

28/03/2024

Konuşma Tanıma ve Sentezleme Teknolojisi

Bilgisayar Mühendisliği 3. Sınıf
Öğrencisi
Batuhan GENÇ
21360859081



**BURSA TEKNİK
ÜNİVERSİTESİ**

1.Konuşma Tanıma Teknolojisi

- Konuşma Tanıma Nedir1
- Uygulama Alanları2
- Konuşma Tanıma Türleri.....3
 - Konuşmacıya Bağlı, Konuşmacıdan Bağımsız, Sürekli Tanıma4
 - Ayırık Konuşma Tanıma, Geniş Kelime Dağarcığı,Komuta ve Kontrol5
 - NLP Konuşma Tanıma, Uzak Alan Konuşma Tanıma, Yakın Alan6
 - Gömülü ve Bulut Tabanlı, Derin Öğrenme Tabanlı, Hibrit Sistemler7
- Konuşma Tanıma Teknolojisi Kullanan Programlar8
 - Dragon NaturallySpeaking9
 - IBM Watson Speech to Text10

2.Konuşma Sentezleme Teknolojisi

- Konuşma Sentezi Nedir11
- Konuşma Sentezi Nasıl Çalışır12
 - Metinden Kelimelere13
 - Kelimelerden Fonemlere14
 - Sesbirimlerden Sese15
- Şarkı Söyleyen İlk Bilgisayar16

Konuşma Tanıma Nedir?



Konuşma tanıma, bilgisayarların insan konuşmasını anlama ve işleme yeteneğini ifade eder. Bu teknoloji, ses sinyallerini analiz ederek konuşulan sözcükleri tanımlar ve anlamak için dilbilimsel kuralları ve matematiksel algoritmaları kullanır. Genellikle **doğal dil işleme (NLP)** alanında bir alt dal olarak kabul edilir.

Uygulama Alanları



- **Sanal Asistanlar:** Siri , Alexave Google Assistant gibi sesle etkinleştirilen asistanlara güç vermeyi içerir.
- **Transkripsiyon hizmetleri:** Konuşulan içeriğin belgeler, altyazılar veya başka amaçlar için yazılı metne dönüştürülmesini içerir.
- **Sağlık:** Doktorların ve hemşirelerin hasta notlarını ve kayıtlarını eller serbest olarak dikte etmelerini sağlar.
- **Otomotiv:** Müzik çalmaktan navigasyona kadar araçlarda sesle etkinleştirilen kontrollerin etkinleştirilmesini kapsar.
- **Müşteri Hizmetleri:** Çağrı merkezlerinde sesle etkinleştirilen IVR'lara güç vermeyi kapsar.
- **Eğitim:** Dil öğrenme uygulamalarında kolaylaştırmak, telaffuza yardımcı olmak ve anlama alıştırmaları içindir.
- **Oyun:** Daha sürükleyici bir deneyim için video oyunlarında sesli komut yetenekleri sağlamayı içerir.



Konuşma Tanıma Türleri

İşlevler, konuşma tanımlama ve konuşmacı tanıma sistemlerini içerir. Mevcut konuşma tanıma yazılımlarının çeşitliliği, farklı ihtiyaçlara ve kullanımlara hitap eder.

Listelenen 12 tür konuşma tanıma--->



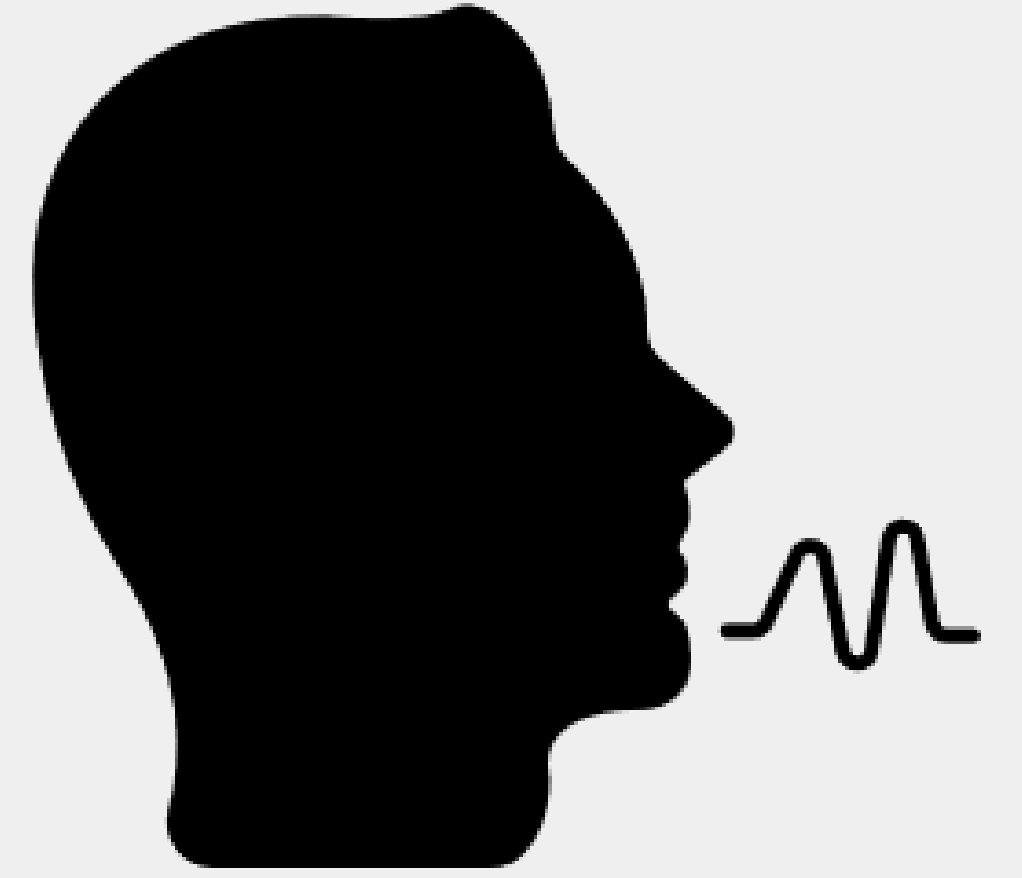
Konuşmacıya Bağlı Konuşma Tanıma:

Konuşmacıya Bağlı Konuşma Tanıma sistemleri, bireysel bir kullanıcının benzersiz ses özelliklerini öğrenir ve bunlara uyum sağlar



Konuşmacıdan Bağımsız Konuşma Tanıma:

Konuşmacıdan Bağımsız Konuşma Tanıma sistemleri, önceden eğitime ihtiyaç duymadan herhangi bir kullanıcıdan gelen konuşmayı anlar ve işler.



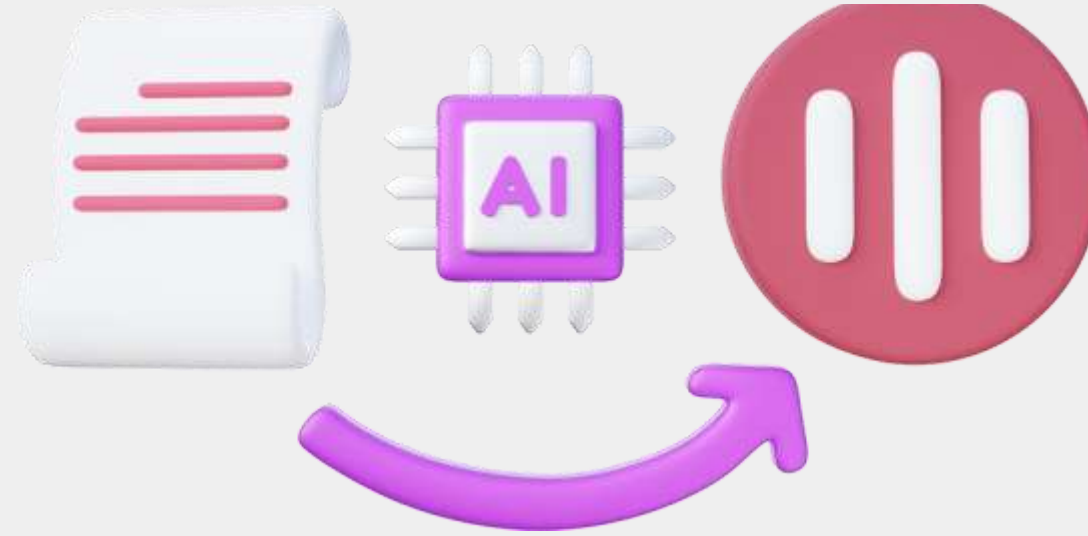
Sürekli Konuşma Tanıma:

Sürekli Konuşma Tanıma sistemleri, doğal, akıcı konuşmayı doğru bir şekilde işler ve yazıya döker.



Ayrık Konuşma Tanıma

Ayrık Konuşma Tanıma sistemleri, kullanıcıların doğru tanıma için kelimeleri aralarında duraklamalarla ayrı ayrı konuşmasını gerektirir.



Geniş Kelime Dağarcığı Sürekli Konuşma Tanıma (LVCSR):

Geniş Kelime Dağarcığı Sürekli Konuşma Tanıma (LVCSR) sistemleri, çok çeşitli kelime dağarcığı ile konuşmayı doğal bir akışta işler ve anlar.



Komuta ve Kontrol Konuşma Tanıma:

Komuta ve Kontrol Konuşma Tanıma sistemleri, belirli sesli komutları tanır ve ilgili eylemleri veya kontrolleri yürütür.



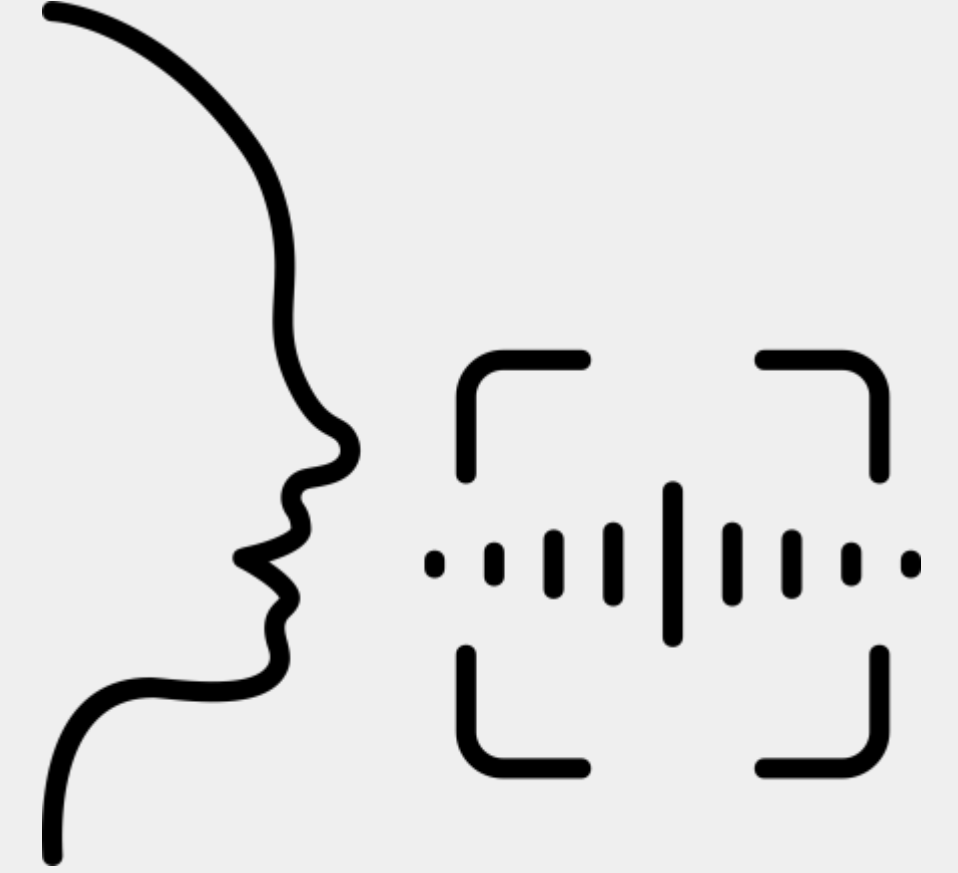
Doğal Dil İşleme (NLP)-Gelişmiş Konuşma Tanıma:

Doğal Dil İşleme (NLP)-Gelişmiş Konuşma Tanıma sistemleri, gelişmiş NLP teknikleri kullanarak konuşulan dili yorumlar ve analiz eder.



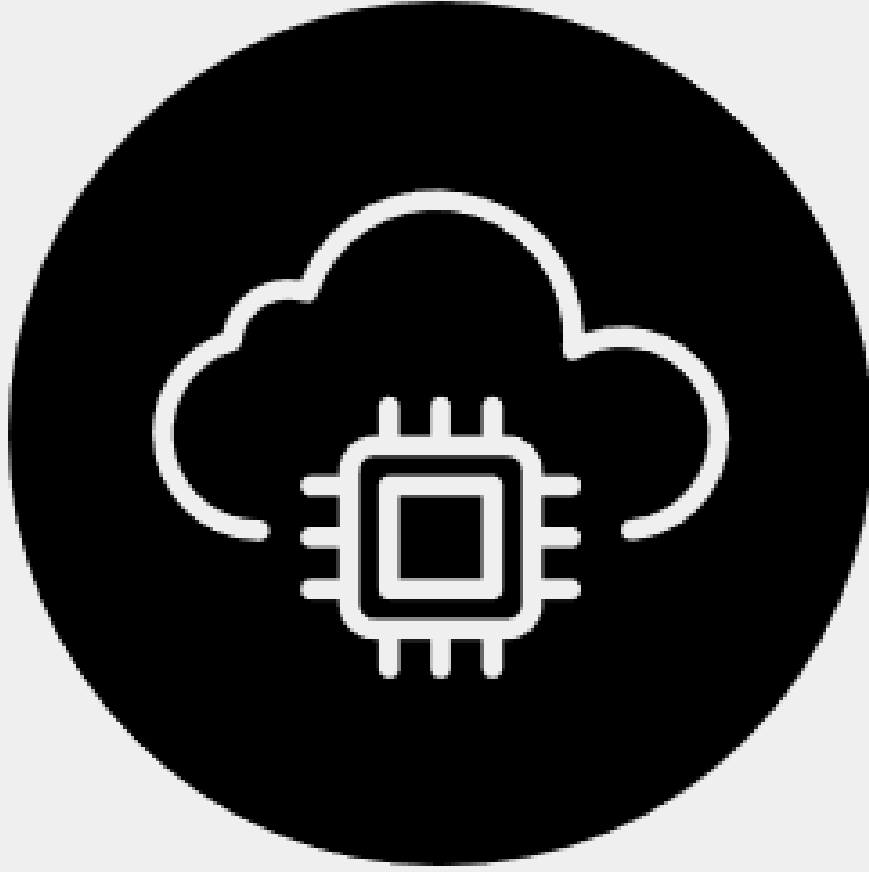
Uzak Alan Konuşma Tanıma:

Uzak Alan Konuşma Tanıma sistemleri, konuşmayı uzaktan doğru bir şekilde yakalar ve işler, arka plan gürültüsünün ve oda akustığının üstesinden gelir.



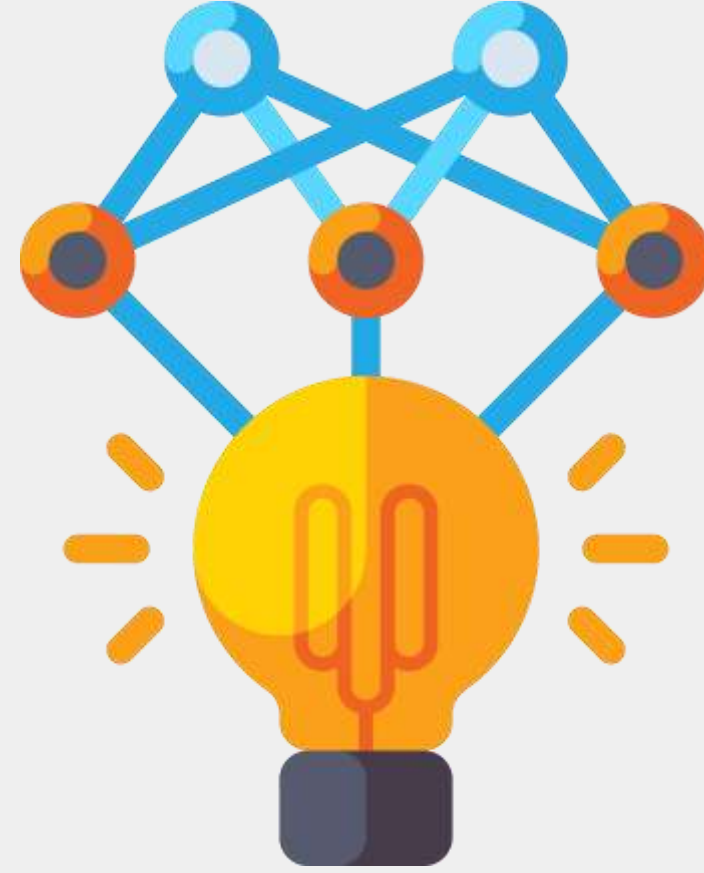
Yakın Alan Konuşma Tanıma:

Yakın Alan Konuşma Tanıma sistemleri, konuşmayı yakın bir mesafeden, tipik olarak mikrofonun birkaç fit yakınından doğru bir şekilde işleme konusunda uzmanlaşmıştır.



