

Hesaplamalı Anlambilim Ders Notları

# Büyük Dil Modelleri

Mehmet Fatih AMASYALI  
Hesaplamalı Anlambilim  
Ders Notları  
BÖLÜM 8



Yıldız Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

## İçerik

- Otoregresif dil modelleri
- Uyarlanmış büyük dil modelleri
- Yöntemler
- Tartışma



NOVA Research Lab

## Dil modellerini nasıl eğitelim?

- Büyük dil modelleri (LLM) token tahmini ile eğitilirler (next, masked, changed). Artık oto-regresif modeller (next thing prediction) standart\*. Çünkü yeni görevlere çok daha kolay uyum sağlıyor.
- LLM'ler eğitim kümesinde 0 hataya erişebilirler mi? Düşük hataya erişebilmeleri için benzer anlama/kullanıma sahip kelimelerin temsilleri ?



[\*] <https://arxiv.org/abs/2204.02311>

NOVA Research Lab

## Amaçlanmadan ortaya çıkan kabiliyetler Emergent properties

- Autoregressive LLM's are ??? [\*]
- Yeni bir görev için modelde güncelleme / finetune yapmadan sonuç alabilmek

Language Models are Few-Shot Learners

Tom R. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Snibb*
Jared Kaplan*	Pradyumn Agarwal	Arvind Nishankumar	Prafulla Shyam
Amanda Askell	Saarthak Agarwal	Arif Herbert-Voss	Gretchen Krueger
Rewon Child	Affra Ranesh	David M. Eagle	Jeffrey Wu
Christopher Hesse	Mark Chen	Eric Sigler	Matteo Litalini
Benjamin Chess	Jack Clark	Christopher Berner	Scott Gray
Sam McCandlish	Alec Radford	Rya Sankar	Dario Amodei
OpenAI			



[\*] <https://arxiv.org/pdf/2005.14165.pdf>

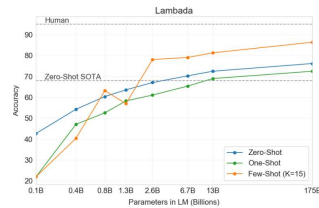
NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

Traditional fine-tuning (not used for GPT-3)

## Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.



Lambada görevi: Cümlelerin son kelimesini tahmin,  
uzun bağımlılıkları bulma kabiliyeti

[\*] <https://arxiv.org/pdf/2005.14165.pdf>

## X-shot

The three settings we explore for in-context learning

## Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



## One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.



## Few-shot

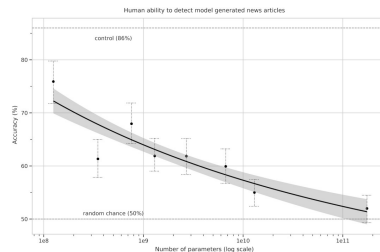
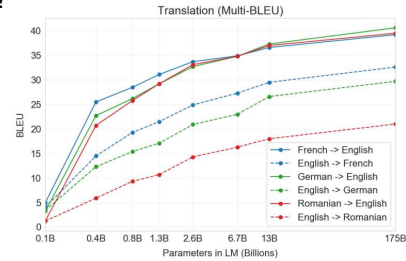
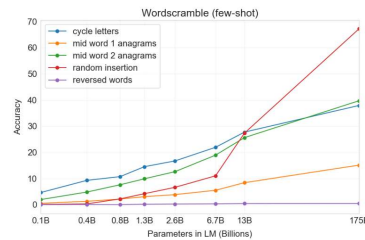
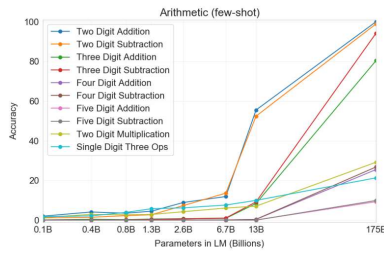
In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

Başka ?

