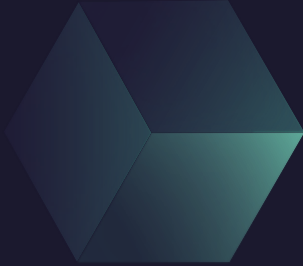


Ders 3

Doğal Dil İşlemeye Giriş



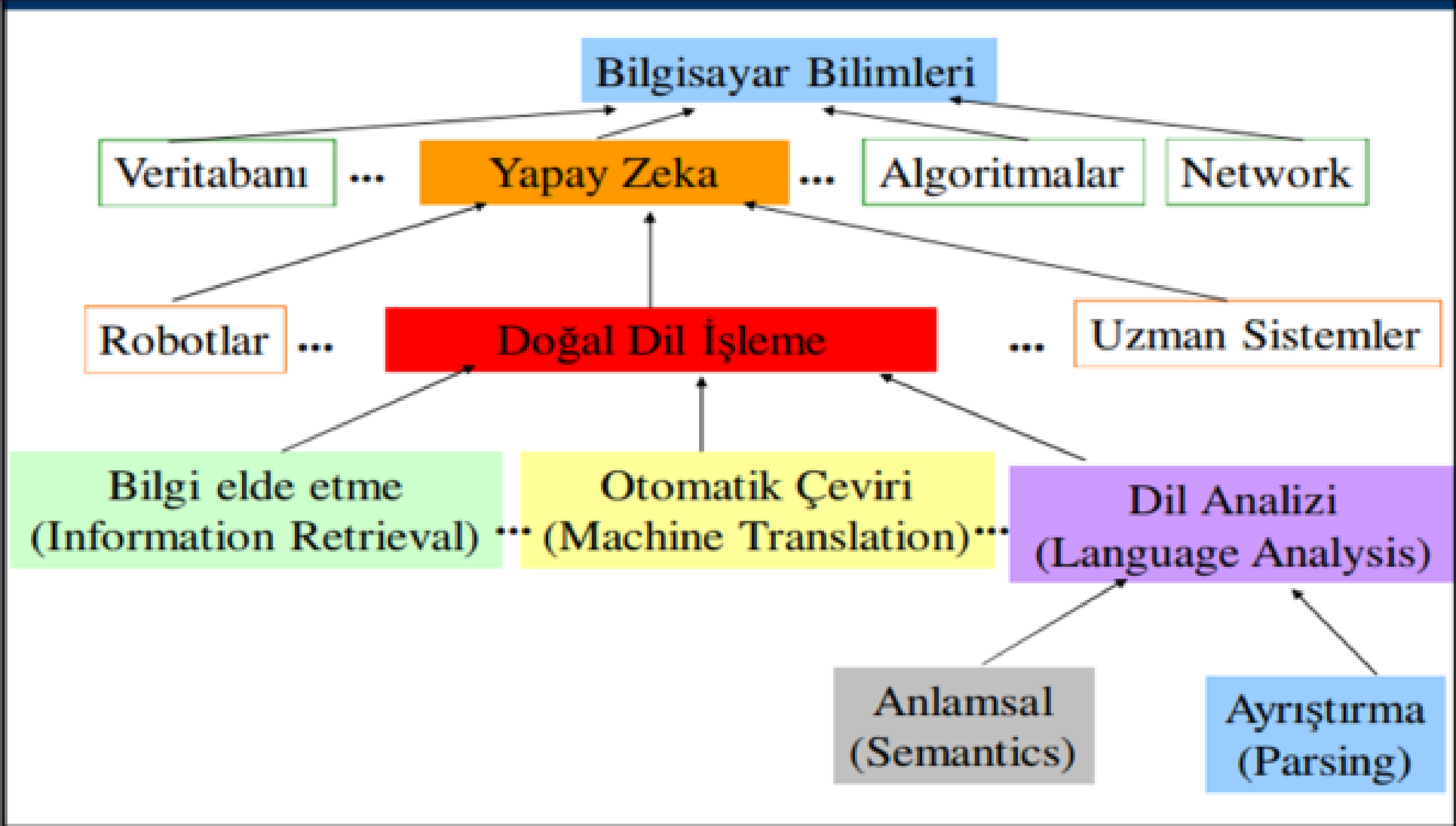
Doğal Dil İşleme (NLP) Nedir?