Gömülü ve Bulut Tabanlı Konuşma Tanıma:

Gömülü Konuşma Tanıma sistemleri, bir cihazda yerel olarak çalışır ve internet bağlantısına ihtiyaç duymadan sesli komutları işler.



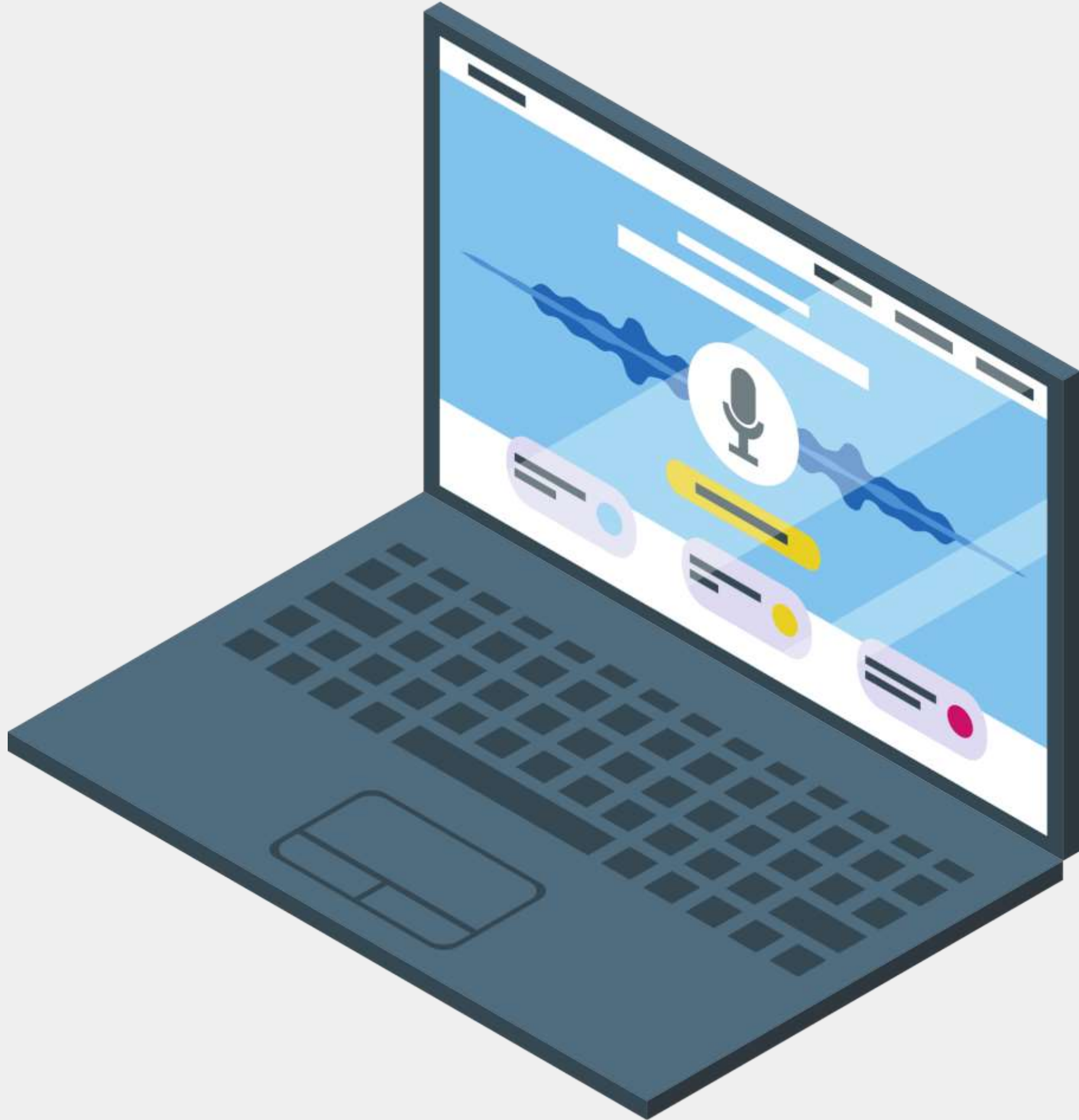
Derin Öğrenme Tabanlı Konuşma Tanıma:

Derin Öğrenme Tabanlı Konuşma Tanıma sistemleri, insan konuşmasını yüksek doğrulukla analiz etmek ve yorumlamak için gelişmiş sinir ağlarını kullanır.



Hibrit Sistemler:

Hibrit Sistemler, doğruluğu ve performansı artırmak için çeşitli konuşma tanıma teknolojilerinin güçlü yönlerini birleştirir.



Konuşma Tanıma Teknolojisi Kullanan Programlardan Bazıları:

- Google Assistant
- Apple Siri
- Amazon Alexa
- Microsoft Cortana
- Dragon NaturallySpeaking
- IBM Watson Speech to Text

Dear Mr. Bossman,

I am writing to address the staffing shortfall in our company you to propose several strategies that we believe help address this issue.

Firstly, we propose the use of incidental supplemental workers to fill in the gaps in our workforce. These workers can be hired on a part-time or temporary basis to cover any staffing shortfalls that may arise. We believe that this will help us maintain a consistent level of service to our customers and to ensure that our operations continue to run smoothly. Secondly, we propose increasing our recruiting efforts to attract more qualified candidates to our company. This can be achieved through targeted advertising, job fairs, and outreach to educational institutions. By expanding our pool of potential hires, we can increase our chances of finding the right fit for our company.

