Tıp Alanında Doğal Dil İşleme Destekli Dijital İkiz Tasarımı

Emel KAYACI, Mehmet Anıl TAYSİ
Bilgisayar Mühendisliği Bölümü, Ege Üniversitesi 35100 Bornova İzmir
kayaciemel18@gmail.com, anil_taysi@hotmail.com

Prof. Dr. Aybars UĞUR, aybars.ugur@ege.edu.tr

Teslim Tarihi: 25.02.2022

Özet. 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 gerçekletirdiğimiz takdirde, figürümüzün bir nevi dijital ikizini ortaya çıkarmış oluyoruz.

Anahtar Kelimeler: Doğal Dil İşleme, Yazılım Mühendisliği, Yapay Zeka, Görüntü İşleme, Generative Adversarial Networks, Dijital İkiz, Sağlık Bilimleri, Seq2Seq

Abstract. The most basic element in our project is to create a structure that can detect diseases in the field of medicine and make diagnosis suggestions for these diseases, while communicating with the user as much as possible from normal communication. Besides the natural language processing parts, visuality is an important factor in achieving this structure. Our aim is to minimize the feeling of the user or patient talking to the robot while using our system. Therefore, based on a photograph (or more than one) of an important figure in the field, we aim to visualize the dialogues on the natural language processing side with their facial expressions and mouth movements. If we realize this situation, we reveal a kind of digital twin of our figure.

Keywords: Natural Language Processing, Software Engineering, Artificial Intelligence, Image Processing, Generative Adversarial Networks, Digital Twin, Health Sciences, Seq2Seq

İÇİNDEKİLER

1. Giriş

Teknolojinin ilerlemesiyle birlikte yapay zeka algoritmalarının da gücü katlandı ve katlanmaya devam ediyor. Bu bağlamda, artık hiç varolmamış insan yüzleri, görsel üzerine çeşitli yöntemlerle mimik ve konuşma taklidi gibi uygulama alanları ortaya çıktı.

Ayrıca, artık cep telefonlarımızda, dijital bankacılık gibi sektörlerde de 'dijital asistan' adı altında doğal dil işleme desteği olan yapay zekalar işlerimizi kolaylaştırabiliyor. Projemizin amacı, yukarıda belirtilen örneklerin harmanlanmasıdır. Bir yapay zeka (sohbet botu) ile yazışma hissiyatı bazen her ne kadar insanları hayrete düşürüyor olsa da, bir makine ile konuşulduğu için bıraktığı etki sınırlı kalabiliyor.

Fakat buna gerçek yüz içeren bir görsel aracılığı ile yapay zekanın ürettiği söylemleri mimikleri ile, dudak hareketleri ile, ses ile karşıya arktardığımızda bunun çok daha etkileyici olması kaçınılmazdır. Bizim bu projedeki temel hedefimiz, uygulamanın kullanıcılarına hem görsel kanalla hem de işitsel kanalla ulaşıp, yapay zekanın ikna edebilirlik düzeyini yükseltmektir. Yani kısacası, bir doktorun dijital ikizi şeklinde davranan bir sistem geliştirmeyi hedefliyoruz.

Özellikle insan ilişkileri konusunda aktif ve/veya kamuya hitap eden bireyler için bizim 'dijital ikiz' çalışmamız gerçekten çok etkili olabilir. Özellikle Tıp alanında doktorların aynı sorular ile farklı kelimlerle veya farklı şekillerde karşılaştıkları aşikardır.

Böyle bir sistem, temel seviyedeki bir çok hasta sıkıntılarını anlayabileceğinden ötürü, bir danışman olarak kullanılabilir. Bu danışmanı etkili yapan şey ise daha önceki bölümde de belirttiğimiz gibi hastanın ekranda sadece yazıları değil, doktorun bir replikasını da görecek olmasıdır.

2. Literatür Çalışması

2.1 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 canladırmayı 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.

2.2 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]

2.3 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 gerçekletirdiğimiz 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. Şekil 1'de de görüldüğü üzere görüntünün 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 özneden kaynak görüntüye aktarılmasıyla ilgilidir.

Şekil 1. *Yüz canlandırma prensibi*

Şekil 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 canladırma amacı güden ve Şekil 2'de görünen UniFaceGAN adlı bir tür GAN olan bu derin öğrenme algoritması/sistemi, yukarıdaki paragrafta bahsedilen yöntem için özel geliştirilmiş bir teknolojidir.

Şekil 2. UniFaceGAN ve çeşitli algoritmalarının kıyaslanması

Yukarıdaki görüntüye, UniFaceGAN'ın orijinal makalesinde yer verilmektedir. Yüz canladırma 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 kadar araştırmalarımız mevcut kaynak görüntümüzün uygulanabileceği hazır bir videonun varlığıyla elde edilen yüz canladırma 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.

