

# ChatGPT 를 활용한 의학논문 작성법

Jy-yong Sohn (손지용)

연세대학교 응용통계학과 조교수



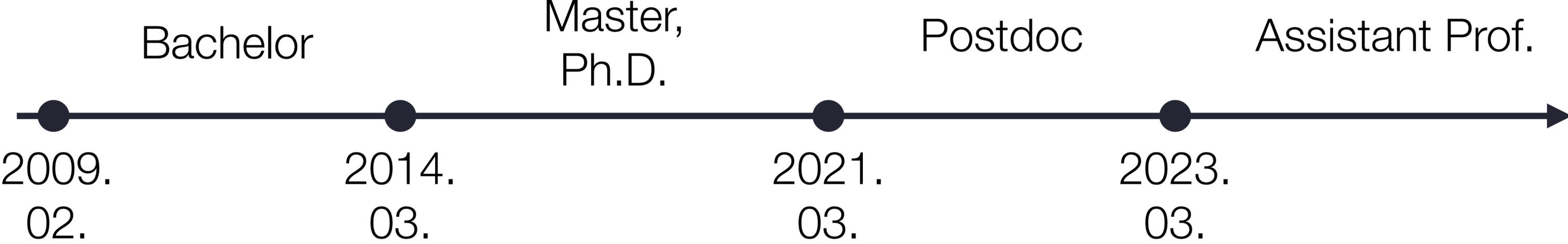
# Basic Terminologies

- Artificial Intelligence (AI): 인공지능
- Machine Learning (ML): 기계학습
- Natural Language Processing (NLP): 자연어처리 분야
  - Large Language Models (LLMs): 거대언어모델
    - Generative Pre-trained Transformer (GPT): OpenAI가 만든 LLM 시리즈
    - ChatGPT: GPT version 3.5부터 제공된 chatting 형태의 서비스

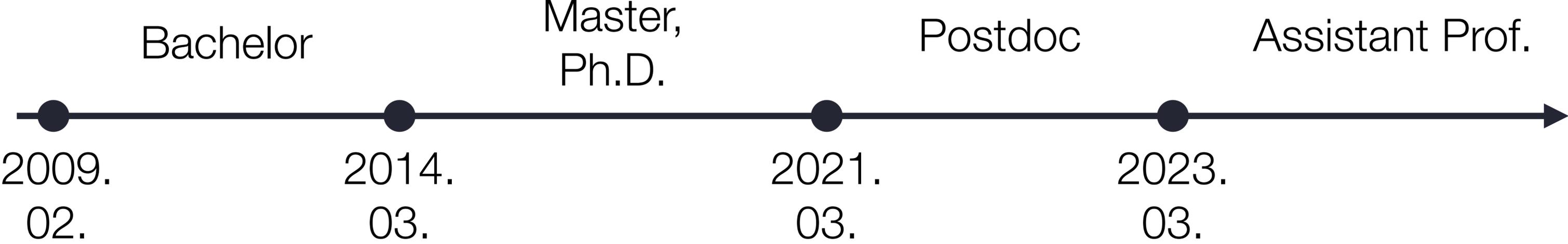
# Table of Contents

- A Brief Introduction
- Large Language Models (LLMs) Basics
- LLMs for Medical Domain

