

Bölüm
7

**Derin Öğrenme Tabanlı
Modern Dil Modelleri**

Derin Öğrenme Tabanlı Modern Dil Modelleri



- Word Embeddings (Kelimelerin Vektör Temsilleri)
- Recurrent Neural Networks (RNN)
- Long Short-Term Memory (LSTM)
- Transformers Modelleri: BERT, GPT, LLAMA

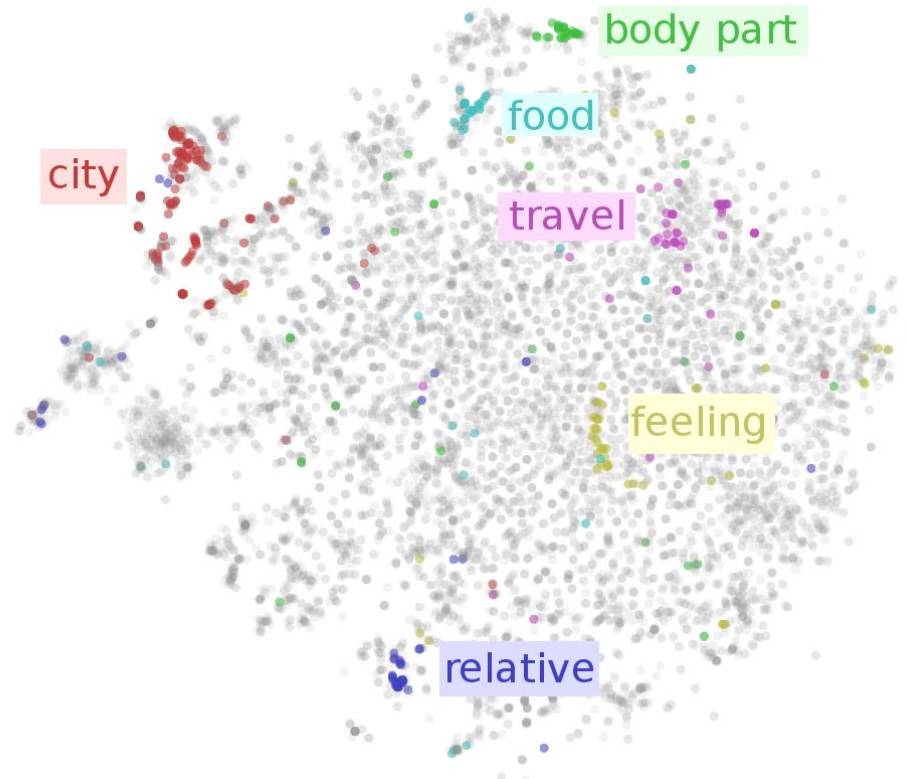


BTK
AKADEMİ

Word Embeddings

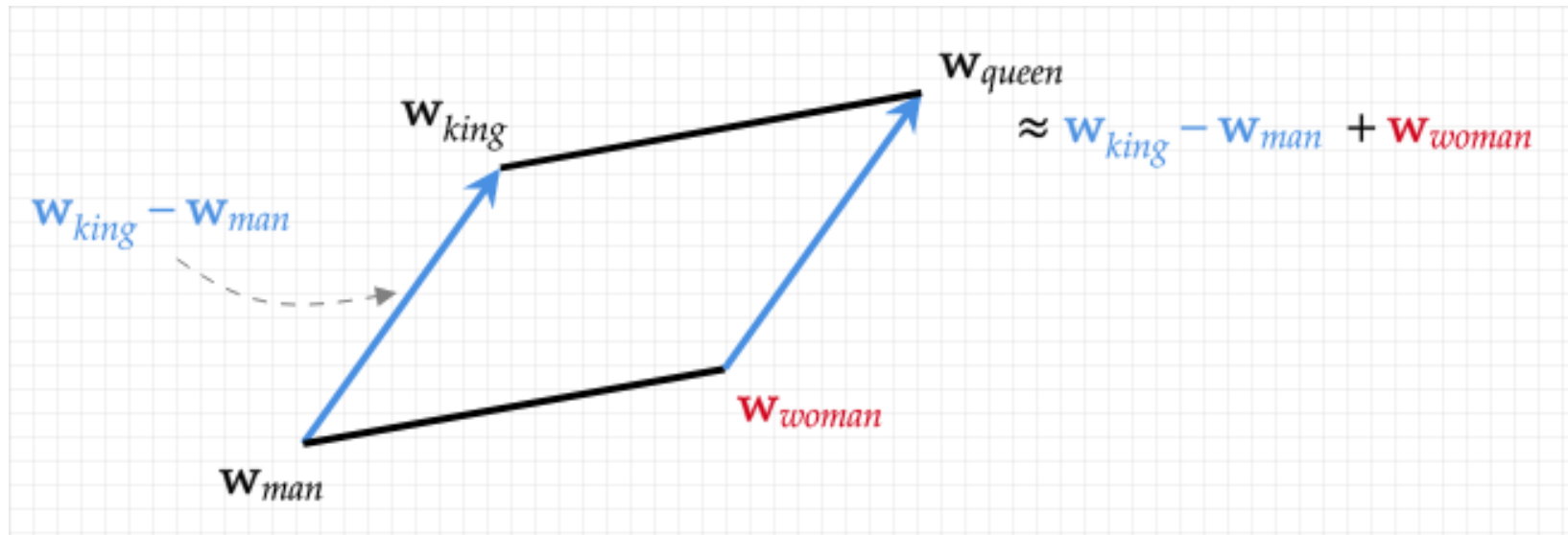
Word Embeddings (Kelimelerin Vektör Temsilleri)

- Word embeddings, kelimeleri sayısal vektörlerle ifade eden bir yöntemdir.
- Bu temsiller, kelimeler arasındaki anlamsal ilişkileri öğrenmeyi sağlar.
- Aynı anlamda veya benzer anlamda kullanılan kelimeler, vektör uzayında birbirine yakın olur.



Word Embeddings (Kelimelerin Vektör Temsilleri)

- Neden kelimeleri vektörlerle temsil etmeliyiz?
 - Dilsel Anlamı Yakalama
 - Matematiksel İşlemler
 - Verimli Temsil



<https://informatics.ed.ac.uk/news-events/news/news-archive/king-man-woman-queen-the-hidden-algebraic-struct>

Word Embeddings (Kelimelerin Vektör Temsilleri)

- **Word2Vec Nedir?**
 - **Özellikleri ve Avantajları**
 - Anlamsal yakınlıkları kelime komşuluklarına göre öğrenir.
 - Kelimeler arasında anlamlı matematiksel işlemler yapmayı sağlar.
 - **Kim Geliştirdi?**
 - Google, Tomas Mikolov ve ekibi.
 - **Kullanılan Veri Seti:** Google News corpus.
 - **Kaynak:** Haber metinleri
 - **Büyükölük:** Yaklaşık 100 milyar kelime
 - **Kapsam:** Siyaset, ekonomi, spor, teknoloji, vb.
 - **Kelimeler ve Kelime Çiftleri:** 3 milyonun üzerinde farklı kelime ve kelime çifti

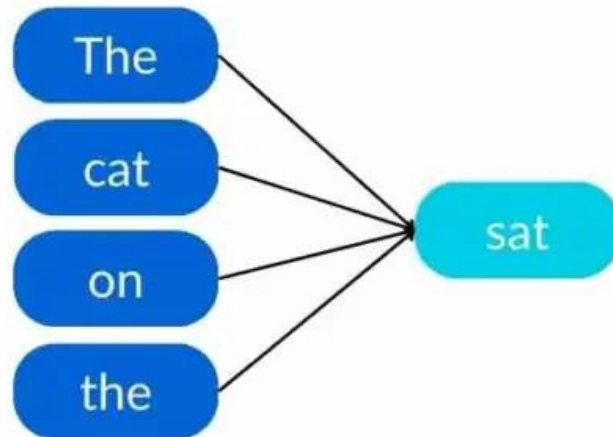
Word Embeddings (Kelimelerin Vektör Temsilleri)

- Word2Vec Nedir?
 - Temel Modelleri: CBOW (Continuous Bag of Words) ve Skip-gram.

