



Literatür Taraması ve Sektör Araştırması

Tıp Alanında Doğal Dil İşleme Destekli Dijital İkiz Tasarımı

Emel KAYACI – 05180000087

Mehmet Anıl TAYSI – 05170000022

16.1.2022

İçindekiler

1. Giriş	1
2. Doğal Dil İşleme Alanında Literatür ve Sektör Araştırması	1
3. Yüz Canlandırma Alanında Literatür ve Sektör Araştırması	3
4. Sonuç	6
5. Kaynakça	7

1. Giriş

Tezimiz için literatür taraması ve sektör araştırması yaparken dikkat ettiğimiz en önemli husus, doğal dil işleme tarafı için ayrı, yüz canlandırma alanında ayrı araştırma yapmaktır. Bunun sebebi, tezimizin bu iki ana konunun birleşimiyle oluşmasıdır. Temelde doğal dil işleme tarafından kullanacağımız/geliştireceğimiz chatbot teknolojisinin yüz canladırmaı destekleyip onun çalışmasını sağlayacağından, literatür taraması hususunda kullanıcıyla etkileşimi de içeren bu ilk konuya ağırlık verdik. Sağlık sektörü için chatbot geliştirmeye yönelik halihazırda çok fazla çalışma olmasına karşın, GAN teknolojisi kullanarak yapılan yüz canladırma uygulamalarını sektör özelinde değerlendirmedik, buna uygun çok spesifik bir çalışmaya rastlamadık.

2. Doğal Dil İşleme Alanında Literatür ve Sektör Araştırması

Chatbot veya diğer bir adıyla chatterbot 1966'dan beri geliştirilmekte olup bilgisayarların kullanıldığı neredeyse her alanda kullanılmaktadır.

Müşteri hizmetleri alanında bir arama ürün ve servis tipine göre 8 ile 45 dakika arasında sürebilmektedir. Müşteri hizmetleri çağrı merkezi işe aldıkları her temsilci için 4.000 dolara kadar ve bu temsilcilerin eğitimi için daha da fazla para harcamaktadırlar. Ayrıca sonrasında %30-45 oranında çalışan devri yaşamaktadırlar. Bu durum sadece Amerika Birleşik Devletleri'nde yıllık satışlarda ortalama 62 milyar dolarlık zarara neden olmuştur. [1]

Covid-19'da sağlık alanında chatbot geliştirilmesinde oldukça önemli bir rol oynamıştır. Örneğin Whatsapp, kullanıcıların COVID-19 hakkındaki sorularını yanıtlayan bir sohbet robotu hizmeti oluşturmak için Dünya Sağlık Örgütü (WHO) ile iş birliği yapmıştır.

Ya da 2020'de Hindistan Hükümeti, Whatsapp üzerinden çalışan ve insanların Coronavirus (COVID-19) pandemisi hakkında bilgilere erişmesine yardımcı olan MyGov Corona Yardım Masası adlı bir sohbet robotunun geliştirilmesinde destek sağlamıştır.

Karma yöntemli bir çalışma, insanların teknolojik karmaşıklığın yeterince anlaşılmaması, empati eksikliği ve siber güvenlikle ilgili endişeler nedeniyle sağlık hizmetleri için sohbet robotlarını kullanmakta hâlâ tereddüt ettiğini gösterdi. [2]

Bizim asıl amaçlarımızdan biri de burada yer alan empati eksikliği problemine çözüm getirmektedir. Projemizde, kullanıcının yapay zeka yerine gerçek bir insanla konuşmasını hissettirecek iki önemli unsur bulunmaktadır: Bunlardan ilki konuşmanın içeriğinin (veri

setinin) empati içerecek şekilde oluşturulmuş olması ikincisi ise GAN ile üretilmiş sanal yapay insan yüzü ile bir animasyon yerine gerçekten insan ile konuşuyormuş hissiyatı yaratmasıdır.