# A Brief Introduction

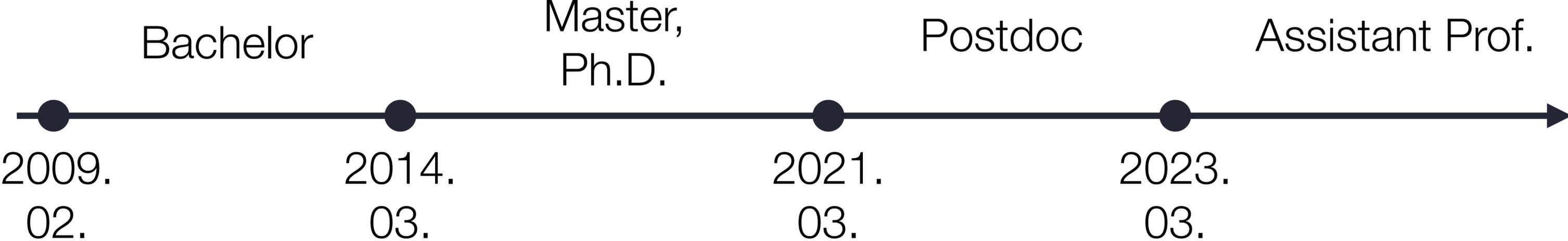


# A Brief Introduction

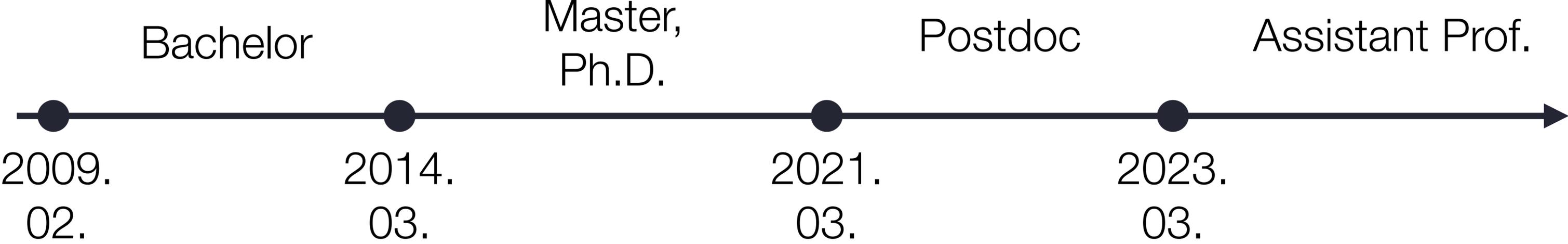


**KRAFTON**

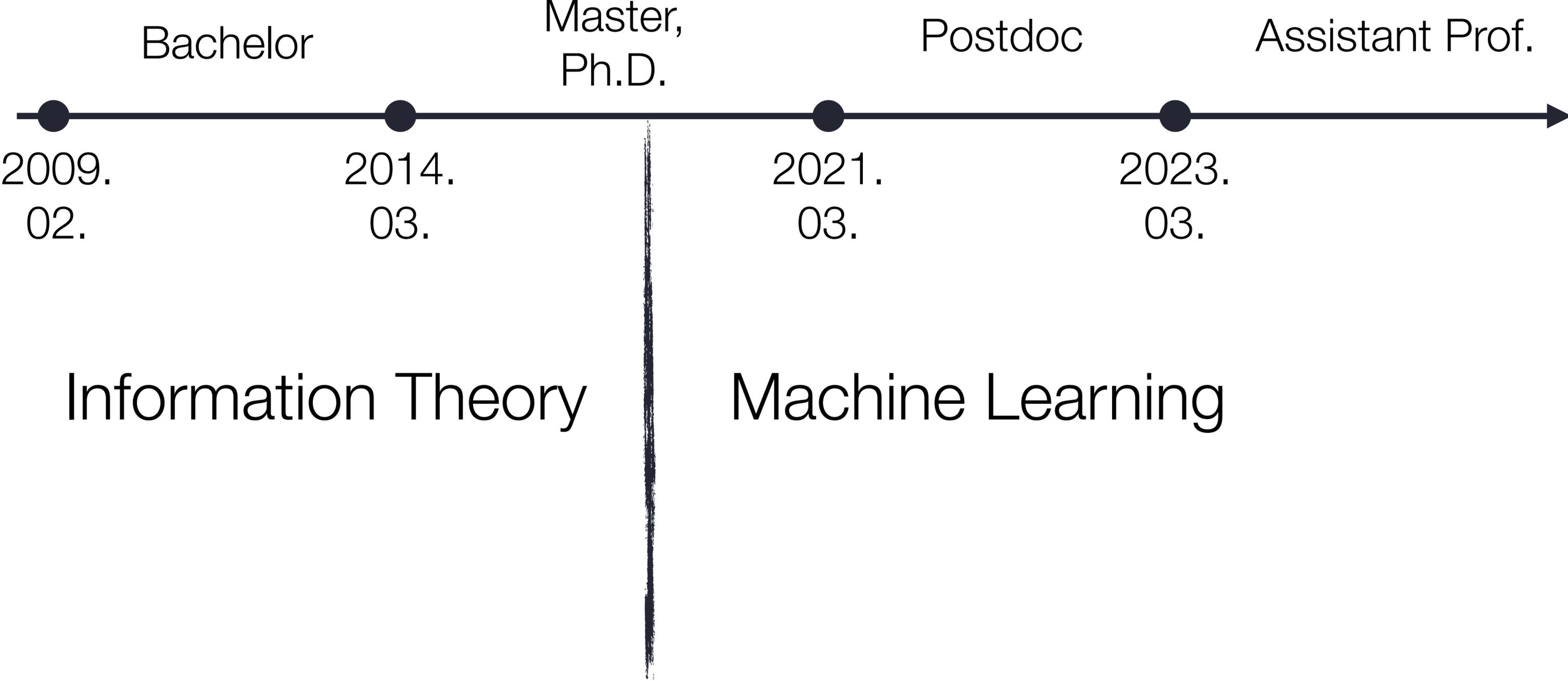
# A Brief Introduction



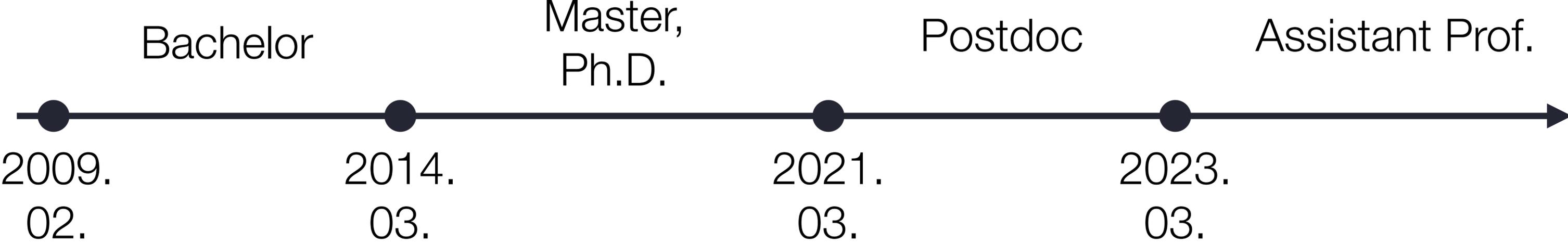
# A Brief Introduction



# A Brief Introduction



# A Brief Introduction



Information Theory and Machine Learning (ITML) lab

[itml.yonsei.ac.kr](http://itml.yonsei.ac.kr)

# A Brief Introduction

- Current Research Topics
  - Theory: 정보이론, 기계학습이론, 최적화이론 (for understanding AI)