Example Sentence: The cat sat on the mat.

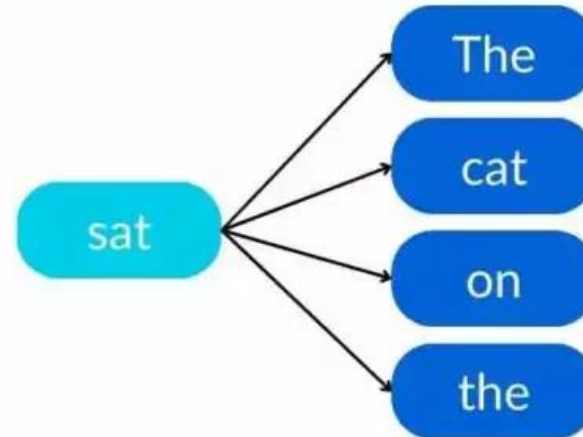
Continuous Bag-of-Words (CBOW)

Goal: Given context words,
predict the target word.



Skip-gram Model

Goal: Given a word,
predict the surrounding context words.



Word Embeddings (Kelimelerin Vektör Temsilleri)

- **FastText Nedir?**
 - **Özellikleri ve Avantajları**
 - Kelimeleri karakter seviyesinde de öğrenir, nadir kelimeler için daha güçlü bir modeldir.
 - **Kim Geliştirdi?**
 - Facebook AI.
 - **Kullanılan Veri Seti**
 - Wikipedia, Common Crawl gibi büyük veri setleri.

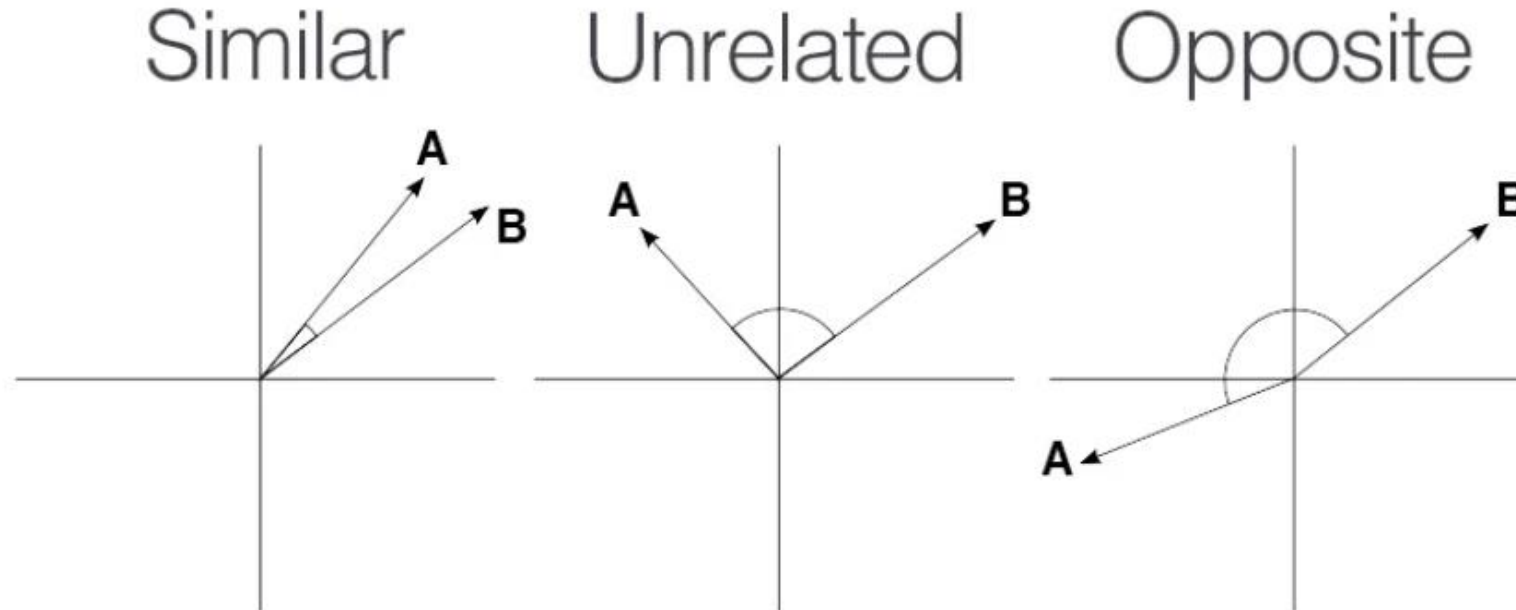
Word2Vec ve FastText Karşılaştırma

Özellik	Word2Vec	FastText
Temel Yöntem	Kelimeler arasındaki komşuluğu kullanır	Kelimeleri karakter düzeyinde temsil eder
Kapsam	Sabit kelime vektörleri	Alt-kelime bilgisiyle nadir kelimeleri daha iyi öğrenir
Kim Geliştirdi?	Google	Facebook
Kullanılan Veri	Google News Corpus	Wikipedia, Common Crawl

Word Embeddings (Kelimelerin Vektör Temsilleri)

- **Vektör Uzayı ve Anlam Yakınlığı Nedir?**

- Kelimelerin vektör temsilleri, anlamlarına göre vektör uzayında konumlandırılır.
- Anlamca benzer kelimeler (örneğin, "kedi" ve "köpek") vektör uzayında birbirine yakın olurken, farklı anlamdaki kelimeler (örneğin, "kedi" ve "araba") uzakta yer alır.
- Yakınlık ve uzaklık ayırımı Kosinüs Benzerliği ile yapılabilir.



Word Embeddings (Kelimelerin Vektör Temsilleri)

- Benzerliklerin Görselleştirilmesi (t-SNE, PCA)





BTK
AKADEMİ

Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNN)

- Zaman Serisi ve Sekans Verisi
- Standart Sinir Ağları Zaman Serilerinden Neden Yetersiz
- RNN Nedir?
- RNN Mimarisi Nedir ve RNN Nasıl Çalışır?
- Vanishing Gradient Problemi
- RNN ile NLP Alanında Yapılan Uygulamalar

Recurrent Neural Networks (RNN)

- Zaman Serisi ve Sekans Verisi Nedir?
 - Zaman Serisi Verisi
 - Sekans Verisi (Sıralı Veri)
 - Doğal Dil
 - DNA Dizileri

Speech recognition



Sentiment classification

“There is nothing to like
in this movie.”

