
- [4] K. Marchisio, W.-Y. Ko, A. Bérard, T. Dehaze, S. Ruder, "Understanding and Mitigating Language Confusion in LLMs", arXiv preprint arXiv:2406.20052, 2024.
- [5] A. Grattafiori, A. Dubey, A. Jauhri, et al., "The Llama 3 Herd of Models", arXiv preprint arXiv:2407.21783, 2024.
- [6] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, "Bag of Tricks for Efficient Text Classification", arXiv preprint arXiv:1607.01759, 2016.
- [7] M. Abdin, J. Aneja, H. Behl, et al., "Phi-4 Technical Report", arXiv preprint arXiv:2412.08905, 2024.
- [8] A. Holtzman, J. Buys, L. Du, M. Forbes, Y. Choi, "The Curious Case of Neural Text Degeneration", ICLR, 2020.
- [9] O. Khade, S. Jagdale, A. Phaltankar, G. Takalikar, R. Joshi, "Challenges in Adapting Multilingual LLMs to Low-Resource Languages using LoRA PEFT Tuning", arXiv preprint arXiv:2411.18571, 2024.
- [10] H. Zhang, M. Liu, C. Li, Y. Chen, J. Xu, M. Zhou, "A Reinforcement Learning Approach to Improve Low-Resource Machine Translation Leveraging Domain Monolingual Data", LREC-COLING, 2024.
- [11] W. Kwon, Z. Li, S. Zhuang, Y. Sheng, L. Zheng, C. H. Yu, J. E. Gonzalez, H. Zhang, I. Stoica, "Efficient Memory Management for Large Language Model Serving with PagedAttention", ACM SIGOPS, 2023.
- [12] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", Text Summarization Branches Out, ACL, 2004.
- [13] F. Can, S. Kocberber, E. Balcik, C. Kaynak, H. C. Ocalan, and O. M. Vursavas, "Information retrieval on Turkish texts", *Journal of the American Society for Information Science and Technology*, vol. 59, no. 3, pp. 407–421, 2008.
- [14] D. Torunoğlu, E. Çakırman, M. C. Ganiz, S. Akyokuş, and M. Z. Gürbüz, "Analysis of preprocessing methods on classification of Turkish texts", in *Proc. 2011 International Symposium on Innovations in Intelligent Systems and Applications*, pp. 112–117, 2011.