  - Application: LLM/GPT를 의료, 법률 등에 활용하는 방법

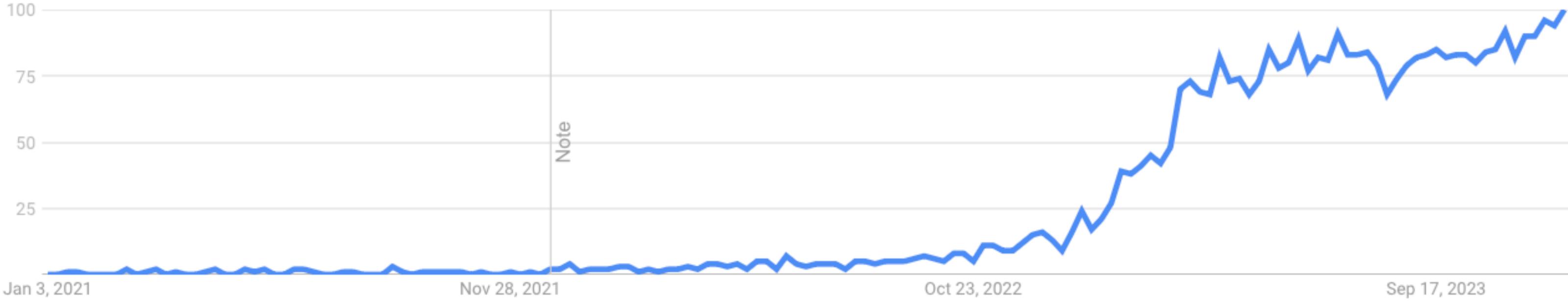


linq

 EverEx

# LLMs Dominated ML/AI Research in 2023

# Google Searches (Keyword: Large Language Models)



<https://trends.google.com>

# LLMs Dominated ML/AI Research in 2023

## ChatGPT Sprints to One Million Users

Time it took for selected online services to reach one million users



\* one million backers \*\* one million nights booked \*\*\* one million downloads  
Source: Company announcements via Business Insider/LinkedIn