DNA sequence analysis

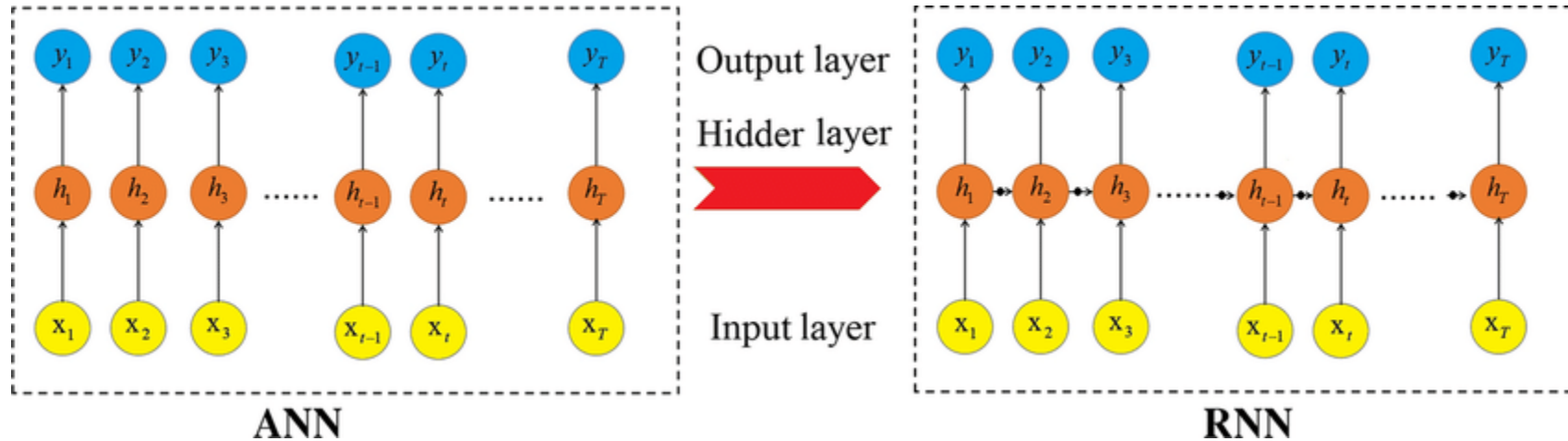
AGCCCCTGTGAGGAACTAG

Recurrent Neural Networks (RNN)

- **Dil ve Zaman Serisi Verilerinde Dizisel Bağımlılık Nedir?**
 - **Dizisel Bağımlılık:** Zaman serisi ve dil verilerinde, her bir öge (kelime veya veri noktası) sırasıyla önceki ögelere bağımlıdır.
 - **Doğal Dil Örneği:** Cümlede bir kelimenin anlamı, önceki kelimelerle ilişkili olabilir. Örneğin, "Ben kahve **içiyorum**." cümlesinde "içiyorum" kelimesinin anlamı, önceki kelimelerden etkilenir.
 - **Zaman Serisi Örneği:** Hava sıcaklığının yarın ne olacağı, bugünkü ve önceki günlerin sıcaklığına bağlı olabilir.

Recurrent Neural Networks (RNN)

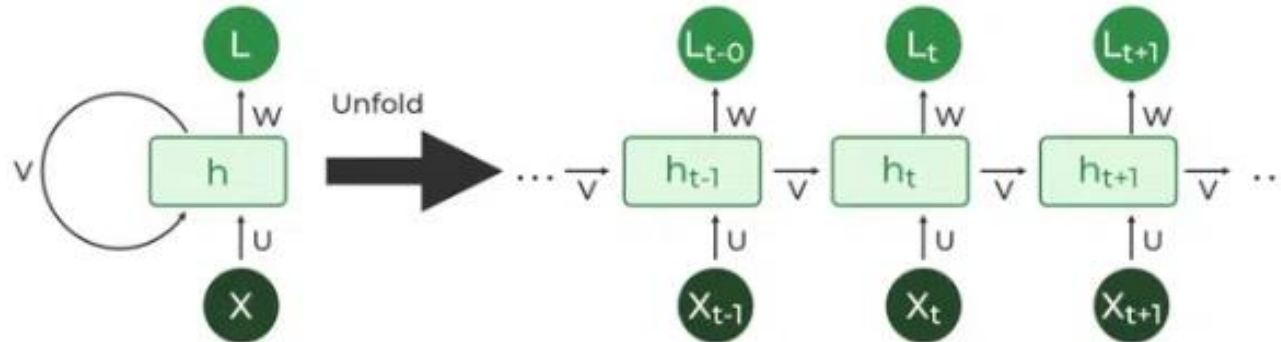
- Standart Sinir Ağları Sekans Verilerinde Neden Yetersizdir?
 - Sabit Girdi/Çıktı
 - Zaman Bağımlılığı
 - Geçmiş Bilgiyi Kaybetme



https://www.researchgate.net/figure/Comparison-between-artificial-neural-network-ANN-and-recurrent-neural-network-RNN_fig5_344946727

Recurrent Neural Networks (RNN)

- RNN'ler, sekans verilerini işlemek için özel olarak tasarlanmış sinir ağlarıdır.
- Her zaman adımında, **önceki zaman adımındaki bilgiyi** saklayarak ve sonraki adımlarla bu bilgiyi güncelleyerek çalışırlar.
- **RNN'in Temel Özellikleri**
 - Zaman Boyutunda Tekrar
 - Sekans Verisi İçin Uygun



RNN Mimarisi Nedir ve Nasıl Çalışır?

- Ağ Yapısı
- Zaman Boyutunda Tekrar Yapısı

Adımlar:

1. **Girdi:** Dizideki bir öge (örneğin, bir kelime).
2. **Gizli Durum:** Önceki adımda üretilen bilgi.
3. **Çıktı:** Girdi ve gizli duruma dayanarak üretilen çıktı.

Gizli Durum Denklemi:

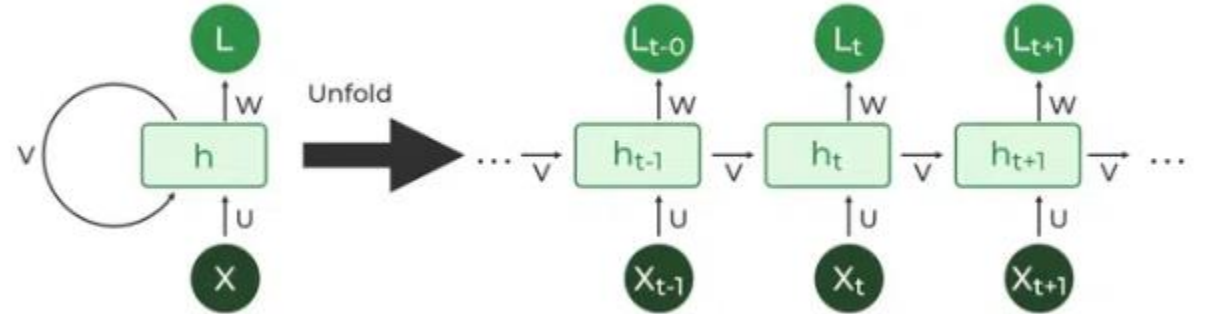
- $h_t = f(W \cdot x_t + U \cdot h_{t-1} + b)$
- $y_t = g(V \cdot h_t)$

h_t : Şu andaki gizli durum

x_t : Şu andaki giriş

h_{t-1} : Önceki gizli durum

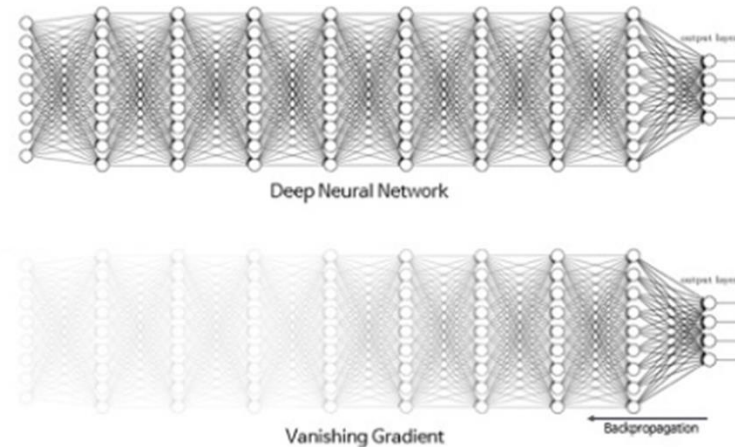
y_t : Çıktı



<https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>

Vanishing Gradient Sorunu Nedir?

