

# CASE STUDY

## 1.PROJE AÇIKLAMASI

TranscriptTransformer projesi, konuşma metinlerini eğitim materyaline dönüştüren kapsamlı bir NLP sistemidir. Sistem, öncelikle DOCX formatındaki girdi dosyasını okuyarak işleme başlar. Bu okuma işlemi sırasında python-docx kütüphanesi kullanılır ve metin paragraflar halinde alınır. Ham metin alındıktan sonra, preprocess\_text metodu ile temizleme ve normalizasyon işlemleri gerçekleştirilir. Bu aşamada gereksiz boşluklar, özel karakterler temizlenir ancak noktalama işaretleri ve cümle yapısı korunur.

Sistemin kalbi olan NLP işlemleri için NLTK ve SpaCy kütüphaneleri kullanılır. NLTK, temel metin işleme görevleri için kullanılırken (tokenizasyon, stop words temizleme), SpaCy daha gelişmiş NLP görevleri için kullanılır (named entity recognition, dependency parsing). Sistem başlatılırken gerekli NLTK kaynakları ve SpaCy modeli (en\_core\_web\_sm) yüklenir. Ayrıca, anahtar kavramlar ve ilişkili terimler bir sözlük yapısında tanımlanır.İçerik analizi aşamasında, metin önce ana bölümlere ayrılır. Bu bölümlendirme, önceden tanımlanmış anahtar kelimelere göre yapılır. Her bölüm için önemli cümleler çıkarılır, özetler oluşturulur ve eğitim notları hazırlanır.

Öğrenme hedefleri, metnin içeriğine göre dinamik olarak belirlenir. Sistem, her bölüm için anahtar noktaları tespit eder ve bunları eğitim notlarına dönüştürür.Metin işleme sürecinde, gelişmiş NLP teknikleri kullanılarak önemli kavramlar ve ilişkiler tespit edilir. Örneğin, cümlelerin önem skorları hesaplanır, tekrar eden içerikler temizlenir ve tutarlı bir anlatım sağlanır. Sistem ayrıca, her bölüm için tartışma soruları üretir ve aktivite önerileri sunar. Bu işlemler sırasında, metnin eğitim materyali formatına uygun hale getirilmesi için çeşitli şablonlar ve kalıplar kullanılır.Kalite kontrol mekanizmaları, sistemin önemli bir parçasıdır. word\_count\_verification metodu ile minimum kelime sayısı kontrolü yapılır (en az 3900 kelime hedeflenir). İçerik tutarlılığı kontrol edilir, tekrarlar önlenir ve çıktının eğitim materyali standartlarına uygunluğu denetlenir. Hata durumlarında detaylı geri bildirimler sağlanır ve işlem güvenliği için try-except blokları kullanılır.

Çıktı formatlaması, python-docx kütüphanesi kullanılarak yapılır. Başlıklar, alt başlıklar, paragraflar ve diğer metin öğeleri profesyonel bir şekilde formatlanır. Sistem, çıktıyı DOCX formatında kaydeder ve bu dosya doğrudan eğitim materyali olarak kullanılabilir.

Çıktı, öğrenme hedefleri, ana konular, bölüm özetleri, eğitim notları ve detaylı içerik bölümlerini içerir.Performans optimizasyonu için çeşitli teknikler kullanılır. Generator yapıları ile bellek kullanımı optimize edilir, gereksiz işlemler önlenir ve verimli veri yapıları tercih edilir.

Sistem modüler bir yapıda tasarlanmıştır, bu sayede yeni özellikler eklemek veya mevcut özellikleri değiştirmek kolaydır. Ayrıca, farklı dillerde içerik işleme potansiyeli vardır ve sistem gerektiğinde bu yönde geliştirilebilir.

Sonuç olarak, TranscriptTransformer projesi, modern NLP teknikleri kullanarak konuşma metinlerini yapılandırılmış eğitim materyallerine dönüştüren kapsamlı bir sistemdir. Otomatik içerik yapılandırma, kalite kontrol mekanizmaları ve profesyonel çıktı formatlaması ile eğitim materyali hazırlama sürecini önemli ölçüde kolaylaştırır ve standardize eder.

## **2.PROJE DENEYİMLERİ**

Yapay zeka ile ilgili çalışmalarında daha önce doğal dil işleme(NLP) tekniklerini çok kullanmamıştım sadece küçük bir bankacılık chat botu ve durum duygu analizi üzerine çalışmıştım.Bu projede ise bunlardan farklı olarak daha büyük bir transcript metninden bir eğitim materyali çıkarmam isteniyordu.Bu doğrultuda hazır NLP ve LLM modelleri kullanarak bunu yapmaya çalıştım.En zorlandığım kısımlar ise eğitim aşamasında modele doğru promptları vermek oldu bunun seçiminde hata yaptığımda ana metinden çok bağımsız sonuçlarla karşılaştım.Sonuç olarak hazır model üzerinden transformu gerçekleştirebildim.Benim için güzel bir deneyim oldu bu projeden sonraki hedefim kendimi geliştirmek için kendi verilerim için transfer learning yaparak bir LLM modeli üretmek olacak.