# Battle of Top Tech Companies

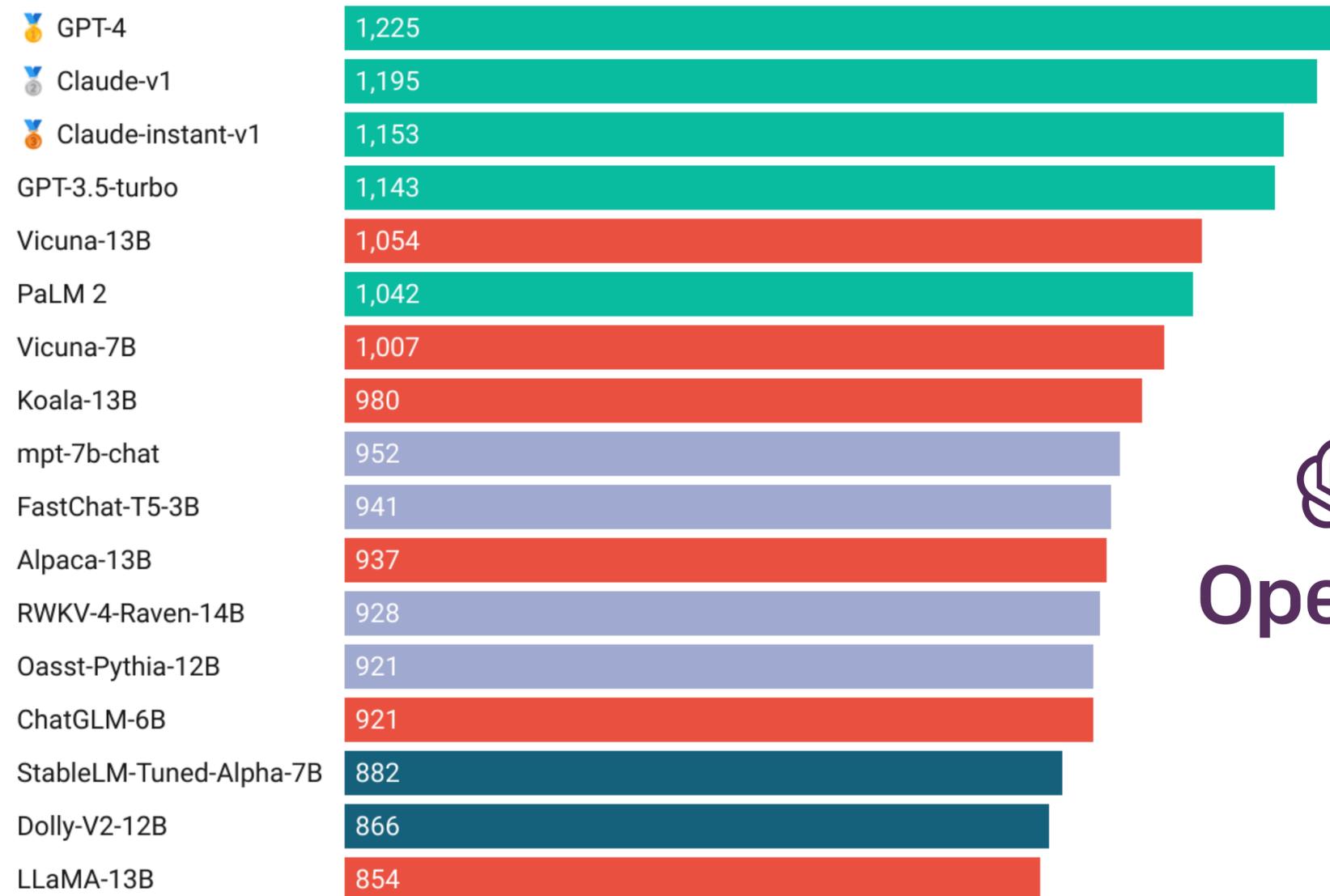


# Battle of Top Tech Companies

## Elo Ratings of LLMs

Project by LMSYS Org to rank LLMs via voting of community members (Timeframe: April 24 - May 22, 2023)

Proprietary Open Source/Research Apache 2.0 Other



Originally published by Newcomer in the post "14 Charts That Tell the Story of AI Right Now"

Chart: newcomer.co • Source: LMSYS Org • Created with Datawrapper

# Battle of Top Tech Companies

Rank* (UB) ▲	Model ▲	Arena Elo ▲	95% CI ▲	Votes ▲	Organization ▲	License ▲	Knowledge Cutoff ▲
1	<a href="#">GPT-4-Turbo-2024-04-09</a>	1258	+3/-3	44592	OpenAI	Proprietary	2023/12
2	<a href="#">GPT-4-1106-preview</a>	1252	+2/-3	76173	OpenAI	Proprietary	2023/4
2	<a href="#">Gemini 1.5 Pro API-0409-Preview</a>	1249	+3/-3	61011	Google	Proprietary	2023/11
2	<a href="#">Claude 3 Opus</a>	1248	+2/-2	101063	Anthropic	Proprietary	2023/8
3	<a href="#">GPT-4-0125-preview</a>	1246	+3/-2	70239	OpenAI	Proprietary	2023/12
6	<a href="#">Bard (Gemini Pro)</a>	1208	+5/-6	12387	Google	Proprietary	Online
6	<a href="#">Llama-3-70b-Instruct</a>	1208	+3/-3	75844	Meta	Llama 3 Community	2023/12
7	<a href="#">Reka-Core-20240501</a>	1199	+4/-4	18735	Reka AI	Proprietary	Unknown
8	<a href="#">Claude 3 Sonnet</a>	1200	+2/-3	84252	Anthropic	Proprietary	2023/8
10	<a href="#">GPT-4-0314</a>	1189	+3/-3	53446	OpenAI	Proprietary	2021/9
10	<a href="#">Command R+</a>	1189	+3/-3	50490	Cohere	CC-BY-NC-4.0	2024/3