- Vanishing Gradient Sorunu, RNN'lerde eğitim sırasında ortaya çıkan bir problemdir.
- Geriye dönük hata yayılımı (backpropagation) sırasında, gradyanlar çok küçük hale gelir ve bu, uzun süreli bağımlılıkların öğrenilmesini zorlaştırır.
- **Neden Oluşur?**
 - Her zaman adımında zincirleme türev alınır.
 - Derin ağlarda bu türevler zaman içinde küçülebilir ve neredeyse sıfıra yaklaşır.
 - Bu durumda, önceki adımlardaki bilginin etkisi kaybolur.
- **Sonuç**
 - Kısa dönem bağımlılıkları öğrenir, ancak uzun dönem bağımlılıkları öğrenmekte zorlanır.



RNN ile NLP Alanında Yapılan Uygulamalar

- Dil Modelleme
- Makine Çevirisi
- Duygu Analizi
- Konuşma Tanıma
- Metin Üretimi



BTK
AKADEMİ

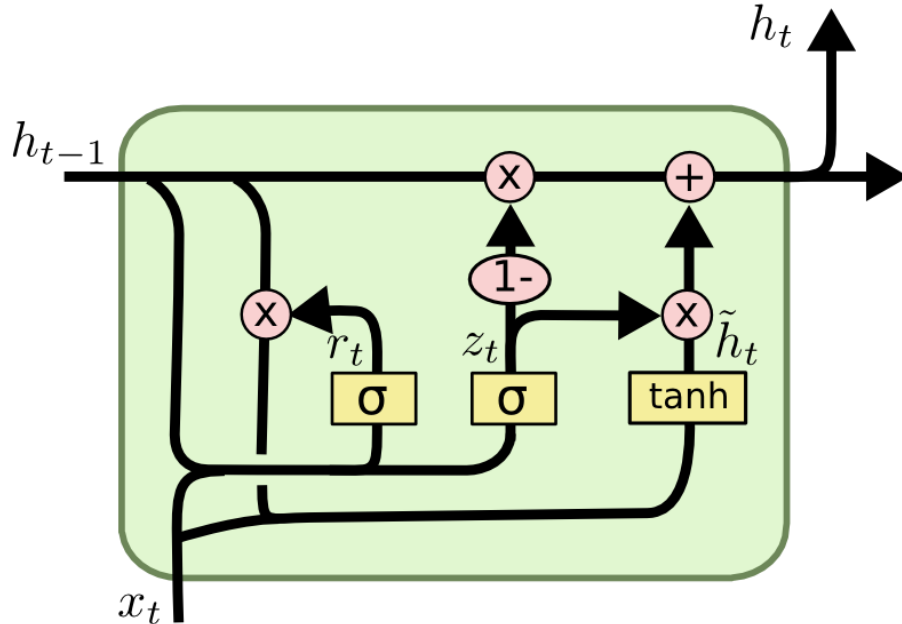
Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM)

- LSTM Nedir?
- LSTM'in RNN Üzerindeki İyileştirmesi
- LSTM Mimarisi, Bileşenleri ve İşleyişi
- LSTM Kullanım Alanları
- RNN vs LSTM

Long Short-Term Memory (LSTM)

- **Long Short-Term Memory (LSTM)**, Recurrent Neural Network (RNN) türlerinden biridir.
- Uzun vadeli bağımlılıkları öğrenebilmek için özel olarak tasarlanmıştır.
- **LSTM'nin Ana Amacı**
 - Zaman bağımlı verilerde uzun dönem bağımlılıkları öğrenmek.
 - Geleneksel RNN'lerin yaşadığı **vanishing gradient** sorununu çözmek için geliştirilmiştir.



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

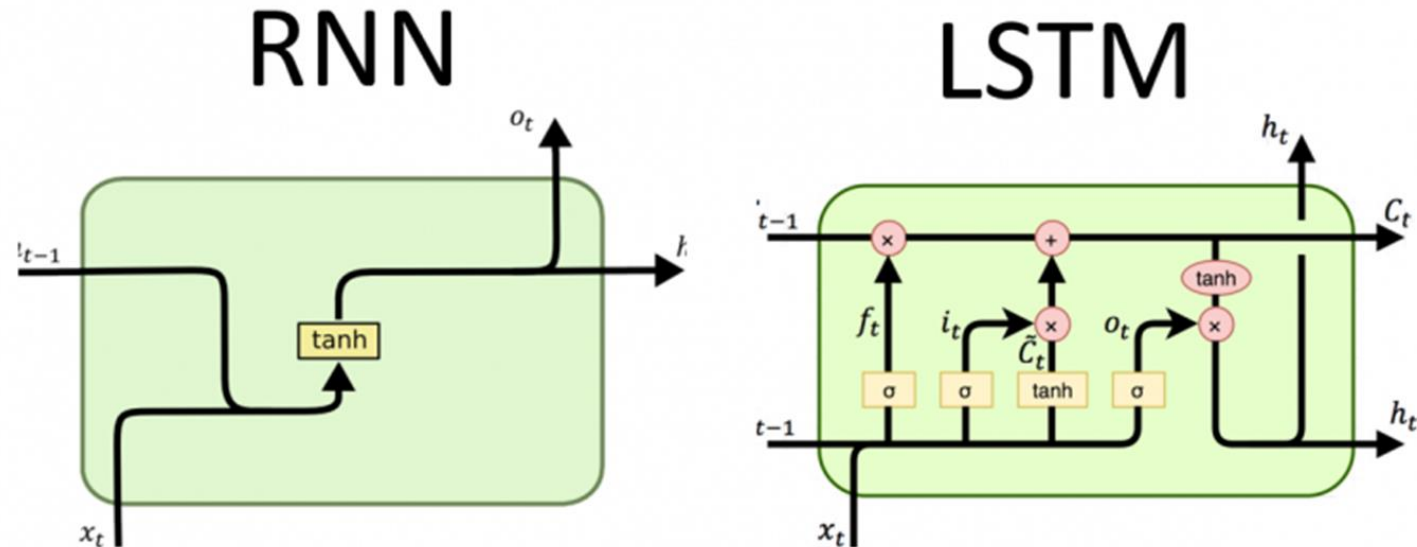
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

LSTM'in RNN Üzerindeki İyileştirmesi Nedir?

- RNN'lerde Yaşanan Sorunlar
 - Vanishing Gradient Sorunu
 - Uzun Vadeli Bağımlılıkları Öğrenememe
- LSTM'nin Çözümü
 - Hücre Durumu (Cell State)
 - Kapılar (Gates)



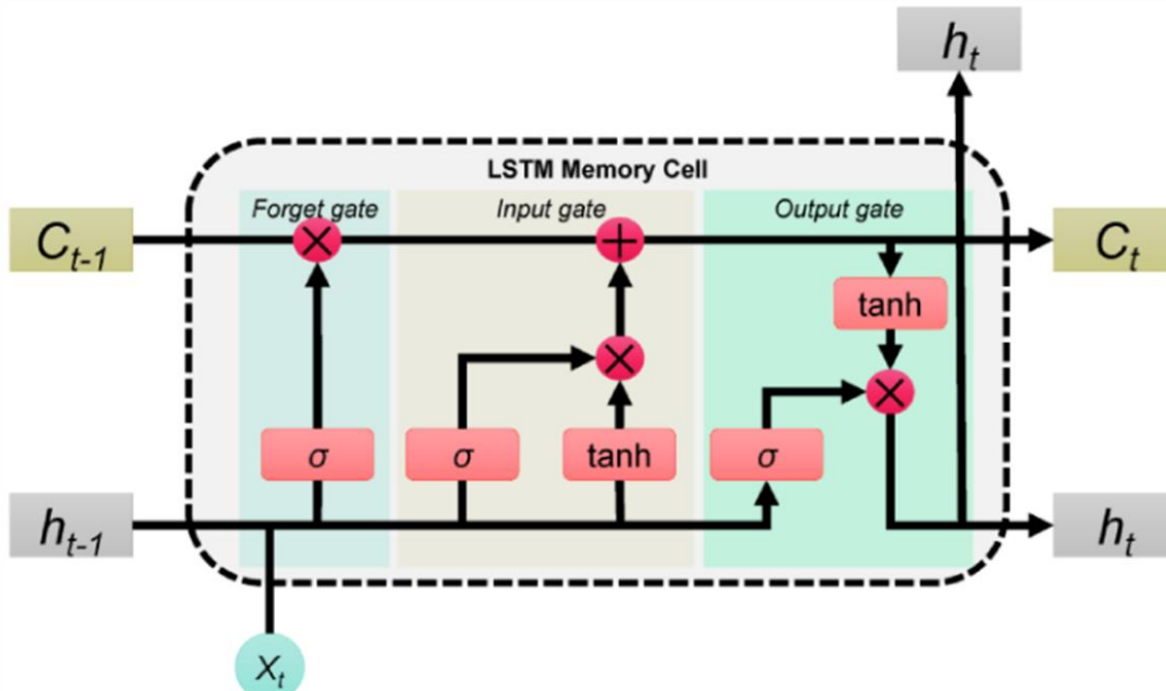
LSTM Mimarisi, Bileşenleri ve İşleyişi Nedir?

