

TÜBİTAK–****2209-A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI****

****ARAŞTIRMA ÖNERİSİ**** FORMU

2024 Yılı

Bahar Dönem Başvurusu

**A. GENEL BİLGİLER**

|  |
| --- |
| Başvuru Sahibinin Adı Soyadı: Emre Dinç – Batuhan Eymen Çorlu |
| **Araştırma Önerisinin Başlığı:** HerkeZ Hata Yapar |
| **Danışmanın Adı Soyadı:Bihter Daş** |
| **Araştırmanın Yürütüleceği Kurum/Kuruluş:Fırat Üniverstesi** |

**ÖZET**

Türkçe özetin araştırma önerisinin (a) özgün değeri, (b) yöntemi, (c) yönetimi ve (d) yaygın etkisi hakkında bilgileri kapsaması beklenir. Bu bölümün en son yazılması önerilir.

|  |
| --- |
| **Özet**  Proje önerisinin Özgün Değeri: Bu proje önerisi Large Language Model kullanılarak Türkçe metinlerde yazım yanlışlarını ve imla kuralı hatalarını bulup modeli TDK websitesinden çekilen verilere göre eğitip çıktılara uygun makalelerde ve diğer özgün yazılarda kullanıcılara kolaylık sağlama daha önce literatürde modelin gerçek verilerle eğitip doğruluk oranı olarak arttırılmış bir versiyonu bulunmamaktadır bu da proje önerisinin özgün değeri olarak kabul edilebilir.  Yaygın sosyal medya uygulamalarına entegre edilebilir olması kullanılabilirlik açısından çok önemli olmaktadır.  Proje önerisi, react kullanılarak server ve clientten oluşan bir uygulama olup server tarafında verilerin Large Language modelde işlenip kullanıcıya client tarafında sunulmasından oluşuyor.  Proje önerisinin yönetimi, uygun modelin bulunması clientte kullanıcıdan verilerin doğru şekilde alınıp işlenip servera gönderilip daha sonra bu verilerin işlenip modele verilmesinden oluşur.  Bu proje önerisi, makale yazıları mail yazıları sosyal medya yazıları veya herahangi bir türkçe metin yazısında kullanıcıların çokca işine yarayacaktır. Çok geniş bir kapsamı olmakla beraber önceden yazılmış metinlerde de kullanmak mümkün oluyor |
| **Anahtar Kelimeler:** TDK verileri, React, Yazım Hataları,Large Language Model |

1. **ÖZGÜN DEĞER**

**1.1. Konunun Önemi, Araştırma Önerisinin Özgün Değeri ve Araştırma Sorusu/Hipotezi**

Araştırma önerisinde ele alınan konunun kapsamı ve sınırları ile önemi literatürün eleştirel bir değerlendirmesinin yanı sıra nitel veya nicel verilerle açıklanır.

Özgün değer yazılırken araştırma önerisinin bilimsel değeri, farklılığı ve yeniliği, hangi eksikliği nasıl gidereceği veya hangi soruna nasıl bir çözüm geliştireceği ve/veya ilgili bilim veya teknoloji alan(lar)ına kavramsal, kuramsal ve/veya metodolojik olarak ne gibi özgün katkılarda bulunacağı literatüre atıf yapılarak açıklanır.

Önerilen çalışmanın araştırma sorusu ve varsa hipotezi veya ele aldığı problem(ler)i açık bir şekilde ortaya konulur.