# 1. LLM Basics

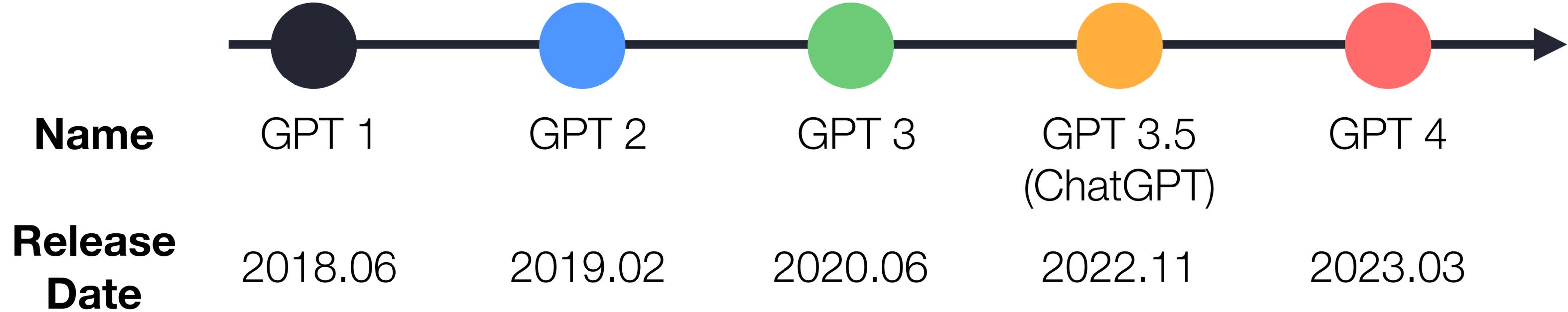


# GPT

- Generative: 새로운 것을 생성하는
- Pretrained: 사전학습 과정을 거친
- Transformer: 특정 뉴럴넷 구조의 이름  
(input: sequence, output: sequence)



# GPT: Generative Pretrained Transformer



- [GPT1] Improving **Language** Understanding by Generative Pre-Training
- [GPT2] **Language Models** are Unsupervised Multitask Learners
- [GPT3] **Language Models** are Few-Shot Learners

# GPT is a Language Model (LM)

- “My name is Jy-yong Sohn and I am an Assistant \_\_\_\_\_”
- Q: What is the most likely missing word?

# GPT is a Language Model (LM)

- “My name is Jy-yong Sohn and I am an Assistant Professor”
- Q: What is the most likely missing word?

# GPT is a Language Model (LM)

- “My name is Jy-yong Sohn and I am an Assistant Professor”
- Q: What is the most likely missing word?
  
- A “**language model**” (LM) is a model that attempts to generate sentences that sound “typical” enough

# GPT is a Language Model (LM)

- “My name is Jy-yong Sohn and I am an Assistant Professor”
- Q: What is the most likely missing word?
- A “**language model**” (LM) is a model that attempts to generate sentences that sound “typical” enough
- Often viewed as “a probability distribution”

$$\Pr(\underbrace{\text{word}_n}_y \mid \underbrace{\text{word}_1, \text{word}_2, \dots, \text{word}_{n-1}}_x)$$

(Completion) (Prompt)

# How is GPT Trained?

Language modeling == next word prediction

# How is GPT Trained?

Language modeling == next word prediction

A paragraph  
crawled from Internet

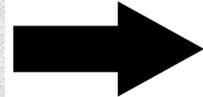
*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*

# How is GPT Trained?

Language modeling == next word prediction

A paragraph  
crawled from Internet

*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*



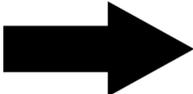
**x:** new, **y:** ML

# How is GPT Trained?

Language modeling == next word prediction

A paragraph  
crawled from Internet

*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*



**x:** new, **y:** ML

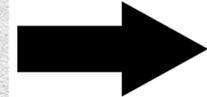
**x:** new ML, **y:** theory

# How is GPT Trained?

Language modeling == next word prediction

A paragraph  
crawled from Internet

*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*



**x:** new, **y:** ML

**x:** new ML, **y:** theory

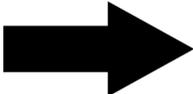
**x:** new ML theory, **y:** and

# How is GPT Trained?

Language modeling == next word prediction

A paragraph  
crawled from Internet

*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*



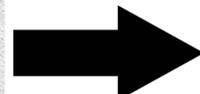
- x: new, y: ML*
- x: new ML, y: theory*
- x: new ML theory, y: and*
- ...
- (~ 500 billion pairs for GPT-3)

# How is GPT Trained?

Language modeling == next word prediction

A paragraph  
crawled from Internet

*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*



- x:** new, **y:** ML
- x:** new ML, **y:** theory
- x:** new ML theory, **y:** and

...