- LSTM'in Bileşenleri

- Hücre Durumu (Cell State)
- Giriş Kapısı (Input Gate)
- Unutma Kapısı (Forget Gate)
- Çıkış Kapısı (Output Gate)

- LSTM'in İşleyişi

- Unutma Kapısı
- Giriş Kapısı
- Hücre Durumunun Güncellenmesi
- Çıkış Kapısı



LSTM Formülleri:

- Forget Gate (Unutma Kapısı): $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- Input Gate (Giriş Kapısı): $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
- Output Gate (Çıkış Kapısı): $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$
- Cell State Update (Hücre Durumu Güncellemesi): $C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$

LSTM Kullanım Alanları

- Doğal Dil İşleme (NLP)
- Konuşma Tanıma
- Zaman Serisi Tahmini
- Müzik ve Metin Üretimi
- Video İşleme

RNN vs LSTM

Özellik	RNN	LSTM
Uzun Dönem Bağımlılık	Uzun vadeli bağımlılıkları öğrenmede zayıf	Uzun vadeli bağımlılıkları etkili bir şekilde öğrenir
Vanishing Gradient	Vanishing Gradient sorunu yaşar	Bu sorunu çözer, uzun vadeli bağımlılıkları tutar
Kapılar (Gates)	Kapılar yok, yalnızca basit bir yapı	Giriş, unutma ve çıkış kapıları ile bilgi akışını kontrol eder
Hafıza Yönetimi	Hafızayı verimli bir şekilde yönetemez	Hücre durumu sayesinde hafızayı verimli bir şekilde yönetir
Kullanım Alanları	Kısa sekanslarda veya küçük veri setlerinde daha hızlı olabilir	Uzun dizilerde daha etkili ve geniş çapta kullanılır



BTK
AKADEMİ

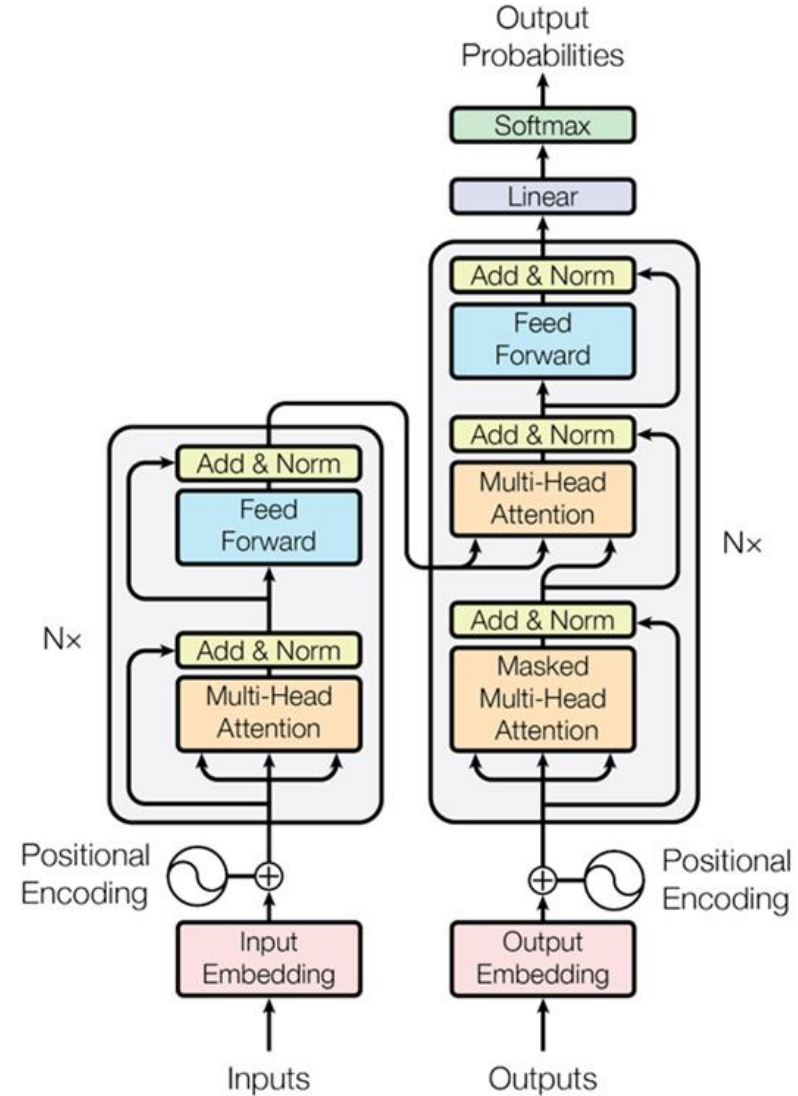
Transformers

Transformers

- Transformers Nedir?
- BERT
- GPT
- LLAMA
- Transformers Modelleri Karşılaştırma