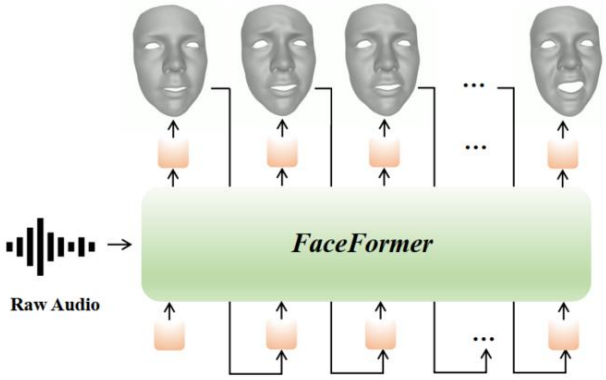
Metin bazlı yüz canlandırma tekniği üzerine Temmuz 2020 yılında "Text and Audio-Based Real-Time Face Reenactment" makalesi ile Amerika Birleşik Devletleri patent dairesine başvuru yapılmıştır. Temel mantığı özetleyecek olursak, girdi metninin karakteristik ve dilsel (linguistic) özellikleri çıkarılır, bir derin öğrenme ağı ile model geliştirilir, metinden oluşabilecek akustik özellikler ile birleştirilerek metine bağlı ağız hareketlerinin oluşturulması sağlanır. [10]

Ses faktörünün kullanımı işin 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 alarak 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 Şekil 3'te görüldüğü üzere bir yüz animasyonu oluşturmaya yönelik çalışmalar yapıldığı gözlenmiştir.



Şekil 3. İlgili makaledeki 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.

2.4 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.

3. Yöntem ve Teknolojiler

Çalışmamızın yapısı gereği iki ana modül içerdiğinden, kullandığımız veya kullanacağımız yöntem ve teknolojileri iki ana başlık altında inceleyeceğiz. Bunlardan ilki, interaktif bir şekilde çalışabilecek arayüze sahip, doğal dil işleme destekli sohbet robotu, ikincisi ise bu sohbet robotunun ürettiği cümlelerden beslenip bunu mimik, tepki gibi unsurlarla belirli bir kişinin yüzünde canlanacak şekilde oluşturacağımız yüz canlandırma (face reenactment) yapısı olacaktır.

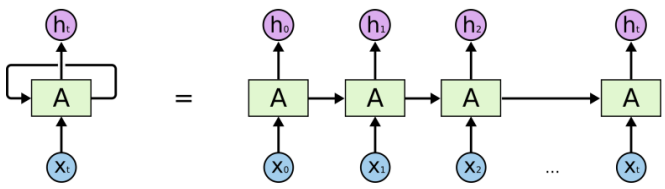
3.1 Sohbet Robotu (Chatbot)

Sohbet robotları, günümüzde çeşitli sektörlerde ve çalışmalarda kullanılan, öneri sistemleri, müşteri destek sistemleri, eğitim sistemleri gibi sistemlerde büyük faydalar sağlamış bir teknolojidir.

Chatbot teknolojisini kurabilmek için derin öğrenme ve NLP (Natural Language Processing) alanlarının birleşiminden oluşan ve DNLP (Deep Natural Language Processing) olarak adlandırılan alan içerisinde yer alan RNN (Recurrent Neural Networks) ve LSTM (Long Short Term Memory) gibi yapıların araştırılması gerekmektedir.

Geleneksel yapay sinir ağlarının önemli bir dezavantajı kısa süreli bir hafızaya sahip olmamalıdır. Örneğin DNLP projelerinden biri olan video örneklerinden video konusunun çıkarılmasında geleneksel yapay sinir ağları başarısız olmaktadır çünkü bu tarz bir projede ard arda gelen resimlerin arasındaki bağlantının da keşfedilmesi gerekmektedir.

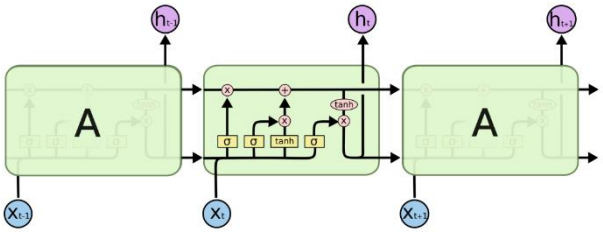
RNN yapıları bu soruna feed-back yapıları sayesinde çözüm getirmektedirler. Buradaki feed-back yapısını birbirine bağlanmış yapılar halinde düşünebiliriz. Şekil 4'te RNN yapısının açılmış hali verilmiştir.