[\*] <https://arxiv.org/pdf/2005.14165.pdf>

NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

Çoktan seçmeli sorular içeren veri kümelerinde X-shot:  
Her cevabın üretilme olasılığı bulunup, büyük olan seçiliyor

CommonsenseQA	Minimal	What blocks sunshine? \n {summer park desktop sea moon}
	Manual	The question is: What blocks sunshine? \n The answer is: {summer park desktop sea moon}
COPA	Minimal	Effect: I coughed. \n {Cause: I inhaled smoke. Cause: I lowered my voice.}
	Manual	I coughed because {I inhaled smoke. I lowered my voice.}
ARC	Minimal	Which biome has the most vegetation? \n {desert forest grassland tundra}
	Manual	The question is: Which biome has the most vegetation? \n The answer is: {desert forest grassland tundra}



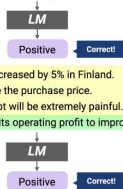
[\*] <https://arxiv.org/pdf/2202.12837.pdf>

NOVA Research Lab

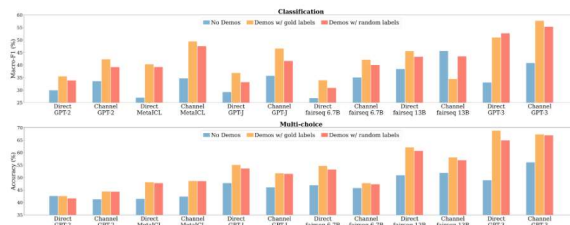
## Hesaplamalı Anlambilim Ders Notları

## Cevaplar yanlış olsa da\*

Circulation revenue has increased by 5% in Finland. \n Positive  
Panostaja did not disclose the purchase price. \n Neutral  
Paying off the national debt will be extremely painful. \n Negative  
The company anticipated its operating profit to improve. \n



Circulation revenue has increased by 5% in Finland. \n Neutral  
Panostaja did not disclose the purchase price. \n Negative  
Paying off the national debt will be extremely painful. \n Positive  
The company anticipated its operating profit to improve. \n



## Hiç görmediği kelimeler olsa da\*\*

A "Burringgo" is a car with very fast acceleration. An example of a sentence that uses the word Burringgo is:  
In our garage we have a Burringgo that my father drives to work every day.

A "Gigamuru" is a type of Japanese musical instrument. An example of a sentence that uses the word Gigamuru is:  
I have a Gigamuru that my uncle gave me as a gift. I love to play it at home.

To "screeg" something is to swing a sword at it. An example of a sentence that uses the word screeg is:  
We screeged at each other for several minutes and then we went outside and ate ice cream.



[\*] <https://arxiv.org/pdf/2202.12837.pdf>

[\*\*] <https://arxiv.org/pdf/2005.14165.pdf>

NOVA Research Lab

## Bu nasıl mümkün oluyor?

- Net bir açıklama yok şimdilik
- <https://ai.stanford.edu/blog/understanding-incontext/>



## Openai api\*

```
import os
import openai
openai.api_key = os.getenv("OPENAI_API_KEY")
response = openai.Completion.create(
    model="text-davinci-003",
    prompt="The following is a list of companies and th
they fall into:\n\nApple, Facebook, Fedex\n\nApple\nCatego
temperature=0,
max_tokens=64,
top_p=1.0,
frequency_penalty=0.0,
presence_penalty=0.0 )
```

### Prompt

The following is a list of  
companies and the categories  
they fall into:  
Apple, Facebook, Fedex  
Apple  
Category:

### Sample response

Technology  
  
Facebook  
Category: Social Media  
  
Fedex  
Category: Delivery



[\*] <https://platform.openai.com/examples/>

## Hesaplamalı Anlambilim Ders Notları

```
response = openai.Completion.create( model="text-davinci-003", prompt= .....
```

## Prompt

```
def foo(n, k):
    accum = 0
    for i in range(n):
        for j in range(k):
            accum += i
    return accum
"""
The time complexity of this function is
```

## Sample response

$O(n \cdot k)$ . This is because the function has two nested for loops, each of which has a time complexity of  $O(n)$  and  $O(k)$  respectively. Therefore, the total time complexity is the product of the two, which is  $O(n \cdot k)$ .

## Prompt

Decide whether a Tweet's sentiment is positive, neutral, or negative.