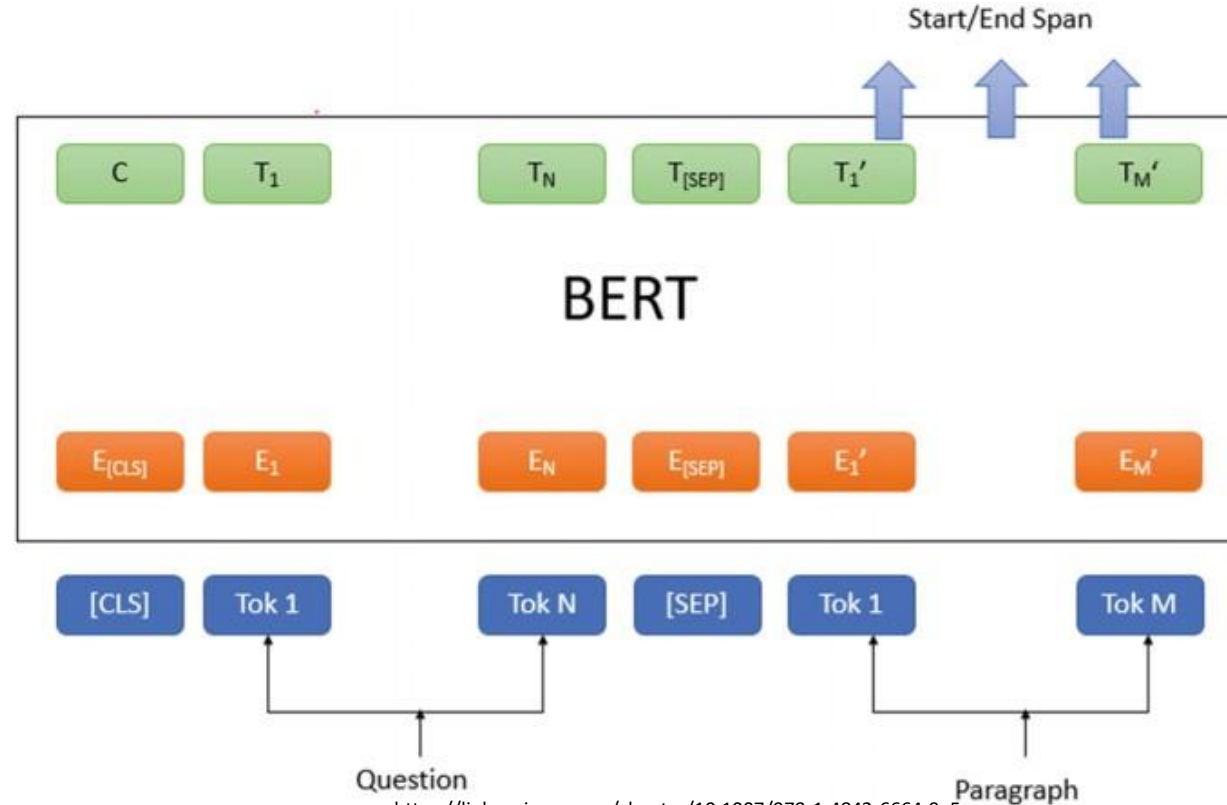
Transformers

- NLP devrimi
- Attention is All You Need
- Bağlamı Daha İyi Anlama
- Paralel İşleme Yeteneği
- Çeşitli NLP Görevlerinde Kullanım
- Önceden Eğitilmiş Modellerin Yeniden Kullanımı



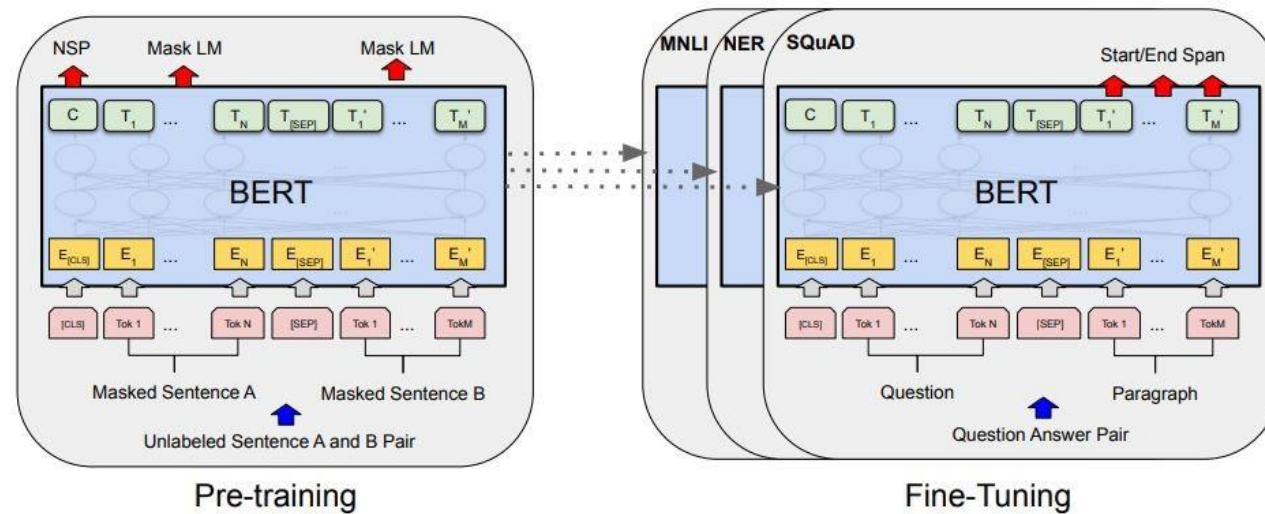
Transformers Modelleri: BERT

- BERT, dil anlama ve metin işleme görevleri için kullanılan bir dil modelidir.
- BERT, metni hem soldan sağa hem de sağdan sola okuyarak bağlamı anlamaya çalışır. (hem geçmişe hem de geleceğe bakar)
- Metni anlamak için kullanılır, örneğin, soru-cevaplama ve metin sınıflandırma.



Transformers Modelleri: BERT Mimarisi

- **Transformer Encoder**
 - BERT, transformer mimarisinin sadece **encoder** kısmını kullanır.
 - Transformer, dikkat (attention) mekanizmasına dayanan bir modeldir.
- **İki Aşamalı Eğitim**
 - **Ön Eğitim (Pre-training)**
 - Masked Language Modeling (MLM)
 - Next Sentence Prediction (NSP)
 - **İnce Ayar (Fine-tuning)**