Bu alanda araştırdığımız örneklerden bazıları aşağıda yer almaktadır:

Medibot, IOT teknolojisini chatbot ile birleştirip zengin bir işlem seçeneği sunmaktadır. Bulundurduğu sıcaklık, kalp atış hızı sensörü, kandaki şeker miktarını ve kan basıncını ölçmeyi sağlayan araçlar sayesinde hastalık tahmini gerçekleştirebilmekte ve bunu chatbot aracılığıyla kullanıcıya aktarabilmektedir. Ayrıca doktor reçeteleri kendine özgü el yazısı stili nedeniyle bir eczacı veya başka bir doktor tarafından okunabilir, yani sıradan bir insan reçeteyi okumakta oldukça zorlanır. Medibot el yazılarından reçete okuma özelliğini de bulundurmaktadır. [3]

Likita, Afrika'da sağlık hizmetlerini iyileştirmek için geliştirilmiş bir chabot'tur. Afrika sayısız sağlık sorunuyla karşı karşıyadır ve nüfusunun yalnızca %50'sinden azının, özellikle kentsel alanlarda yaşayanların, modern sağlık tesislerine erişimi bulunmaktadır. Bu da chabot'ların yalnızca günlük yaşamı kolaylaştıracak değil yaşam için kritik rollerinin de olduğunu göstermektedir. Likita yaygın rahatsızlıkların teşhisinde ve uygun tedavilerin önerilmesinde, doktor randevularının ayarlanmasında, hastalara ilaç hatırlatılmasında ve sağlıkla ilgili konulardaki soruların yanıtlanmasında yardımcı olmaktadır.

Örneğin sıtma ve tüberküloz Afrika'da oldukça yaygın hastalıklardır. Likita sayesinde bu hastalıklara teşhis konulabilmekte, teşhis konulduktan sonra da hastalığın ilerleyiş durumuna göre çeşitli öneriler sunulmaktadır. [4]

Yaptığımız araştırmalar sonucunda sağlık alanında oldukça fazla chatbot örneğine rastladık. Konumuzun bu avantajından yararlanarak çalışmalarımızı bu sefer de chatbot örneklerinin karşılaştırılmasına yönlendirdik. 27 farklı chatbot örneğinin metot, artı ve eksilerinin tablo biçiminde karşılaştırıldığı makaleyi inceledik.

NLP alanında daha önce yaptığımız çalışmalarda gözlemlediğimiz en önemli unsur yapay zekanın diğer alanlarına göre bu alanda veri setinin gösterim zorluğunun olduğudur. Resimler, piksel değerleri yardımıyla RGB formatında (renkli ise) kolaylıkla gösterilebilirken kelimelerin gösterilmesi için onlarca yol bulunmaktadır. Örneğin kelimeler bazen liste veri yapısıyla bazen de vektörler yardımıyla gösterilebilirler.

Makalede metot, algoritma veya model, veri seti bilgileri sayesinde birçok örneğin hangi teknolojileri kullandığına dair fikir sahibi olmuş olduk. Ayrıca bu teknolojilerin getirdiği avantaj ve dezavantajları da gözlemleyerek implementasyona geçiş yaptığımızda dikkate alacağımız noktaları belirlemiş olduk. [5]

Spor ve sağlık birbirleriyle oldukça iç içe olan kavramlardır. Kişiler her iki alanda ilerleyişlerini görebilmek için en az iki uygulama kullanmak zorundadırlar fakat bazen bu bile ortak bir ilerleyiş bilgsinin oluşması için yeterli olmamaktadır. Bu makalede bu dezavantaja odaklanılıp çözüm getirilmiştir. [6]

Chatbot alanında empati, önemli bir çalışma konusu olmuştur. Özellikle bu makalede yer alan chatbot kanser teşhisi yaptığından empati daha da kritik bir öneme sahiptir. Bu makalede teşhis yapılırken yapay zekanın nasıl insan gibi davranabileceği incelenmektedir. Bunun için duygu analizi (sentiment analysis) kullanılmıştır. [7]