Şekil 4. RNN mimarisinin yapısı

“Hava kapalı ve *yağmur* yağacak, “ şeklinde kısa bir cümlede RNN modeli yağmur kelimesini tahmin edebilirken “Ben Türkiye’de doğdum… Ben aksansız *Türkçe* konuşabiliyorum.” gibi uzun bir cümlede Türkçe kelimesini tahmin edememektedir. Bu sorunun da üstesinden gelmek için RNN mimarisinin gelişmiş versiyonu olan LSTM önerilmiştir. [12]

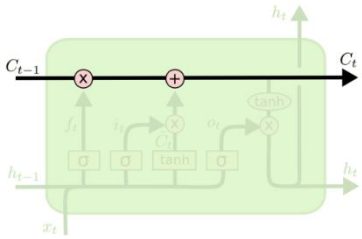
Hochreiter & Schmidhuber tarafından 1997 yılında tanıtılan LSTM mimarisi bu soruna bir ek mekanizma getirip çözmekten ziyade bu sorun hiç yaratılmadan çalışmaktadır. Şekil 5’te LSTM mimarisinin sembolik şekli yer almaktadır. [12]



Şekil 5. LSTM mimarisi

Mimaride belirtilen her bir ok bilgileri vektör olarak taşımaktadır. Sarı ile gösterilmiş kısımlar yapay zeka katmanını, pembe yuvarlak noktalar vektörler ile gerçekleştirilebilecek işlemleri belirtmektedirler.

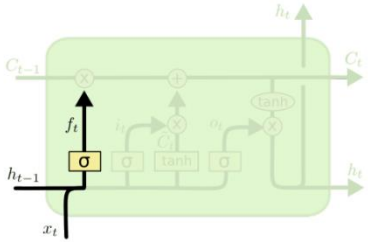
LSTM mimarisinin başarısı cell state olarak adlandırılan ve Şekil 6’da belirtilmiş yapıdan kaynaklanmaktadır. [12]



Şekil 6. Cell state yapısının gösterimi

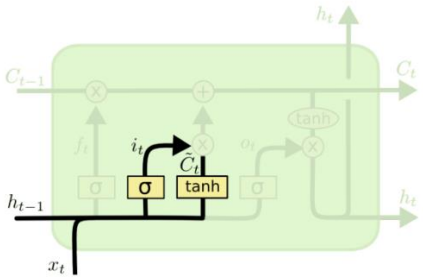
Şekilden de görüldüğü üzere bilgiler bu yapı üzerinde çok az bir değişimle iletilmektedirler. Yapıda gösterilen pembe kısımlar somut olarak su valfı olarak düşünülebilmektedirler. Her bir kısmın cell state üzerinde farklı etkisi bulunmaktadır.

Şekil 7’de bu yapılardan ilki olan forget gate görünmektedir. Sigmoid fonksiyonu 0 ile 1 arasında değerler almaktadır. Fonksiyon 0 değerini aldığı taktirde cell state yapısına input olarak x_t ve h_{t-1}‘den gelen bilgi yansıtılmayacaktır. 1 değeri geldiğinde ise kesinlikle tutulması gereken önemli bilgi tutulacaktır. [12]



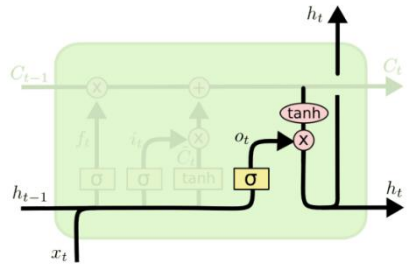
Şekil 7. Forget gate yapısının gösterimi

Örneğin “How is your *sister*?” sorusuna cevap olarak “*She* is fine.” Verilebilmesi için cinsiyet unsuru içeren sister kelimesinin hatırlanması gerekmektedir. Fakat bu öznenin her zaman tutulması bir dezavantajdır çünkü farklı sorular geldikçe her zaman *she* öznesinin kullanılması doğru bir yaklaşım olmayacaktır. Bu nedenle yeni özne gelince bilginin önce forget gate yardımıyla unutulup daha sonra güncellenmesi gerekmektedir. Bu güncelleme işlemi Şekil 8’de gösterilmiş update gate yardımıyla gerçekleştirilir. [12]



Şekil 8. Update gate yapısının gösterimi

Son olarak çıktı olarak tutulan bilgilere göre üretilecek bilgi hakkında çıkarsama yapılması gerekmektedir. Şekil 9’da output gate yapısı hem kendisinden önce gelen LSTM bloklarının, hem yeni girdi değerinin hem de hafızada tutulan önemli bilgiyi harmanlayarak çıktının oluşturulmasını sağlamaktadır. Örnekte “*sister”* kelimesine bağlı olarak cinsiyet kavramı tutulduğundan “*she”* çıktı olarak üretilebilmiştir.