(~ 500 billion pairs  
for GPT-3)

train



# How is GPT Trained?

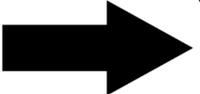
Language modeling == next word prediction

A paragraph  
crawled from Internet

*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*

→ **x:** new, **y:** ML  
**x:** new ML, **y:** theory  
**x:** new ML theory, **y:** and  
...  
(~ 500 billion pairs for GPT-3)

train



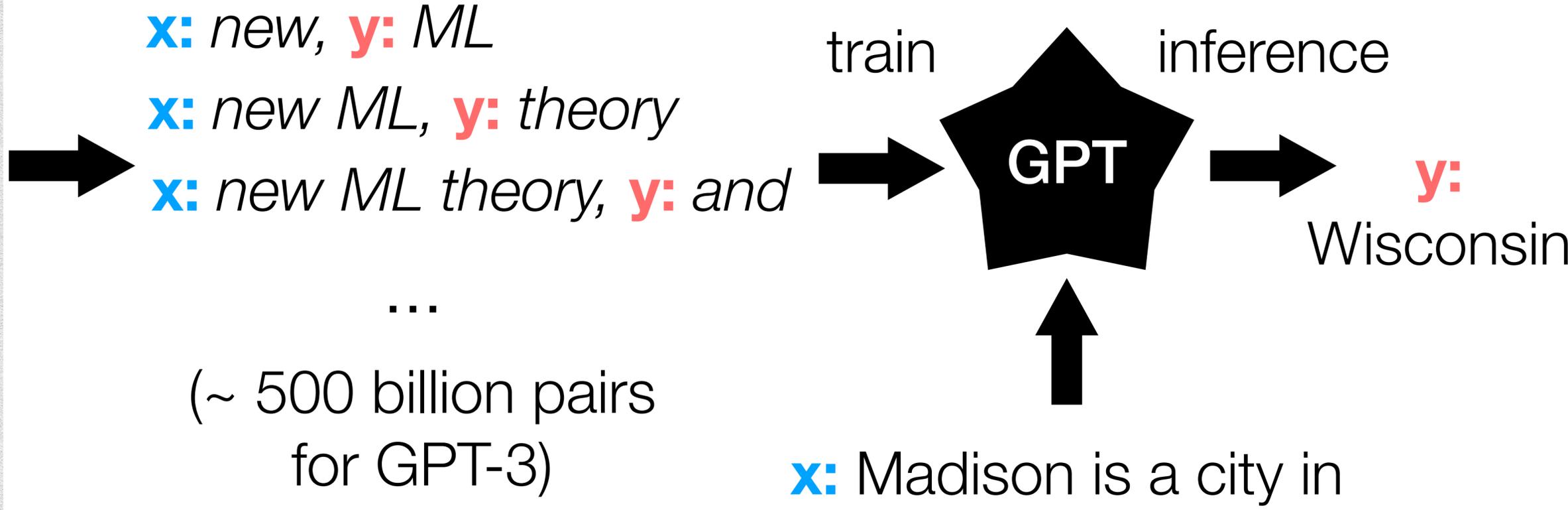
**x:** Madison is a city in

# How is GPT Trained?

Language modeling == next word prediction

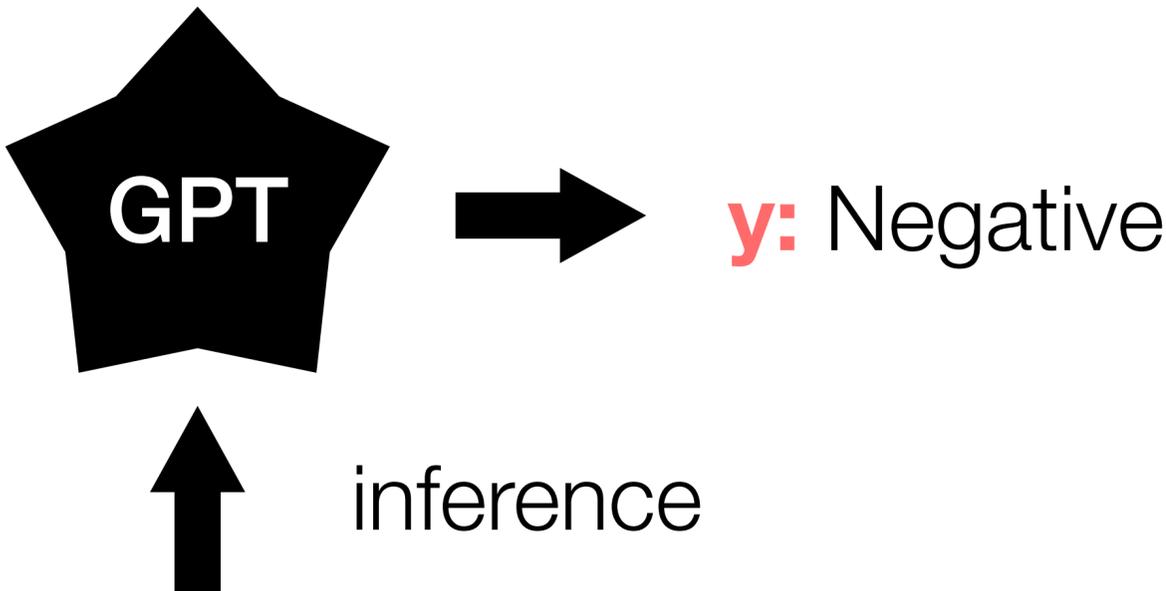
A paragraph  
crawled from Internet

*“new ML theory and methods to enable the rapid training and/or adaptation of predictors using a small number of labeled training examples”*



# How to use GPT

Option #1: zero-shot (do not use any data for the target task)



**x:** <I hate this movie> Is this review comment positive or negative?

# How to use GPT

Option #2: “fine-tune” a pre-trained GPT (use small data for the target task)



- **x:** <I hate this movie> Is this review positive or negative? **y:** Negative.
- **x:** <cool> Is this review positive or negative? **y:** Positive.