3. Yüz Canlandırma Alanında Literatür ve Sektör Araştırması

Önceki bölümde de bahsettiğimiz üzere, bitirme projemizdeki en temel unsur, tıp alanında hastalık tespiti ve bu hastalıklar için teşhis önerisi yapabilen, kullanıcı ile iletişimi sırasında normal iletişimden olabildiğince farksız bir yapı oluşturmaktır. Bu yapının elde edilmesi hususunda doğal dil işleme kısımlarının yanı sıra, görsellik önemli bir faktördür. Amacımız kullanıcı veya hastanın, sistemimizi kullanırken robotla konuşuyor olma hissiyatını minimize etmektir. Dolayısıyla, alanında önemli bir figürün bir fotoğrafından (veya birden fazla) yola çıkarak doğal dil işleme tarafındaki diyalogları mimikleriyle ve ağız hareketleriyle görselleştirmeyi hedefliyoruz. Bu durumu gerçekleştirdiğimiz takdirde, figürümüzün bir nevi dijital ikizini ortaya çıkarmış oluyoruz.

GAN tabanlı mimik taklit etme uygulamaları son yıllarda ‘deepfake’ teknolojisiyle birlikte çok büyük ilerlemeler katetmiştir. Özellikle siyasal olarak önemli figürlerin mimikleri farklı konuşmaları, yani kendisinin daha önce hiç yapmadığı konuşmaları yapıyormuşçasına taklit edilebilmiştir. Bu gelişmeler beraberinde çeşitli etik problemlerini de beraberinde getirmiştir. Biz bu başlık altında, ‘Video generative adversarial network’ alanında yapılmış çalışmalar ve yayınlanmış makalelere yoğunlaşacağız.

Bir görüntü üzerinden video sentezi yapmak çeşitli görüntü işleme uygulamalarının kullanılmasıyla birlikte GAN teknolojisinin birleştirilmesi sonucu oluşturulmuş oluyor. Video sentezi, statik görüntüler yerine video içeriği oluşturmaya odaklanır. Görüntü sentezi

ile karşılaştırıldığında, video sentezinin çıktı videolarının zamansal tutarlılığını sağlaması gerekmektedir. [8]

Yüz değiştirme, bir videodaki veya fotoğraftaki özneyi diğer video videodaki ana karakter yerinde oynatma temeline dayanırken, yüz canlandırma (face reenactment), ifadelerin ve baş pozlarının hedef öznenin kaynak görüntüye aktarılmasıyla ilgilidir.



Görüntü 1. *Yüz canlandırma prensibi [8]*

Görüntü 1’de görüldüğü üzere kaynak olarak alınmış görüntü üzerinde ‘Target Video’ aracılığı ile bir yüz canlandırma işlemi yapılmıştır. Bu uygulama için kullanılabilecek model, sadece tek bir görüntüye bağlı olarak tasarlanabilir, ya da herhangi bir görüntü ile çalışabilecek şekilde çalışabilecek şekilde dizayn edilebilir. [8] Yüz canlandırma işlemi farklı sektörlerin çeşitli uygulama alanlarında kullanılmaktadır.

GAN teknolojisi çıktıktan sonra, bu teknoloji kullanılarak çok farklı kullanım çeşitleri geliştirilmiştir. Tezimiz ile bağlantısı dolayısıyla yüz canlandırma amacı güden UniFaceGAN adlı bir tür GAN olan bu derin öğrenme algoritması/sistemi, yukarıdaki paragrafta bahsedilen yöntem için özel geliştirilmiş bir teknolojidir.



