

Teknofest Türkçe Doğal Dil İşleme Yarışması – Serbest Kategori

TEAM: DATA DESTROYERS

Week 1

13 Temmuz 2024 – 20 Temmuz 2024

- Github Reposunun oluşturulması.
(Berke Sevim)
- Github dökümantasyonu için Github Readme.md dosyası oluşturulacak.
(Berke Sevim)
- Google Drive açılması
(Berke Sevim)
- Projeyle ilgili müşteri hizmetleri ses verilerinin araştırılması.
(Büşra Sulukan)
- Projeyle ilgili platformu oluşturmak için web arayüzlerinin araştırılması.
(Gözde Ceren Yıldız)
- Piyasada projeye benzer toolların araştırılması.
(Büşra Sulukan)
- Türk pazarının bu alandaki istek, ihtiyaçları veya açıklarının araştırılması
(Ortak)
- Projenin Türkçe diline ait katkılarının vurgulanması.
(Ortak)
- <https://huggingface.co/TURKCELL/Turkcell-LLM-7b-v1> modelinin test edilmesi.
(Berke Sevim)

Not: Model test edilemedi. Çok güçlü GPU gerektiren altyapı gerekiyor. Alternatif modeller araştırılacak.

- Hugging Face Text-to-Speech modellerinden birini kullanarak Türkçe çağrı merkezi ses verilerinin oluşturulması.(Büşra Sulukan.)

Not: Tamamlanılan görev dağılımlarını kırmızı ile boyayabilirsiniz.

Week 2

20 Temmuz 2024 – 27 Temmuz 2024

- Projeyle ilgili platformu oluşturmak için web arayüzünün tasarımının taslağının oluşturulması.
(Gözde)
- Alternatif Open Source TPU Tabanlı API desteği sağlayan LLM'lerin araştırılması
(Berke)
- Fine-Tuning yapılacak LLM modellerinin araştırılması.
(Berke-Büşra)
- Fine-Tuning işlemi için Open Source Turkish Datasetlerinin araştırılması.
(Büşra-Gözde)
- Türkçe deyimler-atasözleri data olarak çekilmesi ve etiketlenmesi.
(Gözde-Büşra)
- Alternatif olarak belirlenen Open-Source LLM Modelinin Fine Tuning edilmesi.
(Berke)
- Projenin FastAPI'ye deploy edilmesi.
(Büşra)
- Ses dosyalarının iki farklı cinsiyete ait olması. (Bunu kendimiz ses kaydı ile oluşturacağız.)

Week 3

28 Temmuz 2024 – 03 Ağustos 2024

- Çekilen ve oluşturulan datasetlerin tekrar Fine-Tuning edilmesi ve Türkçe NLP Litatürüne katkı sağlanması. (Microsoft Azure ücretsiz kullanıcılara GPU tabanlı Notebook desteği sağlayamadığı için kota koyduğu için bu işlemi gerçekleştiremedik.
- Web Arayüzünün FastAPI'ye bağlanması.

Bu maddelerden sonra projenin bitirilmesi.

Week 4

03 Ağustos 2024 – 06 Ağustos 2024

- Projenin sunum dosyasının hazırlanması
- Github Reposunun ve readme.md dosyasının güncellenmesi.