Thirdly, we propose reviewing our current benefits package to ensure that it is competitive in the current job market. By offering attractive benefits, we can improve our employee retention rates and reduce turnover. This will help us retain our top talent and reduce the need for new hires.

Lastly, we propose offering overtime pay to our existing staff to help us cover any immediate staffing shortfalls. By compensating our employees for the extra work they put in, we can incentivize them to work overtime and to ensure that our operations continue to run smoothly.

Dear Mr. Bossman,

I am writing to address the staffing shortfall in our company and to propose several strategies that we believe will help us address this issue.

Firstly, we propose the use of incidental supplemental workers to fill in the gaps in our workforce. These workers can be hired on a part-time or temporary basis to cover any staffing shortfalls that may arise. We believe that this will help us maintain a consistent level of service to our customers and to ensure that our operations continue to run smoothly. Secondly, we propose increasing our recruiting efforts to attract more qualified candidates to our company. This can be achieved through targeted advertising, job fairs, and outreach to educational institutions. By expanding our pool of potential hires, we can increase our chances of finding the right fit for our company.

Thirdly, we propose reviewing our current benefits package to ensure that it is competitive in the current job market. By offering attractive benefits, we can improve our employee retention rates and reduce turnover. This will help us retain our top talent and reduce the need for new hires.

Lastly, we propose offering overtime pay to our existing staff to help us cover any immediate staffing shortfalls. By compensating our employees for the extra work they put in, we can incentivize them to work overtime and to ensure that our operations continue to run smoothly.

Watson Speech to Text Demo

In this demo, you will see Watson Speech to Text transcribe a support call from a call center. On the left, notice how the base model does not recognize certain product names. But, as seen on the right, a user can improve transcript accuracy by using the Watson Speech to Text training tools. At the bottom, notice how the confidence scores improve with the trained model.

Play audio 

Recognize Speakers ⓘ



Base Model

JSON 

With User Training ⓘ

JSON 

Speaker 1: Welcome to the bottom online how may I help you.

Speaker 2: Order trip for robot vacuum cleaner last week which got delivered yesterday I can cleaner appears to be damaged I was wondering if I can get a refund.

Speaker 1: If the product is damaged you can return the product to get a full refund do you want me to process the return.

Speaker 2: That would be awesome hi Michael walk water
~~disappears in the same order I would like to keep that~~

Alternatives ⓘ

Word Timings ⓘ



to 100.0%

the 97.0%, is 1.0%, use 1.0%

pot 41.0%, bottom 20.0%, bought 11.0%, product 9.0%

, part 6.0%, bar 2.0%, pod 2.0%, potter 1.0%

online 90.0%

how 91.0%

may 100.0%

Speaker 1: Welcome to the bottom online how may I help you.

Speaker 2: Order trip for robot vacuum cleaner last week which got delivered yesterday I can cleaner appears to be damaged I was wondering if I can get a refund.

Speaker 1: If the product is damaged you can return the product to get a full refund do you want me to process the return.

Speaker 2: That would be awesome hi Michael walk water
~~disappears in the same order I would like to keep that~~

Alternatives ⓘ

Word Timings ⓘ



to 100.0%

the 97.0%, is 1.0%, use 1.0%

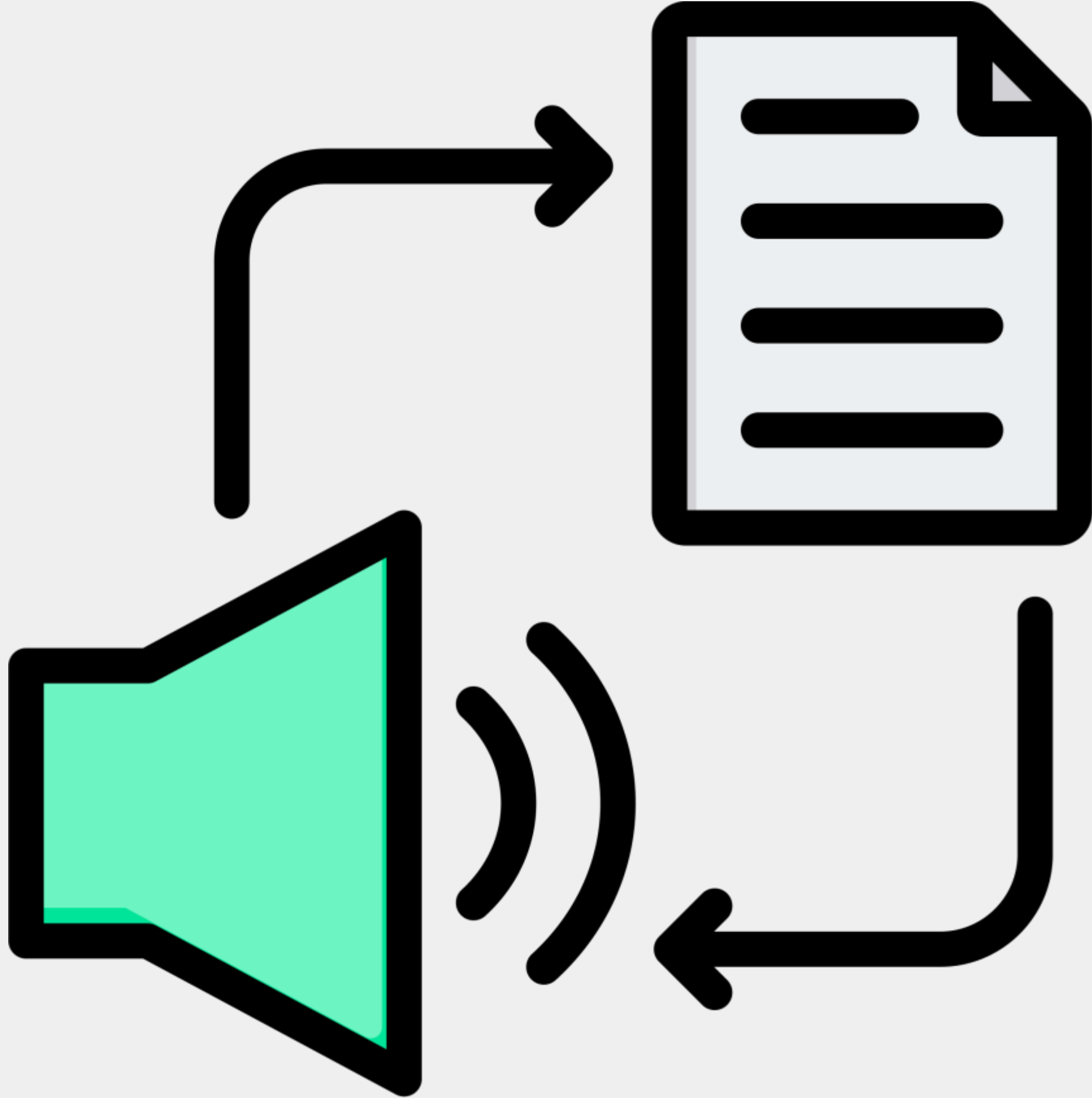
pot 41.0%, bottom 20.0%, bought 11.0%, product 9.0%