|  |
| --- |
| Proje Önerisinin Özgün Değeri:  Günümüzde doğal dil işleme alanında yapılan araştırmalar, dil hatalarını tespit etme ve düzeltme konusunda heyecan verici gelişmelere yol açmıştır. Bu alandaki çalışmaların örneklerinden biri,(Hece Tabanlı İstatistiksel Yöntemler İle Yazım Hatası Bulma Ve Düzeltme)**1** Türkçe dilinde hece tabanlı n-gram modelleri üzerine odaklanmıştır.  Akademik literatürde, Türkçe dilinin biçimbirimleri açısından bitişken yapısından ötürü kelime tabanlı n-gram modellerin yetersiz kalabileceği belirtilmiştir. Bu nedenle, hece tabanlı n-gram modellerinin geliştirilmesi, Türkçe dilindeki yazım hatalarını belirleme ve düzeltme süreçlerinde daha etkili bir yaklaşım sunmuştur.  Ancak, mevcut çalışmaların ötesine geçerek daha gelişmiş bir yaklaşım benimsemek isteyenler için, büyük dil modelleri (large language models) önemli bir potansiyel sunmaktadır. Özellikle, OpenAI gibi platformlar tarafından geliştirilen ve geniş çapta dil anlama yeteneğine sahip olan bu modeller, dil hatalarını tespit etme ve düzeltme konularında yeni bir ufuk açabilir.  Kendi proje önerimde, bu akademik çalışmalardan ilham alarak, büyük dil modellerini kullanarak daha kapsamlı ve etkili bir yazım denetimi sistemi geliştirmeyi hedefliyorum. Bu modeller, geniş veri kümeleri üzerinde eğitildikleri için dilin çeşitliliğini ve yapısal özelliklerini daha iyi kavrayabilirler. Böylece, Türkçe dilindeki yazım hatalarını belirleme ve düzeltme süreçlerinde daha yüksek doğruluk sağlayabilirler.  Sonuç olarak, doğal dil işleme ve yapay zeka teknolojileri, dil hatalarının belirlenmesi ve düzeltilmesi alanında hızla ilerlemekte ve büyük dil modelleri gibi yenilikçi yaklaşımlar, gelecekte daha akıllı ve etkili dil hizmetlerinin geliştirilmesine imkan tanıyacaktır. |

* 1. **Amaç ve Hedefler**

Araştırma önerisinin amacı ve hedefleri açık, ölçülebilir, gerçekçi ve araştırma süresince ulaşılabilir nitelikte olacak şekilde yazılır.

|  |
| --- |
| Amaç:  Bu proje önerisinin temel amacı, Türkçe metinlerdeki yazım yanlışlarını ve imla kuralları hatalarını belirlemek ve düzeltmek için yapay zeka tabanlı bir çözüm geliştirmektir. Bu çözüm, Large Language Model (Büyük Dil Modeli) teknolojisi kullanılarak Türkçe yazım hatalarını tespit edip düzeltebilecek bir yazılım aracı oluşturmayı hedeflemektedir.  Hedefler:  Veri Toplama ve Hazırlık:  TDK sitesi üzerinden alınan Türkçe yazım kural verileri ve TDK apinin kullanılması ile beraber alınan doğru bilgilerin Large Language Model ile kullanılabilecek hale getirilmesi.  Model Eğitimi ve Geliştirme:  Large Language Model (Büyük Dil Modeli) kullanılarak Türkçe metinler üzerinde çalışacak bir model geliştirilecek.  Model, yazım hatalarını ve imla kuralları ihlallerini doğru bir şekilde tespit edebilmek için eğitilecek ve optimize edilecektir.  Doğrulama ve Performans Değerlendirmesi:  Eğitilen model, belirlenen bir doğruluk metriği üzerinde değerlendirilecek ve performansı test edilecektir.  Modelin doğruluğu ve hızı üzerinde çeşitli testler yapılacak ve gerekli iyileştirmeler uygulanacaktır.  Yazılım Aracının Geliştirilmesi ve Entegrasyonu:  Modelin entegre edilebileceği bir internet tarayıcı aracı geliştirilecek ve farklı platformlarda kullanılabilir hale getirilecektir.  Kullanıcıların kolayca erişebileceği bir arayüz tasarlanacak.  Kullanıcı Geri Bildirimi ve İyileştirme:  Kullanıcıların geri bildirimleri toplanacak ve modelin performansını artırmak için kullanılacaktır.  Yazılım aracının kullanıcı dostu olması ve etkili bir şekilde çalışması için sürekli olarak iyileştirmeler yapılacaktır.  Toplumsal Farkındalık ve Kullanımın Teşviki:  Proje önerisinin tanıtımı yapılacak ve Türkçe yazım hatalarının düzeltilmesinin önemi vurgulanacaktır. Sosyal medya üzerinde ve resmi mesajlaşma uygulamaları üzerinde kullanılabilirliği incelenecek ve kullanıcılara sunulacak  Kullanıcıların projeden en iyi şekilde faydalanmaları için eğitim materyalleri ve rehberler sunulacak ve kullanımın teşviki sağlanacaktır.  Bu hedeflerin başarıyla tamamlanması, projenin amacına ulaşmasını sağlayacak ve Türkçe yazım hatalarının azaltılmasına ve dilin daha doğru kullanılmasına katkıda bulunacaktır. |

1. **YÖNTEM**

Araştırma önerisinde uygulanacak yöntem ve araştırma teknikleri (veri toplama araçları ve analiz yöntemleri dahil) ilgili literatüre atıf yapılarak açıklanır. Yöntem ve tekniklerin çalışmada öngörülen amaç ve hedeflere ulaşmaya elverişli olduğu ortaya konulur.