Tweet: "I loved the new Batman movie!"  
Sentiment:

## Sample response

Positive

## Prompt

You: What have you been up to?  
Friend: Watching old movies.  
You: Did you watch anything interesting?  
Friend:

## Sample response

Yeah, I watched an old classic called Casablanca. It was really good!



[\*] <https://platform.openai.com/examples/>

NOVA Research Lab

## Hesaplamalı Anlambilim Ders Notları

- GPT-3 could also in principle be evaluated in the traditional fine-tuning setting, but we leave this to **future work** \*
- **Bunu daha genel bir yapıda (task agnostic – görev bağımsız) yaparsak neler olur?**
- İstedığımızı, metinle ifade edebilmek sınırsız bir ifade gücü veriyor bize, tabi model yeterli ise.
- Self supervised → RLHF



[\*] <https://arxiv.org/pdf/2005.14165.pdf>

NOVA Research Lab

- Alignment: Modelin istediğimizi, **bizim!** kabul edebileceğimiz bir yolla yapması
- İstedığımızı (tüm dünya bilgisine sahip bir insan gibi cevap ver, ama **zararlı!** olabilecek çıktılar üretme) objective fonksiyon olarak yazmak zor ☹, ama örneklerini verebiliriz.
- Biz kimiz? Kime göre zararlı?



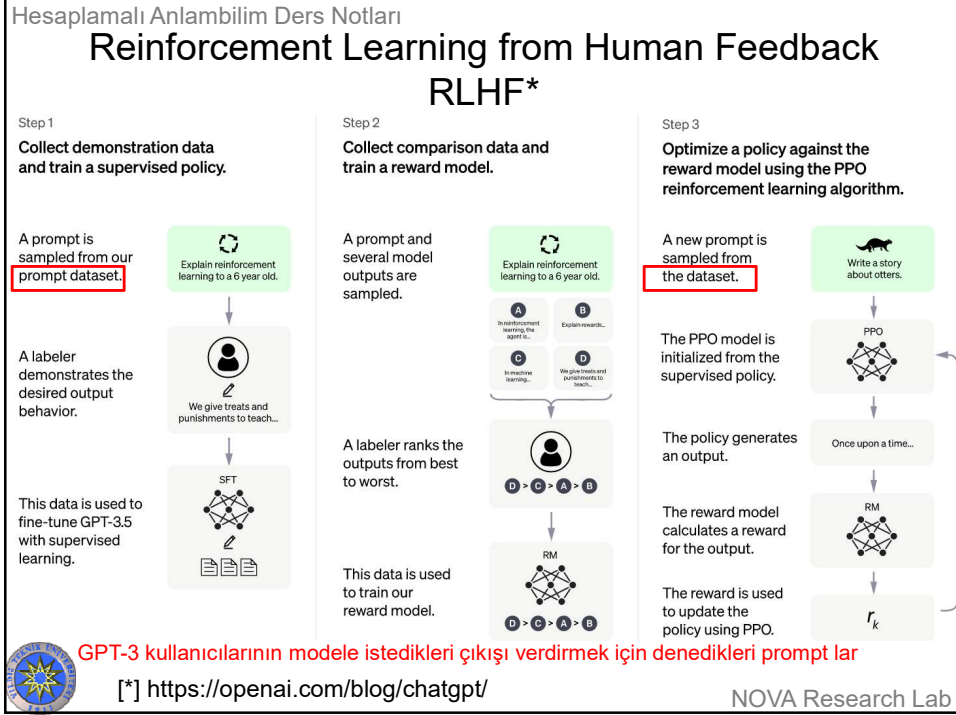
## instructGPT\* - RLHF

1. Model 1: input – output eğitici eğitim (supervised GPT3 finetuning)
2. Veri kümesi A: input – Model 1'in farklı outputları arasında insanların tercihleri (comparison data)
3. Model 2 (Reward Model) : Veri kümesi A üzerinde eğitim (insanların verdikleri bilgilerle, insanların ödül verme mekanizmasını modelleyen)
4. Model 1'i Model 2'nin ödül tahminlerinin kullanarak RL ile eğitim, 2. adıma dön

1. ve 3. adımlarda insan devrede. İlkinde cevap yazıyor, sonrakinde seçiyor. Toplamda 20 bin saat insanlar etkileşime girmiş.