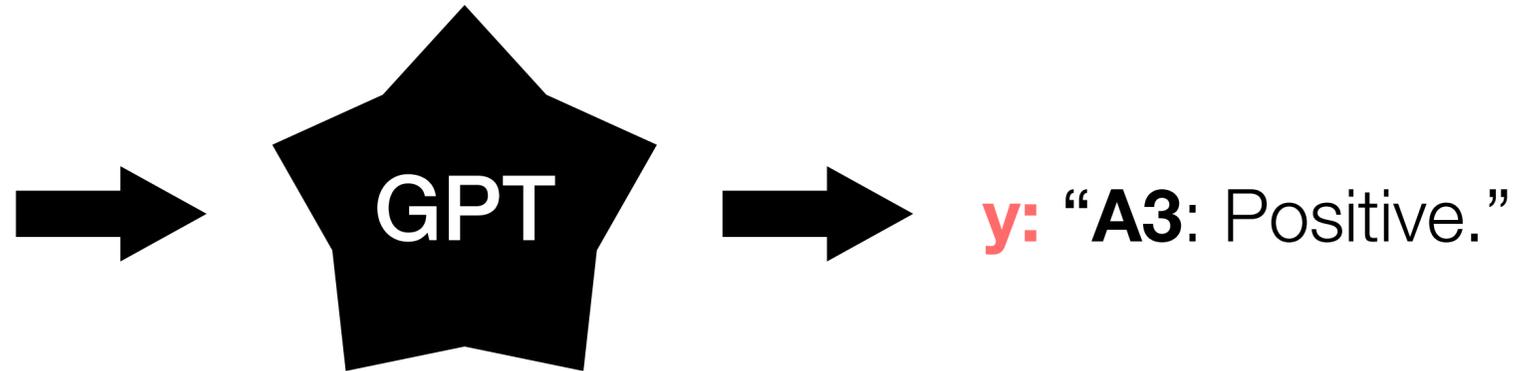
# How to use GPT

Option #3: “in-context learning” (use small data for the target task)

**x:** “**Q1:** <I hate this movie.>  
Is this review comment  
positive or negative?  
**A1:** Negative

**Q2:** <cool.>  
Is this review comment  
positive or negative?  
**A2:** Positive

**Q3:** <Awesome.>  
Is this review comment  
positive or negative?”



# LLM for Various Tasks



## NLP Tasks

- Question Answering
- Summarization
- Translation
- Reading Comprehension
- Solving SAT

# LLM for Various Tasks



## NLP Tasks

- Question Answering
- Summarization
- Translation
- Reading Comprehension
- Solving SAT

## Non-NLP Tasks

Recently, we can put images/audio/etc. to LLMs  
(So called multi-modal LLM)

# Example: Chatbot

근육량을 늘리고, 체지방은 빼주는 일주일 다이어트 식단을 알려줘. 단, 하루에 1500 칼로리가 넘지 않게 부탁해~

오후 1:51

AskUp

일주일 동안 근육량을 늘리고 체지방을 빼기 위해서는 충분한 단백질과 규칙적인 운동이 필요합니다. 1500 칼로리 이내의 식단을 제공드리겠습니다.

아침:  
- 스프린치 계란 후라이와 야채 샐러드  
(스프린치 계란 후라이: 계란 2개, 스프린치 1컵, 체다치즈 1/4컵, 올리브유 약간, 소금과 후추로 간을 맞춘 후 팬에서 구워주세요.)

점심:  
- 그릴에 구운 연어와 새싹 샐러드  
(연어 4온스, 새싹 채소 1컵, 오이 1/2개, 레몬즙 약간, 올리브유 약간, 소금과 후추로 간을 맞추어 그릴에 구워주세요.)