Yöntem bölümünün araştırmanın tasarımını, bağımlı ve bağımsız değişkenleri ve istatistiksel yöntemleri kapsaması gerekir. Araştırma önerisinde herhangi bir ön çalışma veya fizibilite yapıldıysa bunların sunulması beklenir. Araştırma önerisinde sunulan yöntemlerin iş paketleri ile ilişkilendirilmesi gerekir.

|  |
| --- |
| 1. **Literatür Taraması:**    * Daha önce Türkçe veya başka dillerde bu proje önerisi ile alakalı yapılan akademik çalışmaların incelenmesi. 2. **Large Language Model Seçimi:**    * Proje önerisine en uyumlu olarak çalışabilecek Large language modeller taranacak.    * Ve bu LLM üzerinden kullanım senaryolarının araştırılması. 3. **Large Language Modeli Eğitimi:**    * Önceden eğitilmiş bir büyük dil modeli (örneğin, GPT) kullanarak TDK verileri üzerinde feyz alınacak.    * Eğitim için uygun bir derin öğrenme modeli seçilecek ve TDK verileri üzerinde eğitilecek.    * Eğitim süreci için uygun hiperparametreler belirlenecek ve modelin performansını artırmak için gerektiğinde hiperparametre ayarlamaları yapılacak. 4. **Web Extension Oluşturulması:**    * Modelimizin kullanılacağı kullanıcı dostu web extension oluşturulması. 5. **Proje Önerisinin Uygulaması:**    * Eğitilen model, gerçek dünya metinlerinde Türkçe dilbilgisel hataları tespit etmek ve düzeltmek için kullanılacak.    * Modelin geliştirilmiş performansı, gerçek kullanım senaryolarında değerlendirilecek ve sonuçlar gözlemlenecek. 6. **Sonuçların Değerlendirilmesi:**    * Modelin başarı ölçütleri belirlenecek ve performansı bu ölçütlere göre değerlendirilecek.    * Modelin doğruluğu, hızı, ve uygulanabilirliği gibi faktörler üzerinde analiz yapılacak. 7. **Sonuçların Sunumu ve Yaygınlaştırılması:**    * Proje sonuçları, raporlar ve sunumlar aracılığıyla paylaşılacak.    * Modelin yaygınlaştırılması için gerekli adımlar atılacak ve ilgilenen kullanıcılara erişim sağlanacak. |

1. **PROJE YÖNETİMİ** 
   1. **İş- Zaman Çizelgesi**

Araştırma önerisinde yer alacak başlıca iş paketleri ve hedefleri, her bir iş paketinin hangi sürede gerçekleştirileceği, başarı ölçütü ve araştırmanın başarısına katkısı “İş-Zaman Çizelgesi” doldurularak verilir. Literatür taraması, gelişme ve sonuç raporu hazırlama aşamaları, araştırma sonuçlarının paylaşımı, makale yazımı ve malzeme alımı ayrı birer iş paketi olarak gösterilmemelidir.

Başarı ölçütü olarak her bir iş paketinin hangi kriterleri sağladığında başarılı sayılacağı açıklanır. Başarı ölçütü, ölçülebilir ve izlenebilir nitelikte olacak şekilde nicel veya nitel ölçütlerle (ifade, sayı, yüzde, vb.) belirtilir.

**İŞ-ZAMAN ÇİZELGESİ (\*)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **İP No** | **İş Paketlerinin Adı ve Hedefleri** | **Kim(ler) Tarafından Gerçekleştirileceği** | **Zaman Aralığı**  **(..-.. Ay)** | **Başarı Ölçütü ve** **Projenin Başarısına Katkısı** |
| 1 | Literatür Taraması |  | 01-01 | Türkçe yazım hatalarının araştırılması bilgi edinilmesi |
| 2 | Large Language Model Seçimi |  | 02-02 | Large language model hakkında detaylı bilgi edinimi |
| 3 | Large Language Modeli Eğitimi |  | 03-03 | TDK üzerinden yazım kuralları ve kelime yazımlarının araştırılması |
| 4 | Web Extension oluşturulması |  | 04-04 |  |
| 5 |  |  |  |  |

(\*) Çizelgedeki satırlar ve sütunlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

* 1. **Risk Yönetimi**

Araştırmanın başarısını olumsuz yönde etkileyebilecek riskler ve bu risklerle karşılaşıldığında araştırmanın başarıyla yürütülmesini sağlamak için alınacak tedbirler (B Planı) ilgili iş paketleri belirtilerek ana hatlarıyla aşağıdaki Risk Yönetimi Tablosu’nda ifade edilir. B planlarının uygulanması araştırmanın temel hedeflerinden sapmaya yol açmamalıdır.