, part 6.0%, bar 2.0%, pod 2.0%, potter 1.0%

online 90.0%

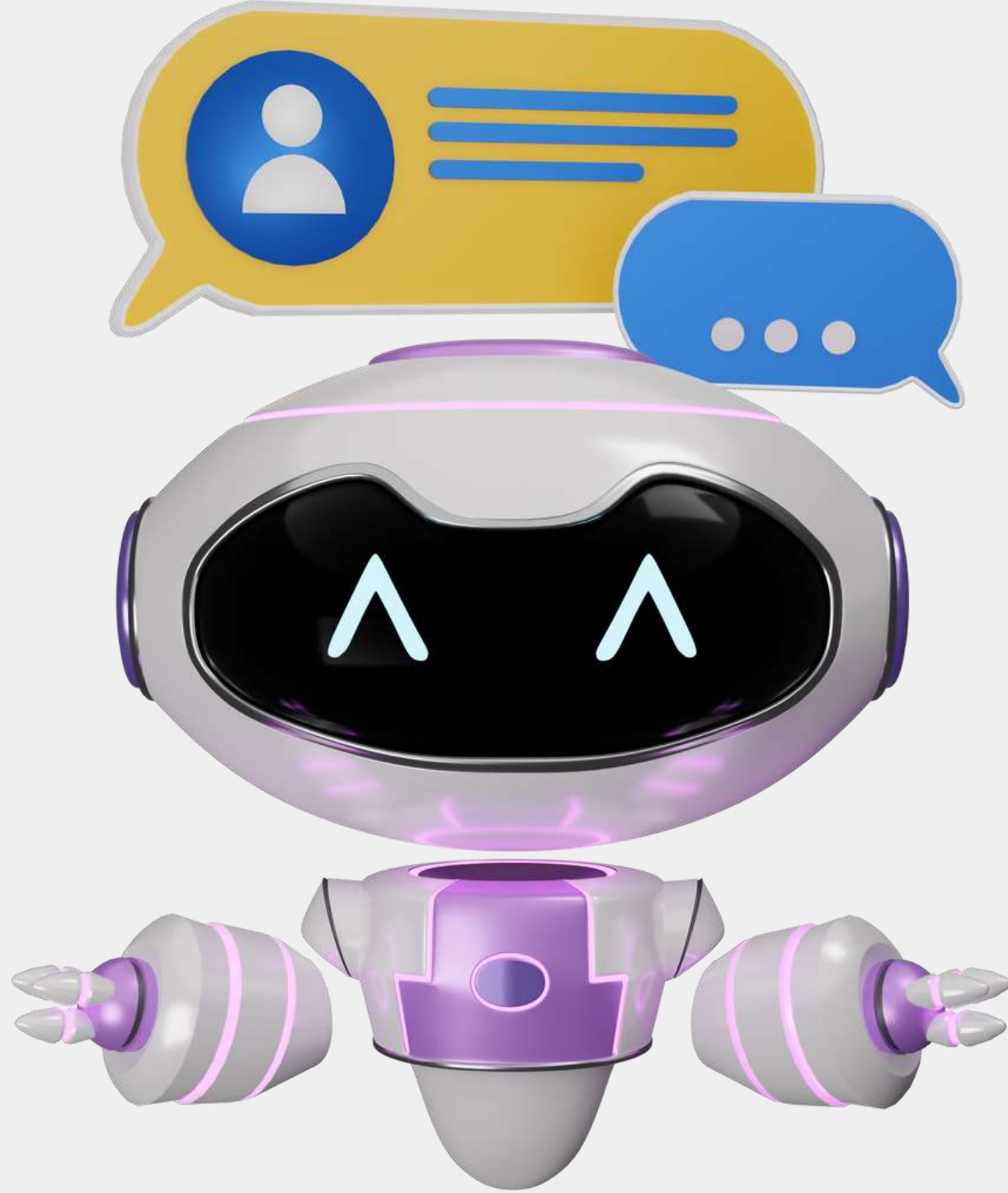
how 91.0%

may 100.0%



Konuşma Sentezi Nedir?

Konuşma sentezleyici , İnsan seslerinin bilgisayar yapay zekası tarafından üretimidir. Genellikle Metin okuma sistemlerinde kullanılan Konuşma sentezleyicisi bu sistemlerde bir dilin tüm ses bilgisi ve benzeri kurallarına göre çevri yapacak şekilde programlanırlar. Bu sentezleyicinin metni doğru telaffuz etmesini sağlar.