Şekil 9. Output gate yapısının gösterimi

Bu mimarilerin yanı sıra BERT gibi daha yeni ve daha hızlı çalışan mimariler de bulunmaktadır. Bu mimariler yalnızca chatbot değil doğal dil çevirisi, duygu analizi gibi çeşitli NLP problemlerinde kullanılmaktadır.

3.2 Yüz Canlandırma

Yüz canlandırma (face reenactment) teknikleri son dönemlerde derin öğrenme alanında üzerine düşülen ve gelişmekte olan yöntemlerdendir. Kullanım yöntemleri olarak, belirli bir girdi görüntüsünün başka bir videodan baz alınarak o videodaki konuşmacı unsurun taklidini yapabilmesi, konuşmayı girdi olarak verilen görüntüdeki kişinin/unsurun yapıyormuş gibi görünmesi sağlama amacı güdülür.

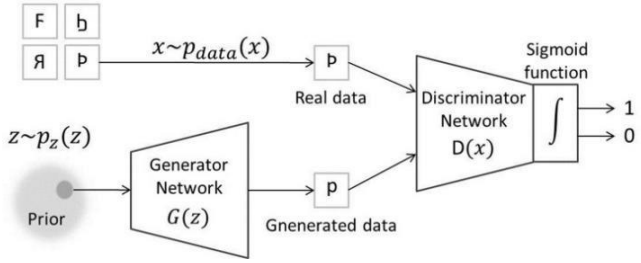
Yüz canlandırma yöntemleri temelde GAN yani Üretici Çekişmeli yapay sinir ağları (Generative Adversarial Networks) araçlarını baz alır, fakat konvülüsyonlu yapay sinir ağları kullanılarak yapılan çalışmalar da vardır. Bizim yöntemimiz, GAN teknolojilerini temel almaktadır.

GAN, temelde iki ana bileşenden oluşan, ve bu bileşenlerin birbirleriyle pozitif veya negatif geribildirimler aracılığıyla çekişmesini sağlayacak bir yapıyla tasarlanmış, üretken yapay sinir ağları kapsamına giren bir tür derin yapay sinir ağları alanıdır. [13]

İlk bileşen Generator (Üretici), GAN'ların üretkenliği sağlayan bileşeni olup, görevi, uygulanılan veriseti çerçevesinde amaca uygun verileri üretmeyi sağlamaktır.

İkinci bileşen ise Discriminator (Ayrıştırıcı) adında, ilk bileşenin oluşturduğu verileri çeşitli metrik ve temel alınan değerlere göre ayrıştırma yapmakla görevlidir. Genellikle ayrıştırıcı bileşen üretici bileşenin ürettiği verileri gerçek hayat verisinden farkını inceler ve bu incelemelere göre üretici bileşene bir dönüt gönderir. Üretici bileşen ise bu geridönütü referans alarak kendini güncellemek ile yükümlüdür. Üretici-çekişmeli ismi de buradan gelmektedir.

Bu iki bileşen birbirleriyle çekişerek kendilerini otomatik bir şekilde eniyilemeye çalışırlar. Şekil 10'te GAN mekanizmasının çalışma mantığı içeren diagrama yer verilmiştir.



Şekil 10. GAN çalışma mantığı diagramı

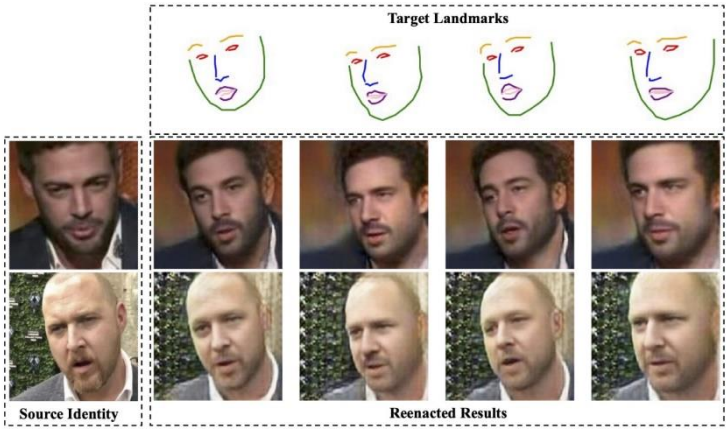
Buradan da görüldüğü üzere, Ayrıştırıcı (Discriminator) bileşen hem gerçek dünya verilerini hem de Üretici bileşenin verilerini girdi olarak alıp, bunları birbiribden ayrıştırmaya çalışmaktadır, buradaki performansına göre Üretici bileşene bir dönüt yollayıp ve Üretici bileşenin kendi parametrelerini değiştirmesini sağlamaktadır ki gerçek dünya verilerine uygun veriler üretilebilsin. Bu durum iki bileşenin birbirlerinin üstüne güç kurma olgusuyla, bir çekişme yaratmaktadır.