**RİSK YÖNETİMİ TABLOSU\***

|  |  |  |
| --- | --- | --- |
| **İP No** | **En Önemli Riskler** | **Risk Yönetimi (B Planı)** |
| 1 | Modelin ücretini projenin karşılayamaması | Daha uygun alternatiflere yönelinecek |
| 2 | Sağlanan modelin piyasadan çekilmesi | Alternatif Model Bulunacak |

(\*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

* 1. **Araştırma Olanakları**

Bu bölümde projenin yürütüleceği kurum ve kuruluşlardavar olan ve projede kullanılacak olan altyapı/ekipman (laboratuvar, araç, makine-teçhizat, vb.)olanakları belirtilir.

**ARAŞTIRMA OLANAKLARI TABLOSU (\*)**

|  |  |
| --- | --- |
| **Kuruluşta Bulunan Altyapı/Ekipman Türü, Modeli**  (Laboratuvar, Araç, Makine-Teçhizat, vb.) | **Projede Kullanım Amacı** |
|  |  |
|  |  |
|  |  |

**(\*)** Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

1. **YAYGIN ETKİ**

Önerilen çalışma başarıyla gerçekleştirildiği takdirde araştırmadan elde edilmesi öngörülen ve beklenen yaygın etkilerin neler olabileceği, diğer bir ifadeyle yapılan araştırmadan ne gibi çıktı, sonuç ve etkilerin elde edileceği aşağıdaki tabloda verilir.

**ARAŞTIRMA ÖNERİSİNDEN BEKLENEN YAYGIN ETKİ TABLOSU**

|  |  |
| --- | --- |
| **Yaygın Etki Türleri** | **Önerilen Araştırmadan Beklenen Çıktı, Sonuç ve Etkiler** |
| **Bilimsel/Akademik**  (Makale, Bildiri, Kitap Bölümü, Kitap) |  |
| **Ekonomik/Ticari/Sosyal**  (Ürün, Prototip, Patent, Faydalı Model, Üretim İzni, Çeşit Tescili, Spin-off/Start- up Şirket, Görsel/İşitsel Arşiv, Envanter/Veri Tabanı/Belgeleme Üretimi, Telife Konu Olan Eser, Medyada Yer Alma, Fuar, Proje Pazarı, Çalıştay, Eğitim vb. Bilimsel Etkinlik, Proje Sonuçlarını Kullanacak Kurum/Kuruluş, vb. diğer yaygın etkiler) |  |
| **Araştırmacı Yetiştirilmesi ve Yeni Proje(ler) Oluşturma**  (Yüksek Lisans/Doktora Tezi, Ulusal/Uluslararası Yeni Proje) |  |

**5. BÜTÇE TALEP ÇİZELGESİ**

|  |  |  |
| --- | --- | --- |
| **Bütçe Türü** | **Talep Edilen Bütçe Miktarı (TL)** | **Talep Gerekçesi** |
| **Sarf Malzeme** |  |  |
| **Makina/Teçhizat (Demirbaş)** |  |  |
| **Hizmet Alımı** | API | Large Language Modeli eğitebilmek için ücretli versiyonuna ihtiyaç var |
| **Ulaşım** |  |  |
| **TOPLAM** |  |  |

**NOT:** Bütçe talebiniz olması halinde hem bu tablonun hem de TÜBİTAK Yönetim Bilgi Sistemi (TYBS) başvuru ekranında karşınıza gelecek olan bütçe alanlarının doldurulması gerekmektedir. Yukardaki tabloda girilen bütçe kalemlerindeki rakamlar ile, TYBS başvuru ekranındaki rakamlar arasında farklılık olması halinde TYBS ekranındaki veriler dikkate alınır ve başvuru sonrasında değiştirilemez.

**6. BELİRTMEK İSTEDİĞİNİZ DİĞER KONULAR**

Sadece araştırma önerisinin değerlendirilmesine katkı sağlayabilecek bilgi/veri (grafik, tablo, vb.) eklenebilir.

|  |
| --- |
|  |

**7. EKLER**

**EK-1: KAYNAKLAR**

* + 1. https://polen.itu.edu.tr:8443/server/api/core/bitstreams/80570324-3bbe-4a9f-87ed-c11411f008f8/content