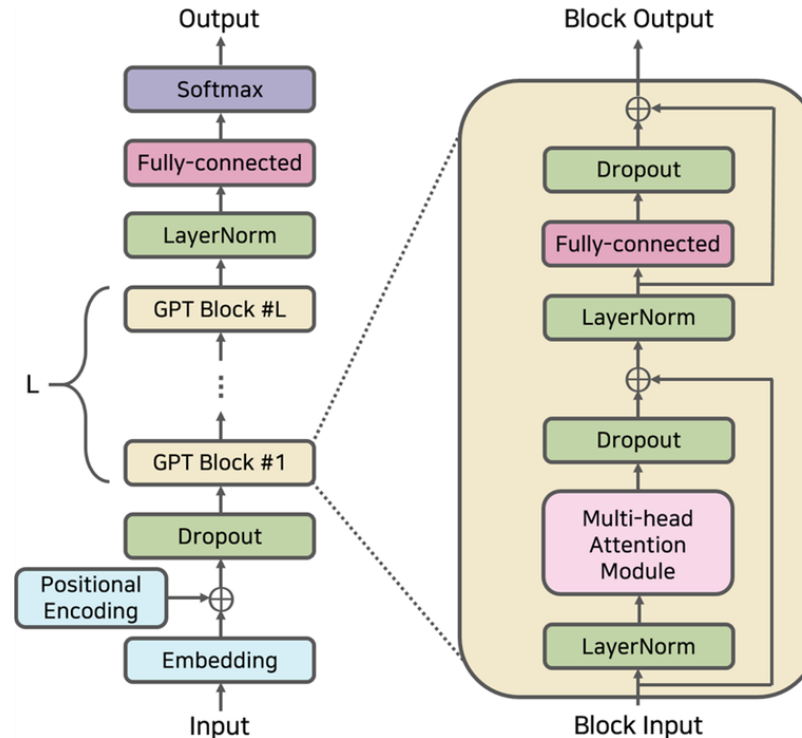


Transformers Modelleri: BERT Özellikleri

- Çift Yönlü (Bidirectional)
- Transfer Öğrenme
- Transformer Encoder Kullanımı

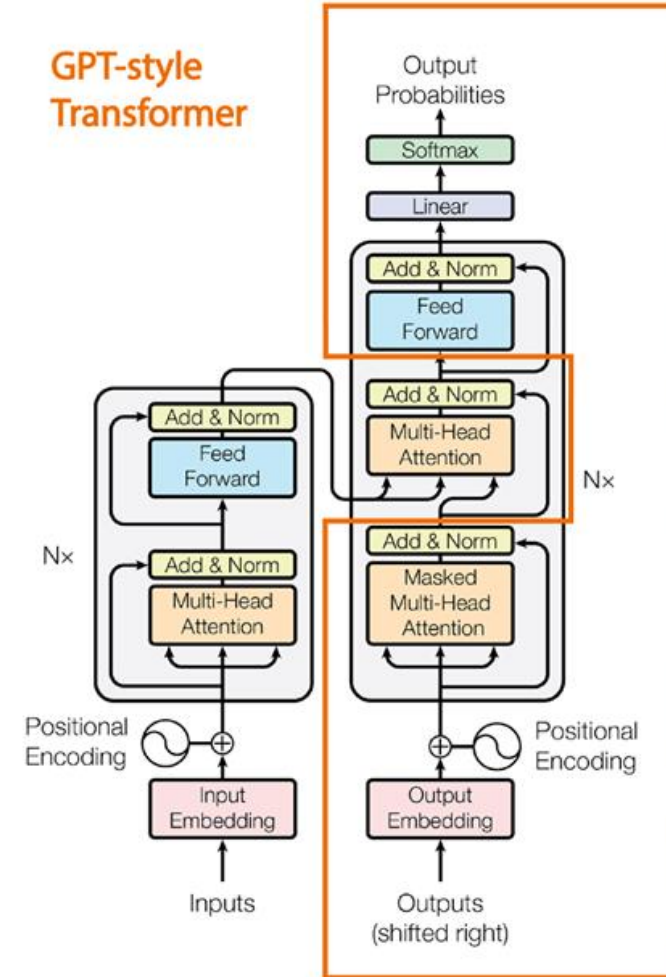
Transformers Modelleri: GPT

- GPT, metin üretme ve dil modelleme için kullanılan bir dil modelidir.
- GPT, metni sadece soldan sağa doğru okur ve bir kelimeyi tahmin etmek için önceki kelimelere dayanır.
- Metin üretme, öykü yazma, yaratıcı içerik oluşturma gibi görevlerde kullanılır.



Transformers Modelleri: GPT Mimarisi

- Transformer Decoder
- Otokorelasyonlu (Autoregressive) Yaklaşım
- Tek Aşamalı Eğitim



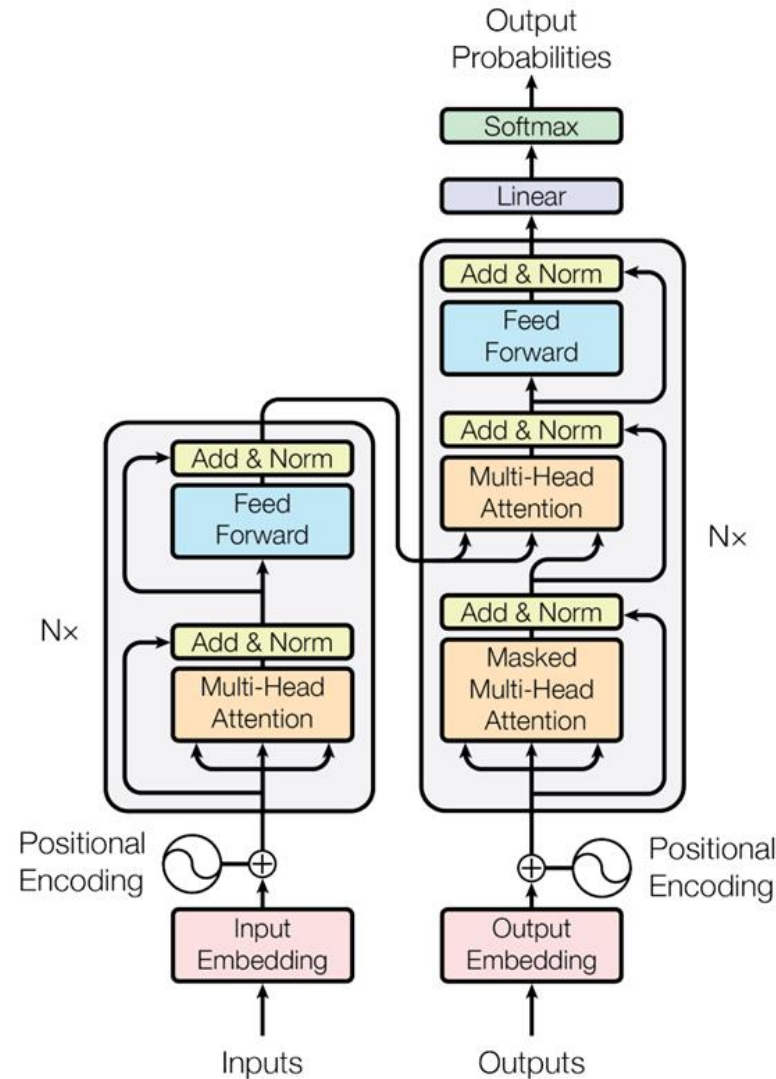
Transformers Modelleri: GPT Özellikleri

- Tek Yönlü (Unidirectional)
- Metin Üretimi
- Transfer Öğrenme
- Çok Büyük Modeller

BERT vs GPT

BERT

Encoder



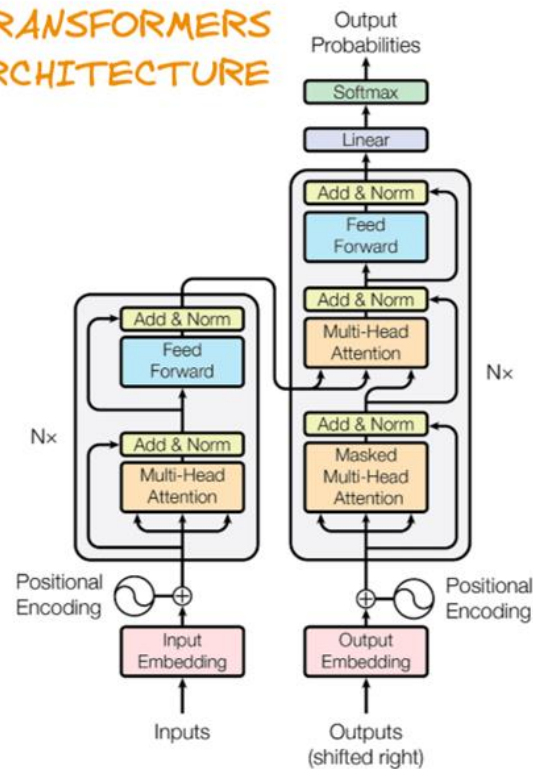
GPT

Decoder

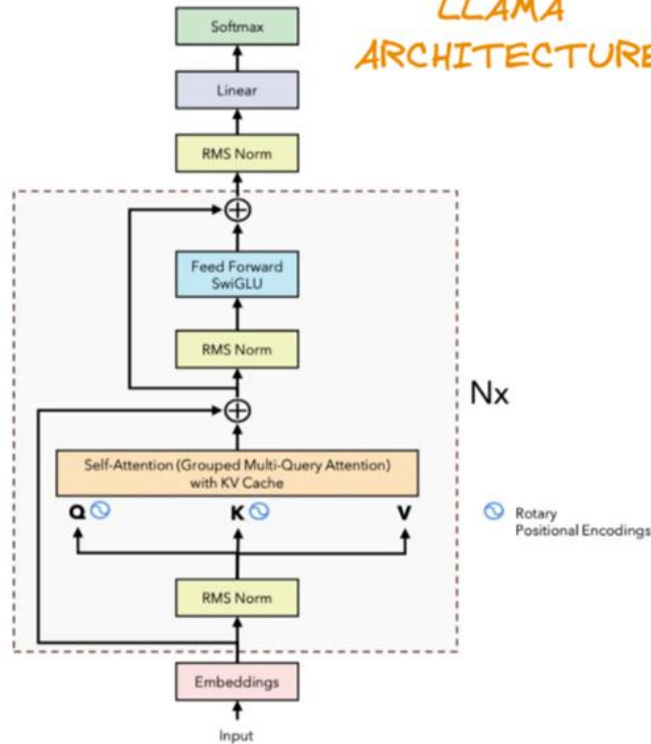
Transformers Modelleri: LLAMA

- LLaMA, Meta (eski adıyla Facebook) tarafından geliştirilen büyük dil modelleri ailesidir.
- LLaMA, büyük veri kümeleri üzerinde eğitilmiş olup, GPT-3 gibi modellerle kıyaslandığında daha az parametre ile benzer performans sağlamayı hedefler.