Bu kısma kadar, yüz canlandırma işleminde kullanacağımız derin öğrenme mimarisi hakkında bilgi verilmiştir, projemizde kullanacağımız yüz canlandırma yöntemi, hazır bir videodan süre gelmeyeceğinden ve sohbet robotunun oluşturduğu metinleri gerçek zamanlı olarak baz alacağından ötürü, literatürdeki çoğu örnekten farklı bir mekanizmaya ihtiyaç duymaktadır. Fakat sonuç itibariyle, temel alınacak GAN mekanizmaları benzeştiğinden, yöntemlerimiz arasında birkaç GAN örneğini vermiş olacağız.

3.2.1 FaR-GAN

Literatür taraması bölümünde de bahsettiğimiz örneklere benzer yapıda çalışan, tek bir kaynak görüntü verilerek istenilen video formatındaki konuşmaları ve mimik hareketlerini, ya da daha genel haliyle yüz hareketlerinin taklit edilmesini sağlayan bir çalışmadır.

Temel olarak, Face Landmark (yüz sembolleri)'ni baz alıp, videodaki sembolleri ile kaynak görüntünde gerçekletirme ve bu bağlamda yüz canlandırma mantığıyla çalışmaktadır.



Şekil 11. FaR-GAN ve hedef yüz sembollerinin gösterimi

Çalışmanın yayınladığı makalede girdi olarak verilen iki görüntünün de aynı örnek videodaki benzeri yüz hareketleri gösterilmiştir. Görüldüğü üzere hedef yüz sembolleri (target landmarks) sadece girdi videosuna bağlıdır ve iki ayrı girdi görüntüsü için de aynı yüz sembolünün yakalanması amaçlanmıştır. Bunu yapabilmesini sağlayan şey ise önceden eğitime tabii tutulmuş GAN modelinin bu semboller doğrultusunda girdi görüntüsünü baz alıp yeniden üretmesidir. [13]

3.2.2 FaceGAN

FaceGAN, 3.2.1'deki çalışmaya ek olarak background manipulation (arkaplan manipülasyonu) konusunda da çözümler içermektedir. Yüz canlandırma tekniği uygulanırken, kaynka görüntünün yüz canlandırmanın uygulandığı girdi videosunu takip ederken yapabileceği kafa hareketleri, görüntüdeki arka planda oynamalara ve sapmalara yol açabilir. Bu sapmaları giderebilmek için, GAN modülünün ürettiği görüntülerde arkaplan üzerine de manipülasyonlar yapmak gerekmektedir. [14] FaceGAN bu konuyla ilgili hesaplamalar içermektedir.

Şu ana kadar yüz canlandırma konusunda yaptığımız gerek literatür taraması, gerekse yöntemlerin tanıtılması bölümünde literatürdeki çeşitli çalışmaların incelenip, kendi çalışmamızda baz alacağımız yöntem ve modeller için netleştirmeler sağlanmıştır. Bu modül, sohbet robotundan hareketle çalışacağından ötürü, bağımsız hareket edememektedir, fakat çalışmamızın en önemli amaçlarından biri olan kullanım esnasında gerçekçiliğin sağlanması hususunda kritik bir yere sahiptir.

3.3 Ekstralar

Çalışmamızı iki temel modüle böldüğümüzden ve bu modüller üzerine araştırmalar yapıp gerçekleştirim yöntemlerimizin detaylarını belirlediğimizden bundan önceki bölümlerde bahsettik. Bu bölümde ise bu iki modülü amacına uygun çalıştırdığımızda, çalışmamızı nasıl servis edeceğimiz konusunda düşündüğümüz aday servis yöntem ve teknolojileri belirteceğiz.

Çalışmamız servis ediliş itibariyle hem bilgisayar yazılımı olarak, hem websitesi olarak hem de mobil ortamda kullanıma uygun olabilir. Fakat gerek derin yapay sinir ağlarının çalışma zamanları hususunda,

hem de boyut hususunda gerçekletirilebilir ve sürdürülebilir bir yapı hedefliyoruz. Dolayısıyla, hem yüz canlandırma hem de sohbet robotu gerçekleştirimleri python üzerinde gerçekleştirileceğinden, öncelikle yapay sinir ağlarımızı optimize etmek ve işlevsel bir yazılım elde etmek amacıyla masaüstü ortamda, python kütüphanelerinin sağladığı ölçüde bir kullanıcı arayüzü tasarımıyla testler yapma planımız bulunmaktadır. Daha sonrası için, bir web sitesi üzerinde çalışmalarımızı yürütebilmek adına, Google, Amazon gibi şirketlerin bulut bilişim ortamlarından da faydalanarak çalışmalarımızın sonuna gelmeyi hedefliyoruz.