Doğal Dil İşleme (NLP) Nedir? başlığı altında şu bilgileri ekleyebiliriz: Doğal Dil İşleme (NLP), bilgisayarların insan dilini anlaması, yorumlaması ve işlemesiyle ilgilenen bir yapay zeka ve dilbilim alanıdır. NLP teknolojisi, insan iletişimi ile bilgisayar anlayışı arasındaki boşluğu doldurmayı amaçlar. NLP'nin kapsamı:

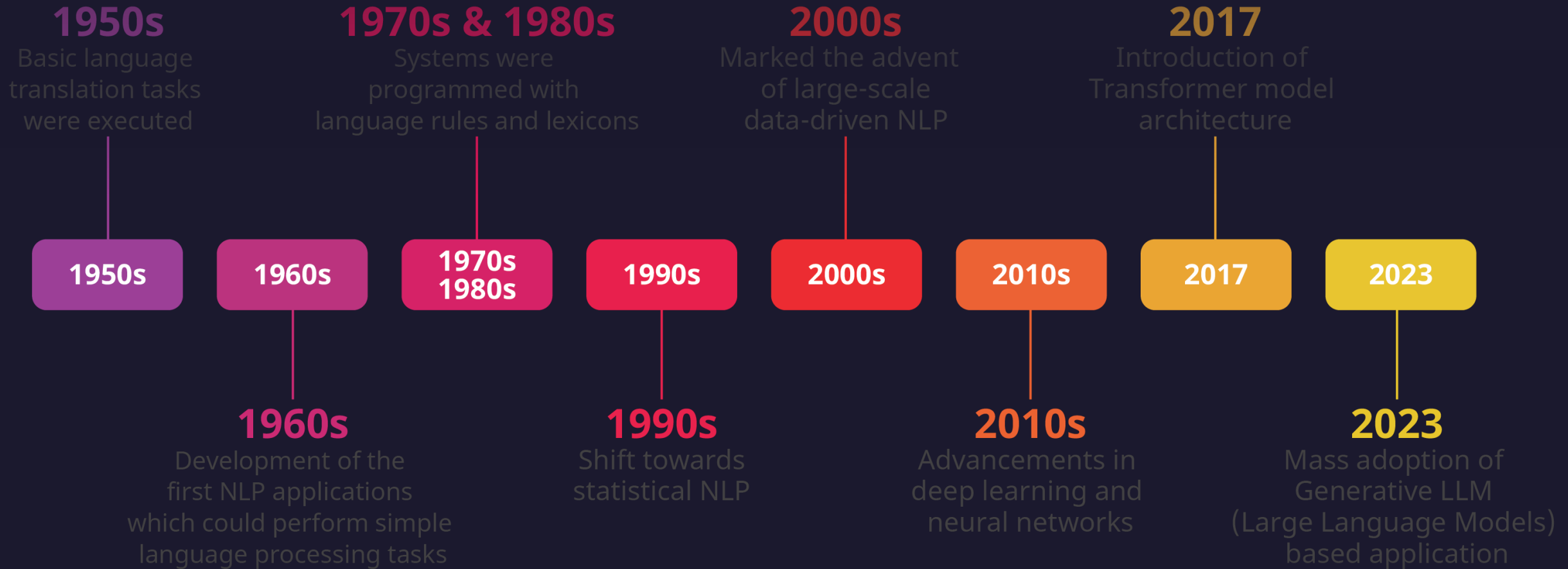
- **Metin verisi:** NLP, yazılı metinlerin analizi ve anlamlandırılmasını içerir.
- **Konuşma verisi:** Ses de NLP'nin bir parçasıdır. Konuşma tanıma ve ses sentezi gibi uygulamalar NLP kapsamına girer.

NLP'nin temel amacı, makinelerin insan diliyle nasıl etkileşim içerisine girdiğini incelemek ve geliştirmektir. Bu teknoloji, günlük hayatımızda sıkça kullandığımız çeviri uygulamaları, sanal asistanlar ve arama motorları gibi birçok uygulamanın temelini oluşturur. NLP, sürekli gelişen bir alan olup, yapay zeka ve makine öğrenimi teknolojilerindeki ilerlemelerle birlikte giderek daha sofistike hale gelmektedir.