TRANSFORMERS ARCHITECTURE

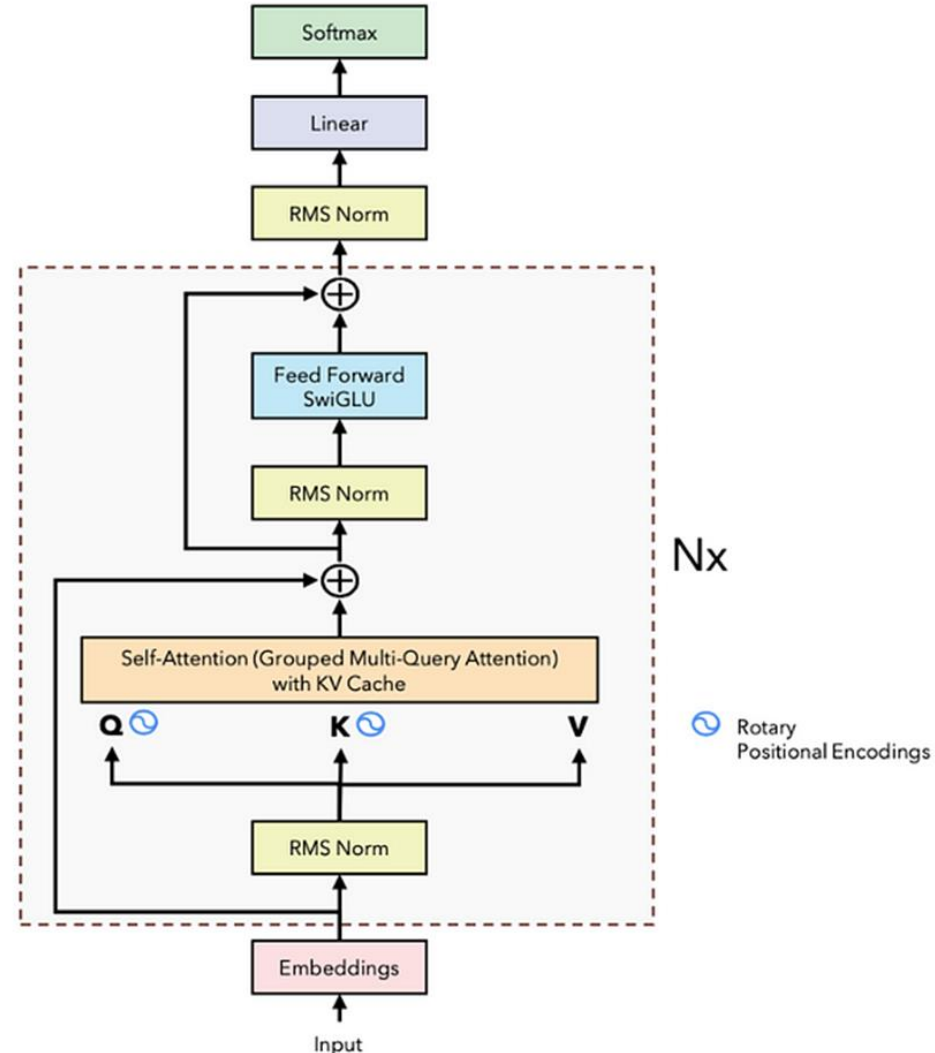


LLAMA ARCHITECTURE



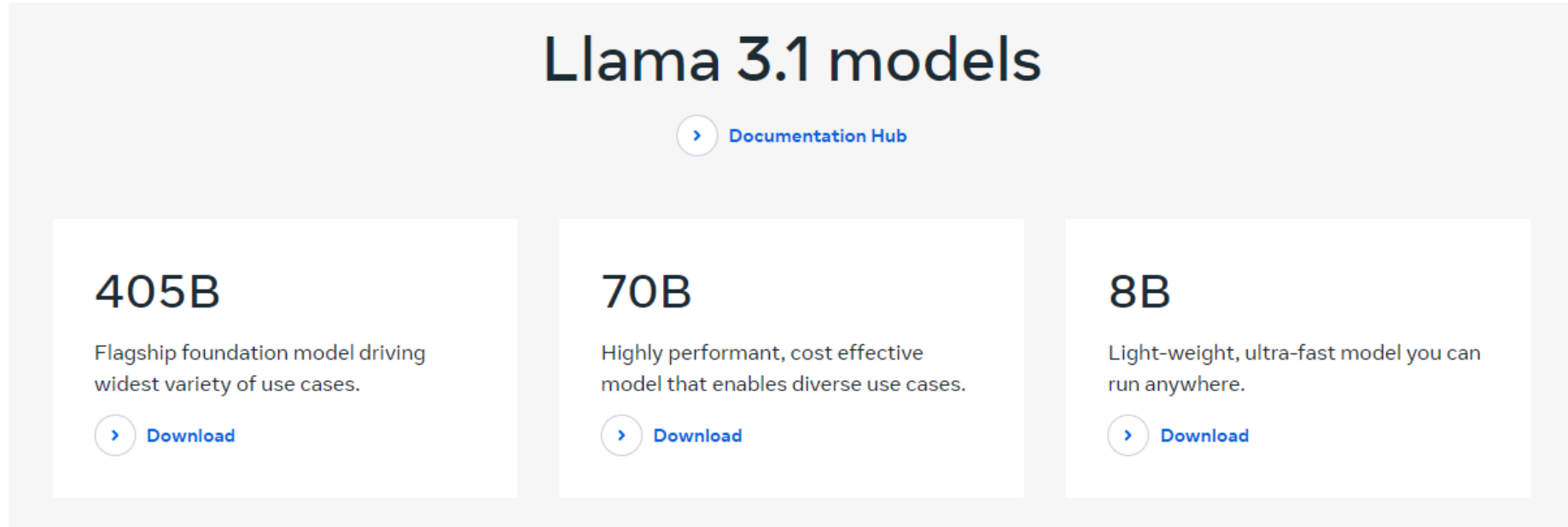
Transformers Modelleri: LLAMA Mimarisi

- Transformer Tabanlı
- Hafif ve Verimli
- Daha Küçük ve Daha Hızlı



Transformers Modelleri: LLAMA Özellikleri

- Verimli Eğitim
- Araştırmacı Odaklı
- Model Boyutları



Llama 3.1 models

[Documentation Hub](#)

<h3>405B</h3> <p>Flagship foundation model driving widest variety of use cases.</p> <p>Download</p>	<h3>70B</h3> <p>Highly performant, cost effective model that enables diverse use cases.</p> <p>Download</p>	<h3>8B</h3> <p>Light-weight, ultra-fast model you can run anywhere.</p> <p>Download</p>
---	---	---

Transformers Modelleri Karşılaştırma

Özellik	BERT	GPT	LLaMA
Eğitim Yönü	Çift yönlü (bidirectional)	Tek yönlü (unidirectional)	Çift yönlü
Kullanılan Transformer	Encoder	Decoder	Encoder + Decoder
Ana Görev	Metin anlama ve sınıflandırma	Metin üretimi ve dil modelleme	Hem metin üretimi hem metin anlama
Eğitim Görevleri	Masked Language Modeling, NSP	Language Modeling	Language Modeling
Kullanım Alanları	Soru-cevap, duygu analizi, NER	Metin üretimi, hikaye yazma, sohbet	NLP araştırmaları, düşük kaynaklı cihazlarda kullanım
Öne Çıkan Özellik	Çift yönlü bağlam öğrenme	Büyük metin üretimi, otokorelasyonlu	Verimlilik ve parametre açısından optimize edilmiş