[\*] <https://arxiv.org/abs/2203.02155>



Hesaplamalı Anlambilim Ders Notları

## RL

- Belli bir durumda ne yapması gerektiğini değil, yaptığıının ne kadar iyi/kötü olduğunu geri bildirim olarak vermek
- Neden?
  - Biz de doğrudan çok emin olmayabiliriz
  - Tek doğrunun olmayışı
- Expedition vs. Exploration ikilemi: Yemeği nerede yiyelim? Bildiğimiz en iyisinde mi? Yeni bir yerde mi?

NOVA Research Lab



## On / off policy / offline RL\*

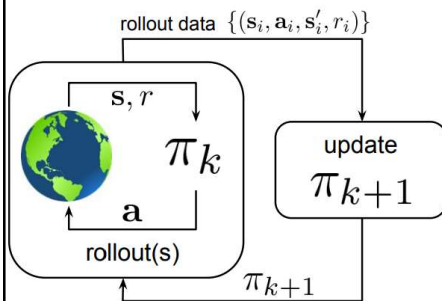
- Ajan, durumu ( $s_t$ ) gözler, buna göre policy  $\pi_\theta(\mathbf{a}|\mathbf{s})$  i kullanıp hareket ( $a$ ) seçer. Bunun sonucunda ödül ( $r$ ) alır ve sonraki duruma ( $s_{t+1}$ ) geçer.
- Policy: durumu harekete dönüştüren bir fonksiyon (amaç bunu optimize etmek)
- Toplanan deneyimlerin formatı :  $\langle \mathbf{s}, \mathbf{a}, \mathbf{s}', \mathbf{r} \rangle$
- Bu deneyimlerle policy eğitilir.
- Yöntemlerin farkı deneyimlerin üretim süreçleri



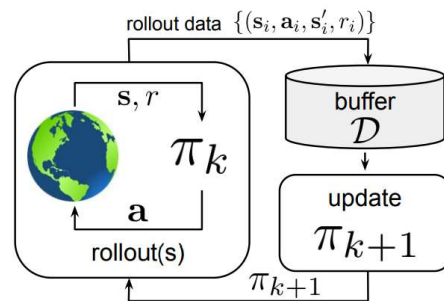
[\*] <https://arxiv.org/pdf/2005.01643.pdf>

NOVA Research Lab

(a) online reinforcement learning

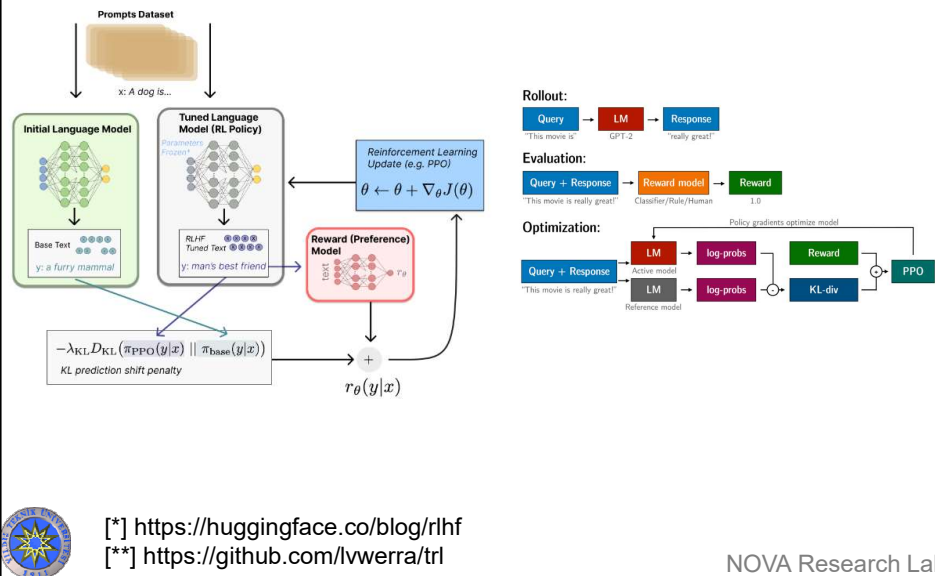


(b) off-policy reinforcement learning



NOVA Research Lab

## Dil modellerini RL ile finetune `[*, **]`



## Bazı Veri kümeleri

- Reddit: Explain I am 5
  - <https://www.reddit.com/r/explainlikeimfive/top/>
- Big-Bench
  - <https://github.com/google/BIG-bench>
- FLAN
  - <https://arxiv.org/abs/2301.13688>
- Collective human preferences