NLP'nin Tarihçesi ve Gelişimi

Genesis and Evolution of Natural Language Processing



NLP'nin Temel Kavramları



Tokenizasyon

Lemmatizasyon

POS (Part-of-Speech)
Tagging

Tokenizasyon

Tiktokenizer

System

Sen doktor asistanısın.

×

User

Bana hasta kayıtlarını getir.

×

Add message

<|im_start|>system<|im_sep|>Sen doktor asistanısın.<|im_end|>
<|im_start|>user<|im_sep|>Bana hasta kayıtlarını getir.
<|im_end|><|im_start|>assistant<|im_sep|>

Token count

25

<|im_start|>system<|im_sep|>Sen doktor asistanısın.<|im_end|>
<|im_start|>user<|im_sep|>Bana hasta kayıtlarını getir.<|im_end|>
<|im_start|>assistant<|im_sep|>

200264, 17360, 200266, 28315, 85571, 472, 12560, 2196
6, 3681, 13, 200265, 200264, 1428, 200266, 33, 1480, 1
1783, 173791, 45919, 116584, 13, 200265, 200264, 17378
1, 200266

Show whitespace

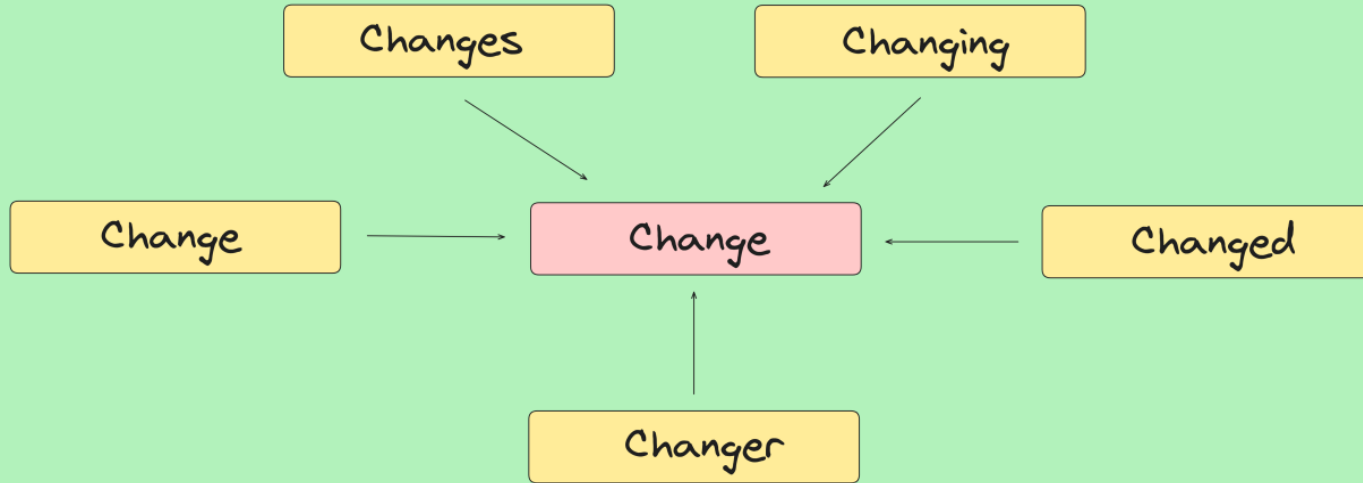
Tokenizasyon, metin verisini daha küçük parçalara (token'lara) ayırma işlemidir. Bu işlem, genellikle bir cümleyi kelimelere veya karakterlere bölmeyi içerir. Tokenizasyon, NLP'de her işlemin ilk adımlarından biridir. Örneğin:

Cümle: "NLP çok ilginç bir alan."

Tokenlar:
["NLP", "çok", "ilginç", "bir", "alan", "."]

<https://tiktokenizer.vercel.app>

Lemmatizasyon



Lemmatizasyon, bir kelimenin çekimli formunu (örneğin, çoğul, geçmiş zaman) temel veya sözlük formuna (lemma) dönüştürme işlemidir. Bu işlem, kelimenin anlamını ve bağlamını korur.