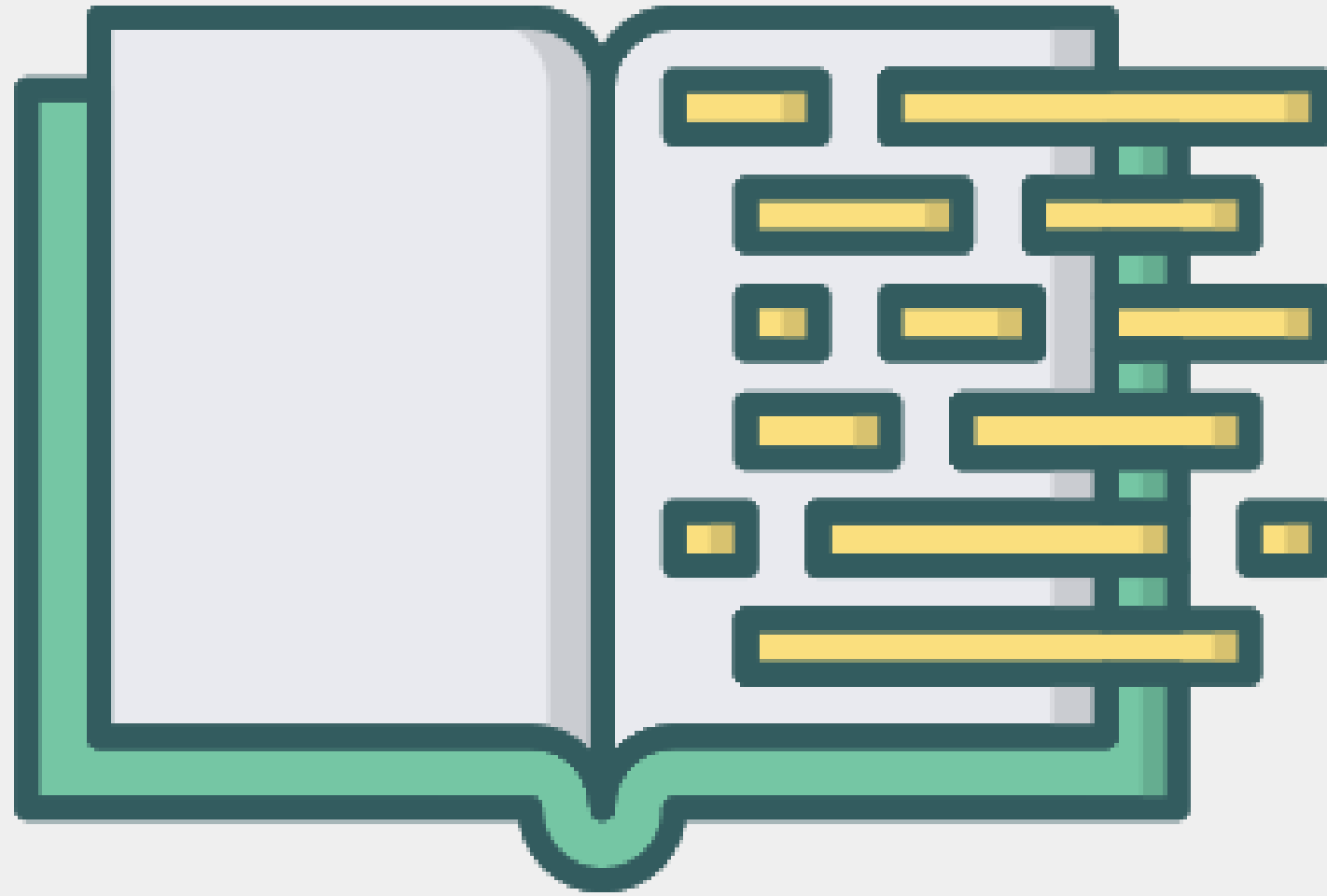


Konuşma Sentezi Nasıl Çalışır

Konuşma Sentezi Üç Aşamada Çalışır:

- Metinden kelimelere
- Kelimelerden fonemlere
- Fonemlerden sese.

Metinden Kelimelere



Konuşma sentezi, bir pasajı okumanın en iyi yolunu seçerek belirsizliği azaltan ön işleme veya normalleştirme ile başlar. Ön işleme, bilgisayarın metni daha doğru okuması için metnin okunmasını ve temizlenmesini içerir. Sayılar, tarihler, saatler, kısaltmalar, akronimler ve özel karakterlerin çevrilmesi gerekir. En olası telaffuzu belirlemek için istatistiksel olasılık veya sinir ağları kullanırlar.

Kelimelerden Fonemlere

Kelimeler belirlendikten sonra, konuşma sentezleyici bu kelimeleri içeren sesler üretir. Her bir kelimenin sesini oluşturan fonemlerin bir listesine ihtiyaçları olacaktır. İngilizce alfabede sadece 26 harf vardır ancak 40'tan fazla fonem vardır.

Teorik olarak, eğer bir bilgisayarın kelime ve fonemlerden oluşan bir sözlüğü varsa, tek yapması gereken bir kelimeyi okumak, sözlükte aramak ve ardından karşılık gelen fonemleri okumaktır. Ancak pratikte bu durum görüldüğünden çok daha karmaşıktır.

Sesbirimlerden sese

Bilgisayar artık metni bir fonem listesine dönüştürmüştür. Peki bilgisayarın farklı dillerde metni konuşmaya dönüştürürken yüksek sesle okuduğu temel fonemleri nasıl elde ediliyor? Bu konuda üç yaklaşım bulunmaktadır.