## Topluluğun tercihleri Collective human preferences

- İnsan geri bildirimi very kümelerini ucuza elde etmek:
- Forumlardan hazır veri toplamak
- Bir soruya verilmiş birçok cevap içinde hangisinin daha iyi olduğunu belirlemek
- Nasıl?
  - Like, vote sayılarına göre tercihleri otomatik toplamak (soru, cevap1, cevap2, hangisi iyi)
  - Stanford human preferences dataset\*: iyilik için zamanı da kullanmak (daha çok oy alan ve eski olan daha iyi, çünkü yeni henüz yeterince görülmemiş olabilir)

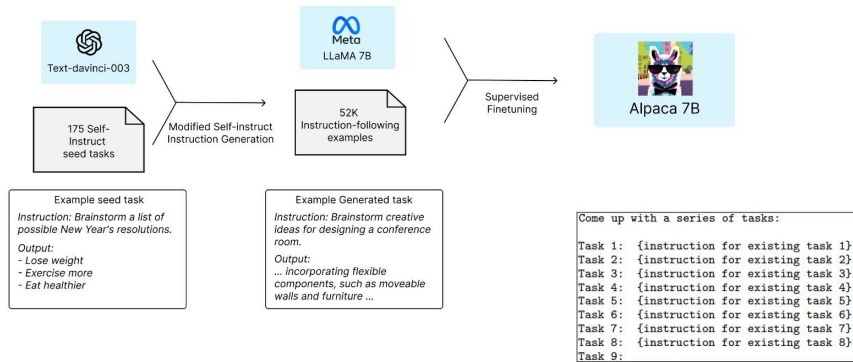


[\*] <https://huggingface.co/datasets/stanfordnlp/SHP>

NOVA Research Lab

## Büyük Dil Modellerini Taklit (Alpaca\*)

- GPT3.5 (>175B) yi kullanıp, self instruction\*\* 'la LLaMA\*\*\* (7B) yı eğitmek, <600\$ (ek insan eforu yok)