Mesela:

- "anlattı" -> "anlatmak"
- "gidiyoruz" -> "gitmek"
- "çocuklar" -> "çocuk"
- "güzeldi" -> "güzel"

POS (Part-of-Speech) Tagging

POS tagging, bir metindeki her kelimenin dilbilgisel rolünü (isim, fiil, sıfat vb.) belirleme işlemidir. Bu işlem, kelimelerin cümle içindeki işlevlerini anlamak için çok önemlidir. Örneğin:

Cümle: "**Küçük kedi bahçede kuşları izliyor.**"

Küçük->SIFAT

kedi->İSİM

bahçede->İSİM+BULUNMA

kuşları->
İSİM+BELİRTME+ÇOĞUL

izliyor->FİİL+ŞİMDİKİ ZAMAN

NLP'nin Uygulama Alanları

Makine Çevirisi: NLP, bir dilden diğerine otomatik çeviri yapan sistemlerin temelini oluşturur.

Duygu Analizi: Sosyal medya yorumları, ürün değerlendirmeleri gibi metinlerdeki duyguları otomatik olarak tespit etmek için kullanılır.

Konuşma Tanıma: Sesli komutları metne dönüştüren sistemlerde NLP teknikleri kullanılır.

Metin Özetleme: Uzun metinleri otomatik olarak özetleyen sistemler NLP tekniklerinden yararlanır.

Soru Cevaplama Sistemleri: Chatbotlar ve sanal asistanlar gibi uygulamalar NLP sayesinde insan benzeri yanıtlar üretebilir.

Spam Filtreleme: E-posta sistemlerinde istenmeyen postaları tespit etmek için NLP kullanılır.

Otomatik İçerik Oluşturma: Haber özetleri, hava durumu raporları gibi metinleri otomatik olarak oluşturmak için NLP teknikleri kullanılır.

Tıbbi Uygulamalar: Hasta kayıtlarının analizi, hastalık teşhisi ve ilaç etkileşimlerinin belirlenmesinde NLP'den yararlanır.

Müşteri Hizmetleri: Otomatik yanıt sistemleri ve müşteri şikayetlerinin analizi için NLP kullanılır.

Eğitim: Öğrenci performansının değerlendirilmesi ve kişiselleştirilmiş öğrenme deneyimleri oluşturmada NLP teknikleri kullanılabilir.

NLP'nin Güncel Zorlukları ve Geleceği

Güncel Zorluklar:

Dil Çeşitliliği: Dünyada binlerce dil ve lehçe var. Her biri için etkili NLP modelleri geliştirmek büyük bir zorluk.

Bağlam Anlama: Makinelerin, insan dilindeki ince nüansları, ironiyi ve bağlamı anlaması hala zor.

Veri Kalitesi ve Miktarı: Yüksek kaliteli ve büyük miktarda eğitim verisi bulmak, özellikle az konuşulan diller için zorlu.

Hesaplama Gücü: Büyük dil modellerinin eğitimi ve çalıştırılması yüksek hesaplama gücü gerektiriyor.

Etik Sorunlar: Veri gizliliği, önyargı ve yanlış bilgi yayma gibi etik konular önemli zorluklar oluşturuyor.

Geleceği:

- **Daha İyi Bağlam Anlama:** Yapay zeka ve makine öğrenimi alanındaki gelişmelerle, NLP modellerinin bağlamı daha iyi anlaması öngörülüyor.
- **Daha Verimli Modeller:** Daha az veri ve hesaplama gücüyle daha etkili sonuçlar üreten modeller geliştiriliyor.
- **Multimodal NLP:** Metin, ses ve görüntüyü birlikte işleyebilen modeller üzerinde çalışılıyor.
- **Kişiselleştirilmiş NLP:** Kullanıcıya özel dil modellerinin geliştirilmesi bekleniyor.
- **Etik NLP:** Önyargıdan arındırılmış, daha şeffaf ve güvenilir NLP sistemleri üzerine odaklanılıyor.
- **Gerçek Zamanlı Çeviri:** Anlık ve doğru çeviri yapabilen sistemlerin yaygınlaşması bekleniyor.