Görüntü 2. *UniFaceGAN ve çeşitli algoritmalarının kıyaslanması [9]*

Yukarıdaki görüntüye, UniFaceGAN’ın orijinal makalesinde yer verilmektedir. Yüz canlandırma uygulaması için çeşitli karşılaştırmalar içermektedir. Buna göre, (a) kaynak fotoğrafımız, (b) yüz canlandırmanın uygulanacağı örnek videodaki ilk kare, (c) ise

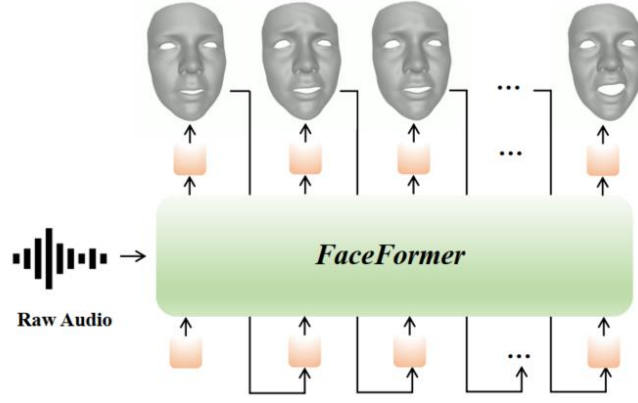
UniFaceGAN’ın ilk framedeki başarısını gösteren görüntüdür. Kalan görüntüler ise sırasıyla, (d) FSGAN. (e) X2face. (f) FTH (Few-shot T. Heads). (g) FOMM (h) Fast Bi-layer. Şeklinde ilerlemektedir. [9]

Buraya kadarki araştırmalarımız mevcut kaynak görüntümüzün uygulanabileceği hazır bir videonun varlığıyla elde edilen yüz canlandırma uygulamalarına dayanmaktadır. Fakat, bizim uygulamamızda oluşturulacak metin doğal dil işleme yoluyla elde edilecek olup, hazır bir video ile desteklenmeyecektir. Dolayısıyla bu aşamadan sonra metin üzerinden yüz canlandırma tekniklerinin incelenmesi gerekmektedir.

Metin bazlı yüz canlandırma tekniği üzerine Temmuz 2020 yılında “Text and Audio-Based Real-Time Face Reenactment” makalesi ile Amerika Birleşik Devletleri patent dairesine başvuru yapılmıştır. Temel mantığı özetleyecek olursak, girdi metninin karakteristik ve dilsel (linguistic) özellikleri çıkarılır, bir derin öğrenme ağı ile model geliştirilir, metinden oluşabilecek akustik özellikler ile birleştirilerek metine bağlı ağız hareketlerinin oluşturulması sağlanır. [10] Ses faktörünün kullanımı için farklı bir boyutudur, metinde uygulanacak vurgular veya mimik hareketlerinin subjektif olması da dikkat edilmesi gereken hususlardır.

Bu noktaya kadar verilen benzer çalışmalar zaten önceden var olan bir videoyu referans olarak çalışmaktadır. Üst paragrafta bahsedilen metin tabanlı yüz canlandırma işlemi uygulamayı başka bir boyuta taşımıştır. Projemizin en kritik noktalarından birisi, yüz canlandırma metodunu bir otomasyon geliştirerek başarmaktır. Bu otomasyon metin tabanlı olabileceği gibi, oluşturduğumuz metinleri otomatik seslendiren bir sistemi kullanarak cümlelerin vurguları ve akustiği gibi verilerden de faydalanmak suretiyle yüz canlandırma işlemini daha başarılı elde edebileceğimizi düşünüyoruz. Bu şu anlama gelmektedir, şu anki literatürde bu alanda bulunan çoğu çalışma face transfer (yüz transferi) kullanılarak yapılmaktadır, ve bu durum bizim tezimiz için çoğu çalışmayı kullanışsız yapmaktadır. Buna karşın, durumun böyle olması diğer çalışmalar ile benzerliği minimal seviyelere çekmektedir.

Yüz canlandırma animasyonu oluşturmak adına yapılan en yeni çalışmalardan biri olan ve 28 Aralık 2021 tarihinde yayınlanan bir makalede [11] Transformers adlı makine öğrenmesi modelini baz alarak, girdi olarak verilen bir ses dosyasından bu ses dosyasına göre bir yüz animasyonu oluşturmaya yönelik çalışmalar yapıldığı gözlenmiştir.