4. Tamamlanan Çalışmalar

Çalışmamızda şu aşamaya kadar, detaylı sektör ve literatür araştırmaları ve çalışmanın gerçekleştirimi için gerekli teknoloji ve yöntemlerin belirlenmesi için çeşitli analizler yapmış olduk. Bu bilgilerin ışığında, çalışmamızın analiz ve tasarımı konusunda ve yüz canlandıma konusunda planlar yapmış bulunuyoruz, ayrıca sohbet robotumuzun tasarımı ve optimizasyonu, yüz canlandırma işlemlerinde kullanacağımız tasarımlardan daha detaylı ve zaman açısından maaliyetli olacağını düşünüyoruz.

4.1 Analiz

Çalışmamızın sistematiğini oluşturan temel unsurlara önceki bölümlerde de değinmiştik. Gerçekleştirim esnasında kullanacağımız programlama dilleri ve kütüphaneleri, modülleri, hem sistemin arka planı için hem de kullanıcının etkileşim içinde bulunacağı kısım için her birinde çeşitli alternatif adaylar olacak şekilde belirlemiş olduk.

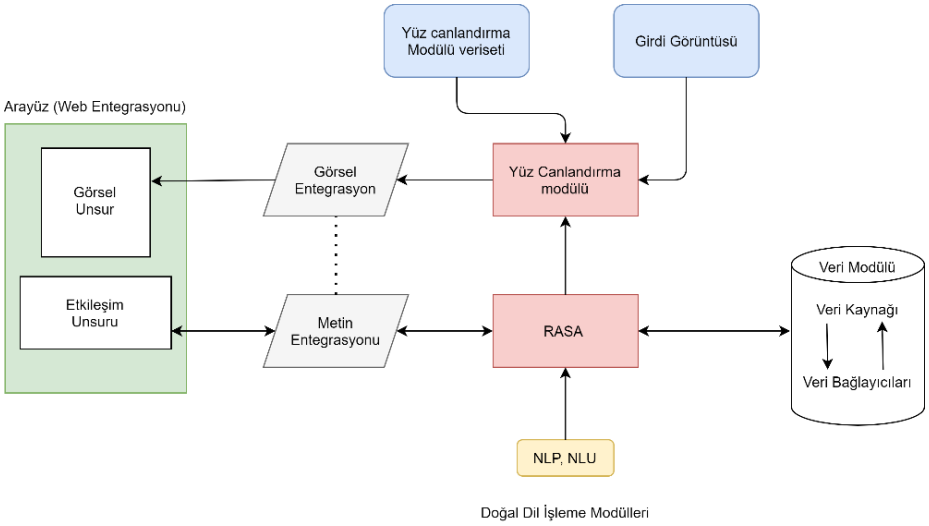
Sohbet robotumuz adına, RASA adında açık kaynak kodlu etkileşimli yapay zeka gelişme ortamını kullanmayı düşünüyoruz, ayrıca bu ortamın sağladığı esneklik olarak, NLP (Natural Language Processing) ve NLU (Natural Language Understanding) birimlerinin modellerini python ortamında kendimiz tasarlayıp, tasarımımızın tamamen Türkçe olmasını sağlayabileceğiz. Sohbet robotu gerçekleştiriminde son aşama olan NLG (natural language generation), geliştiğimizin sistemin kullanıcıya mesaj olarak dönüt vereceği metinleri belirleyecek son kısımdır. Bu kısımda, kullanacağımız verilere ek, bağlamlar (intentions) ve varlıklar (entities) kullanılarak anlatımın güçlendirilmesini mevcut kılabileceğiz.

Yüz canlandırma bileşeni adına, sohbet robotumuzla birebir uyumlu çalışma yapılabilmesi adına literatür çalışması bölümünde yer verdiğimiz “Text and Audio-Based Real-Time Face Reenactment” çalışmasının yanı sıra nu alanda yapılan farklı çalışmaları da inceleyerek yine python ortamında bir model oluşturacağız. Sadece metin üzerinden gerçek zamanlı yüz canlandırma işlemi, yer yer zorluklara yol açabileceğinden, Google tarafından python için özel oluşturulmuş metinlerin seslendirilmesini sağlayan kütüphaneyi türkçe olarak kullanarak, seslendirmedeki unsurları yüz canlandırma tekniklerimize yardımcı olabilmesi adına kullanmaya çalışacağız.

Çalışmamızın kullanıcıyla temasını sağlayacak aday yöntemlere ve teknolojilere 3.3.2 bölümünde değindik. Bu aşama projenin son aşaması olduğundan ve yaptığımız planlar doğrultusunda sonradan nihai karar verileceğinden analize tabii tutulmamıştır.