- **İlk yaklaşım** sesleri söyleyen insanların kayıtları kullanılacaktır.
- **İkinci yaklaşım**, bilgisayarın temel ses frekanslarını kullanarak fonemler üretmesidir.
- **Son yaklaşım**, yüksek kaliteli algoritmalarla doğal seslendirme yaparak insan sesi tekniğini gerçek zamanlı olarak taklit etmektir.



**in late 50's a
computer
sung for the
first time...**

**Şarkı
Söyleyen
İlk
Bilgisayar**

Kaynakça

Yalçın, Nursel «KONUŞMA TANIMA TEORİSİ VE TEKNİKLERİ» Mart 2008 Cilt:16 No:1 Kastamonu Eğitim Dergisi 249-266

https://en.wikipedia.org/wiki/Dragon_NaturallySpeaking

<https://www.ibm.com/demos/live/speech-to-text/self-service>

<https://tr.shaip.com/blog/what-is-audio-annotation/>

<https://transkriptor.com/tr/konusma-tanima/>

<https://transkriptor.com/tr/konusma-tanima-turleri/>

<https://zerzavot.medium.com/speech-recognition-7ef8682806b4>

https://tr.wikipedia.org/wiki/Konu%C5%9Fma_sentezleyici

<https://ilhanmertalan.medium.com/konu%C5%9Fma-sentezleme-derin-%C3%B6%C4%9Frenme-t%C3%BCrk%C3%A7e-d5ec7887de4f>

<https://speaktor.com/tr/konusma-sentez/>

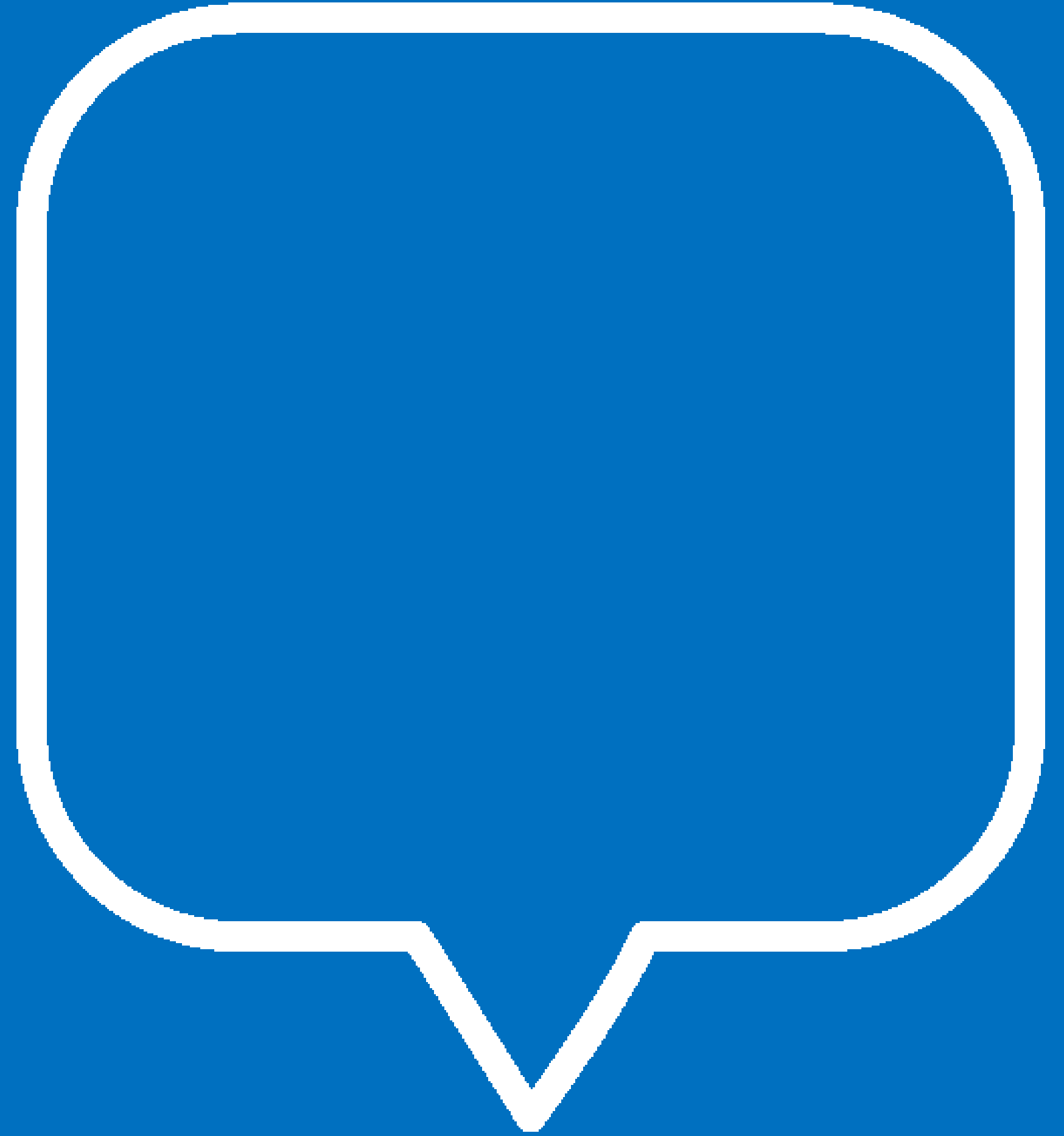
<https://yazokulu.bilimakademisi.org/yapayogrenme/2018/sunumlar/cenkdemiroglu-byoyo18.pdf>

<https://yazokulu.bilimakademisi.org/yapayogrenme/2018/sunumlar/muratsaraclar-byoyo18.pdf>

<https://www.youtube.com/watch?v=5bBVhBnqVA0>

<https://www.youtube.com/watch?v=x3bPY6i8vJE>

Soru Cevaplama Bölümü



Beni Dinlediđiniz İin

Teşekkürler!