Görüntü 3. *İlgili makaledeki [11] yapının Kavram Diyagramı*

Ham ses girişi ve nötr bir 3-boyutlu yüz ağı göz önüne alındığında, uçtan uca Transformer tabanlı mimari ile desteklenen ve FaceFormer olarak adlandırılan bu yapı, doğru dudak hareketleriyle bir dizi gerçekçi 3-boyutlu yüz hareketlerini otoregresif olarak sentezleyebilir. [11]

Önceki kısımlarda da bahsettiğimiz gibi, bir metinden baz alınarak yapılan yüz canlandırma işlemi, akustik ve vurgulama gibi önemli özniteliklerden yoksun olacağından doğru yüz ve dudak hareketlerinden uzak davranışlar sergileyebilir. Bu bağlamda, oluşturulan metinleri ses haline getirip, bu sesi yüz canlandırma işlemi için girdi olarak kullanabiliriz. Araştırmacılar geliştirdikleri FaceFormer[11] projesinde, yüzü nötr bir animasyon şeklinde üretmişlerdir. Fakat bizim çalışmamızda yüz için belli bir referans değerimiz olacaktır. Fakat yüz hareketlerini oluşturmak için kullanılan bu teknoloji, bizim çalışmamızla benzerlik gösterebilecektir.

4. Sonuç

Yaptığımız bu araştırmalar sonucunda, hem daha önceki çalışmalar ile kendi tezimiz arasındaki benzerlikleri incelemiş olduk, hem de proje yapısının kurulmasındaki kilit noktalar hakkında fikir sahibi olmuş olduk. Tezimizin içeriği bilgisayar biliminin iki büyük alt başlığını kapsadığından literatür taraması için bu iki bölümü (doğal dil işleme ve GAN tabanlı yüz canlandırma/yüz replikasyonu) ayrı başlıklar altında inceledik. Chatbot tabanlı animasyon oluşturma çalışmaları literatürde yer bulmakta, fakat yüz canlandırma işlemini metin bazlı bir chatbottan temel alıp çalışan kapsamlı bir araştırmaya rastlamadık.

5. Kaynakça

- [1] A. Freed. (2021). *Conversational AI (s. 4)*. Manning Publications Co.
- [2] Chatbot. (2022, January 15). Wikipedia: <https://en.wikipedia.org/wiki/Chatbot>
- [3] K. Sivaraj, K. Jeyabalasuntharam, H. Ganeshan, K. Nagendran, J. Alosious, J. Tharmaseelan, “Medibot: End to end voice based AI medical chatbot with a smart watch,” January 2021
- [4] Oladapo O. Oyeboode, R. Orji, “Likita: A Medical Chatbot To Improve HealthCare Delivery In Africa,” January 2019
- [5] A. Reyner, W. Tjiptomongsoguno, A. Chen, H. Sanyoto, E. Irwansyah(B), and B. Kanigoro, “Medical Chatbot Techniques: A Review,” November 2020
- [6] S. Rai, A. Raut, A. Savaliya, Dr. R. Shankarmani, “Darwin: Convolutional Neural Network based Intelligent Health Assistant,” 2018 IEEE
- [7] Belfin R V, Shobana A J, Megha Manilal, Ashly Ann Mathew, Blessy Babu, “A Graph Based Chatbot for Cancer Patients,” 2019 ICACCS
- [8] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, Arun Mallya, “Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications”, 2020 arxiv.org
- [9] Meng Cao, Haozhi Huang, Hao Wang, Xuan Wang, Li Shen, Sheng Wang, Linchao Bao, Zhifeng Li, “UniFaceGAN: A Unified Framework for Temporally Consistent Facial Video Editing”, Ağustos 2021 arxiv.org
- [10] Pavel Savchenkov, Maxim Lukin, Aleksandr Mashrabov, “Text and Audio-Based Real-Time Face Reenactment”, United States Patent Application Publication, 23 Temmuz 2020
- [11] Yingruo Fan, Zhaojiang Lin, Jun Saito, Wenping Wang, Taku Komura, “FaceFormer: Speech-Driven 3D Facial Animation with Transformers”, Aralık 2021 arxiv.org