4.2 Tasarım

Bu bölümde analiz kısmında yer verilen planların bir şema üzerinde gösterilerek kısa özet halinde açıklanmasına yer verilmiştir.



Şekil 12. Projenin Çalışma Diagramı

Diagramda görüldüğü üzere, iki ana bileşenimiz, sohbet robutunun beyni olarak nitelendirebileceğimiz RASA modülü ve yüz canlandırma modülüdür. İki modülün de kendi veri kaynağı ve entegrasyon modülleri bulunmaktadır. Ayrıca, RASA üzerinden tasarlayacağımız sohbet robotuna python'da yazacağımız ek NLP ve NLU modüllerini ekleyip Türkçe özelinde hem yapay sinir ağının anlama kapasitesinin güçlendirilmesi, hem de metin oluşturma becerisinin eniyilenmesi açısından ilerleme katedeceğiz.

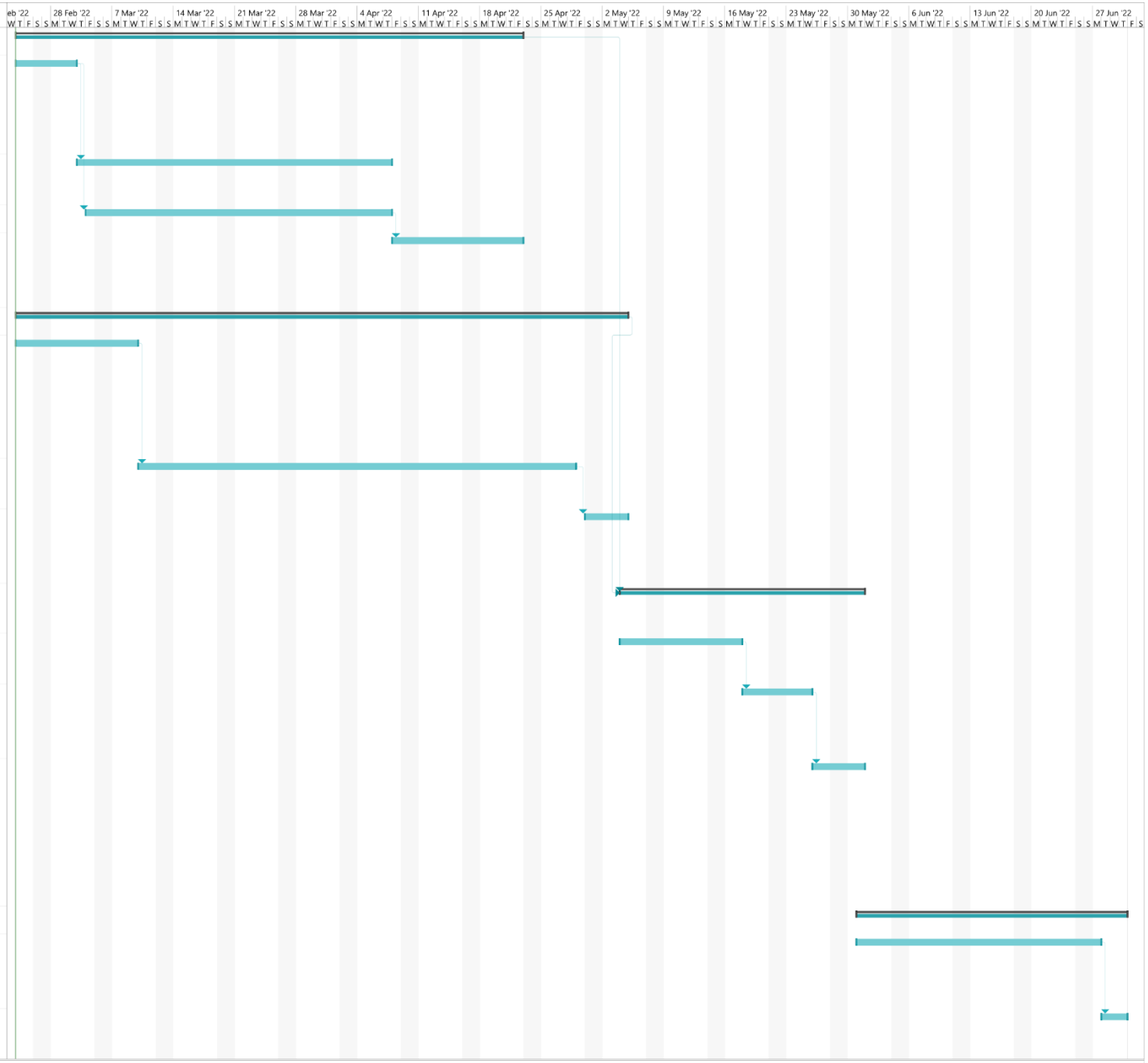
Görsel Entegrasyon ve Metin Entegrasyonu kısımları ortak hareket edebilmelidir, yani sohbet robotunun geri dönütünde etkileşim unsurunda görülecek metinle birlikte yüz canlandırılması yapılmış kişide metin bazlı mimik ve yüz hareketleri gözlemlemiş olacağız.