[\*] <https://crfm.stanford.edu/2023/03/13/alpaca.html>

[\*\*] <https://arxiv.org/abs/2212.10560>

[\*\*\*] <https://arxiv.org/abs/2302.13971v1>

GPT'den yeni görev üretim süreci\*\*

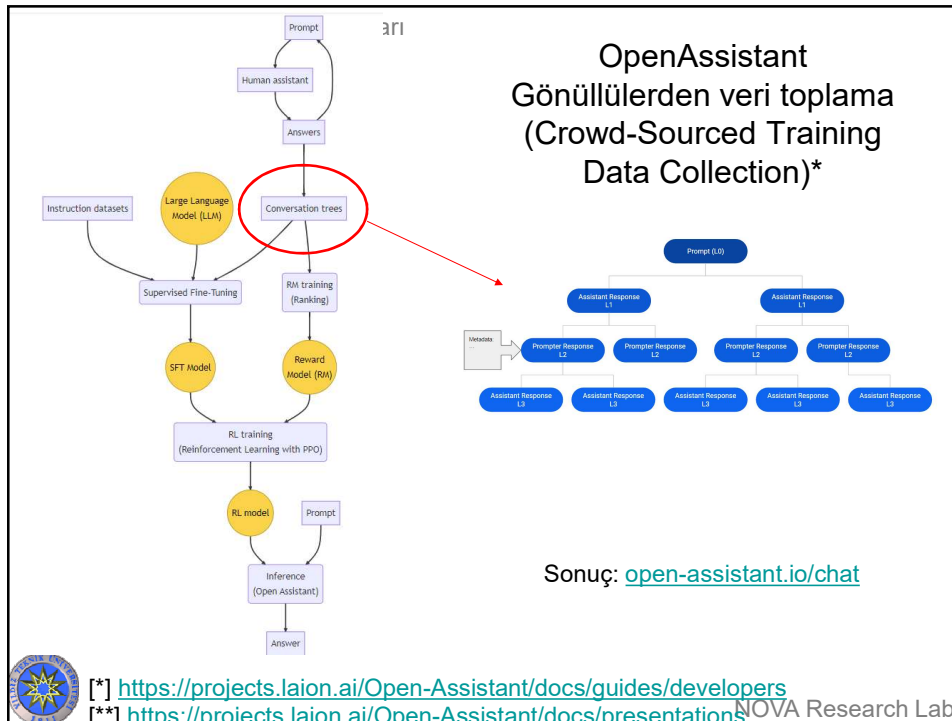
NOVA Research Lab

# chatGPT\*

- Ticari, yöntemi çok bilinmiyor. Bildiklerimiz:
  - Instruct GPT'nin bir versiyonu
  - Diyalog tabanlı (yeni cevap, önceki soru ve cevaplara da bağlı)
  - Daha büyük model, daha çok veri ile eğitilmiş
- Açık kaynak yeniden üretimi:  
OpenAssistant\*

[\*] <https://chat.openai.com/>[\*\*] [github.com/LAION-AI/Open-Assistant](https://github.com/LAION-AI/Open-Assistant)

NOVA Research Lab

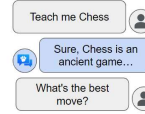


## Nasıl veri toplayalım? Görevler:

### Main Task 1: Reply to conversation



### Main Task 2: Label a text



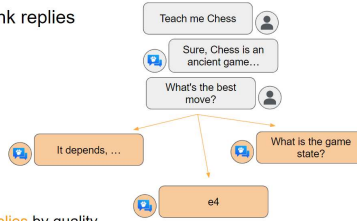
Does **this text** contain:

- Profanity?
- Sarcasm?
- Violence?

Is **this text**:

- Helpful?
- Too short?
- Too vague?

### Main Task 3: Rank replies



## Henüz

- Plan yapma ☹ [\*]
- Tüm işlem tek bir çizgi üzerinde, her şey aynı zorluk seviyesinde
- Nasıl aşılabılır?
  - <https://openreview.net/pdf?id=BZ5a1r-kVsf>
- Aşmalı mıyız?
- Aşılabiliyorsa durdurabilir miyiz?



[\*] Sparks of Artificial General Intelligence: Early experiments with GPT-4

<https://arxiv.org/pdf/2303.12712.pdf>

video: <https://www.youtube.com/watch?v=qblk7-JPB2c>

## Tartışma - 2023

- Hinton: İnsan beyni kadar iyi çalışan sistemler yapmaya çalıştık. Ama şimdi 1T parametrelili bu modeller bizim 100T parametrelili beynimize çok yakın işler yapabiliyorlar. Yani bilgiyi bizden daha iyi temsil edebiliyorlar şu anda. Yakında ...
- Plan yapamıyorlarsa, nasıl kod yazıyorlar? Yepyeni problemleri, çözümünü olanlar kadar iyi çözemiyorlar. (Bu high-tech plagiarism [Chomsky]). Aynı şeyler IBM'in Jeopardy sistemi için de söylenmişti. Abartmayalım.
- Soru: Zorluk seviyesi çok daha fazla olan şeyleri yapabiliyorken daha az olanları yapamamanın, kolayca kandırılabilmesinin açıklaması nedir?
  - High-tech plagiarism, Süper kopyacılar ☺
  - Bizler gibi hiyerarşik kabiliyet kümeleri yok. Başka türlü bir öğrenme bu. Sistematik, düzenli, hiyerarşik değil.
  - Başka ??
- İnsanları tahmin edemiyorsan, insanları tahmin edilebilir hale getir ☺  
Bu tür araçlar insanların tahmin edilebilirliğini arttırır mı?