5. Planlanan Çalışmalar

Yalnızca chatbot üzerine yoğunlaşmış framework çeşitleri bulunmaktadır. Bunlardan en çok kullanılanlar arasında Google Dialogflow, Rasa, IBM Watson Assistant yer almaktadır. Açık kaynak kodlu olan Rasa üzerinde çalışmalarımızın devam ettirilebileceği ve daha detaylı araştırılabileceği düşünüldü. Rasa Core kısmının python dilinde yazılmış olması GAN kısmında da aynı dili kullanacağımız nedeniyle potansiyel seçenek olarak düşünülmektedir.

Planlanan çalışmalarımızda baz alacağımız planımız Gantt Şeması halinde bir sonraki sayfada yer almaktadır.

ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors
1		**1. Chatbot Tasarımı**	**42 days**	**Thu 24.02.22**	**Fri 22.04.22**	
2		1.1 Tasarım için seçilmiş mimari veya framework seçeneklerinin değerlendirilip son kararın verilmesi	5 days	Thu 24.02.22	Wed 2.03.22	
3		1.2 Veri bulunması, eklenmesi, temizlenmesi	26 days	Thu 3.03.22	Thu 7.04.22	2
4		1.3 Modelin geliştirilme	25 days	Fri 4.03.22	Thu 7.04.22	2
5		1.4 Testlerin uygulanması ve modelin son halinin elde edilmesi	11 days	Fri 8.04.22	Fri 22.04.22	4
6		**2. GAN Tasarımı**	**50 days?**	**Thu 24.02.22**	**Wed 4.05.22**	
7		2.1 GAN örnekleri içerisinde projedeki amacımıza uygun seçeneklerin değerlendirilmesi ve seçimin yapılması	10 days	Thu 24.02.22	Wed 9.03.22	
8		2.2 GAN tasarımının gerçekleştirilmesi	36 days	Thu 10.03.22	Thu 28.04.22	7
9		2.3 GAN tasarımının son halinin elde edilmesi, eksiklerin tamamlanması	4 days	Sat 30.04.22	Wed 4.05.22	8
10		**3. Chatbot ve GAN Birleştirilmesi**	**20 days?**	**Wed 4.05.22**	**Tue 31.05.22**	**1;6**
11		3.1 İki temel unsurun birleştirilmesi	10 days	Wed 4.05.22	Tue 17.05.22	
12		3.2 Birleştirme esnasında çıkacak problemlerin değerlendirilmesi ve test süreci	6 days	Wed 18.05.22	Wed 25.05.22	11
13		3.3 Yeni eklenebilecek özelliklerin değerlendirilmesi (Örn. Chatbot için konuşmalara ses desteğinin eklenmesi, görüntüden duygu analizi vs.)	4 days	Thu 26.05.22	Tue 31.05.22	12
14		**4. Deployment**	**23 days**	**Tue 31.05.22**	**Thu 30.06.22**	
15		4.1 Kullanıcıya uygulamanın sunulacağı kanalın değerlendirilmesi (Web, masaüstü vb.)	20 days	Tue 31.05.22	Mon 27.06.22	
16		4.2 Uygulamanın son kullanıcıya ulaşması	3 days	Tue 28.06.22	Thu 30.06.22	15



Project: Project1
Date: Thu 24.02.22

Task
Split
Milestone
Summary
Project Summary
Inactive Task
Inactive Milestone
Inactive Summary
Manual Task
Duration-only
Manual Summary Rollup
Manual Summary
Start-only
Finish-only
External Tasks
External Milestone
Deadline
Progress
Manual Progress

Kaynaklar

[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. Oyebode, 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", August 2021 arxiv.org

[10] Pavel Savchenkov, Maxim Lukin, Aleksandr Mashrabov, "Text and Audio-Based Real-Time Face Reenactment", United States Patent Application Publication, 23 July 2020

[11] Yingruo Fan, Zhaojiang Lin, Jun Saito, Wenping Wang, Taku Komura, "FaceFormer: Speech-Driven 3D Facial Animation with Transformers", December 2021 arxiv.org

[12] Understanding LSTM Networks (2015, 27 August) https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[13] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Networks", 10 June 2014, arxiv.org

[14] Soumya Tripathy, Juho Kannala, Esa Rahtu, "FACEGAN: Facial Attribute Controllable rEenactment GAN", 9 Nov 2020, arvix.org