저녁:  
- 닭가슴살 샐러드와 삶은 계란  
(닭가슴살 4온스, 믹스 샐러드 1컵, 삶은 계란 1개, 올리브유 약간, 레몬즙 약간, 소금과 후추로 간을 맞춘 후 샐러드에 곁들여주세요.)

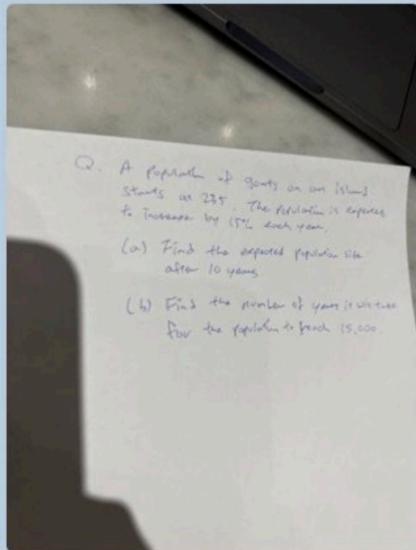
간식:  
- 그릴에 구운 양배추 칩  
(양배추 1컵, 올리브유 약간, 소금으로 간을 맞춘 후 그릴에 구워주세요.)

이 식단은 하루에 약 **1300-1400** 칼로리 정도이므로 간식을 추가로 먹지 않는 것이...

전체보기 >

오후 1:51

오후 4:29



AskUp

이미지 에서 233 글자를 읽었습니다.

오후 4:29

오후 4:29 a 만 풀어줘

AskUp

(a) 10년 후의 예상 인구는 다음과 같습니다.

초기 인구: 235  
연간 증가율: 15%

10년 후의 인구 =  $235 \times (1 + 0.15)^{10}$   
 $\approx 235 \times 4.046$   
 $\approx 951.11$  (소수점 이하 버림)

따라서, 10년 후의 예상 인구는 약 951마리입니다.

오후 4:29

# Example: Chatbot (cont'd)

< AI 여행플래너

 더 브로드 미술관을 포함한 LA여행 일정을 짜줘

 마이리얼트립

첫날:  
로스앤젤레스 도착  
호텔 체크인  
할리우드 명예의 거리 탐험  
TCL 중국어 극장 방문

2일차:  
게티 센터 방문  
할리우드 힐즈 탐험  
넓은 미술관을 방문

3일차:  
그리피스 천문대 방문  
산타모니카 부두 탐험  
라 브레아 타르 구덩이 방문

4일째:  
로스앤젤레스 카운티 미술관 방문  
베니스 해변 보드워크 탐험  
월트 디즈니 콘서트 홀 방문

5일째:  
할리우드 볼 방문  
유니버설 스튜디오 할리우드 탐험  
로스앤젤레스 출발

 메시지 입력 

아가기가 열이 나요.

 굿닥

만약 당신의 아기가 열이 난다면, 즉시 소아과 의사에게 연락하는 것이 중요합니다. 담당 의사가 상황을 평가하고 열을 치료하는 방법에 대한 최상의 조언을 제공할 수 있을 것입니다. 발열의 심각성에 따라, 의사는 이부프로펜이나 아세트아미노펜과 같은 처방전 없이 살 수 있는 약을 추천하거나 다른 치료법을 제안할 수 있습니다. 의사의 지시를 따르고 아기의 체온을 면밀히 관찰하는 것이 중요합니다.

[소아과 예약하기](#) >

무엇이든 물어보세요! 

# Example: Programming

Input:

JY

Please make a python code that generates 1000 samples from 1-dimensional unit-variance zero-mean gaussian distribution and plots the histogram of the generated 1000 samples. We should also include real probability distribution function to compare with the histogram

# Example: Programming

Input:

JY

Please make a python code that generates 1000 samples from 1-dimensional unit-variance zero-mean gaussian distribution and plots the histogram of the generated 1000 samples. We should also include real probability distribution function to compare with the histogram

Output:

python

 Copy code

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Generate 1000 samples from a 1-dimensional Gaussian distribution
samples = np.random.normal(loc=0, scale=1, size=1000)

# Plot histogram of the generated samples
plt.hist(samples, bins=30, density=True, alpha=0.5, edgecolor='black', label='Histogram')

# Plot the real probability distribution function (PDF)
x = np.linspace(-4, 4, 1000)
pdf = norm.pdf(x, loc=0, scale=1)
plt.plot(x, pdf, color='red', label='PDF')
```

# Example: Programming

Input:

JY

Please make a python code that generates 1000 samples from 1-dimensional unit-variance zero-mean gaussian distribution and plots the histogram of the generated 1000 samples. We should also include real probability distribution function to compare with the histogram

Output:

python

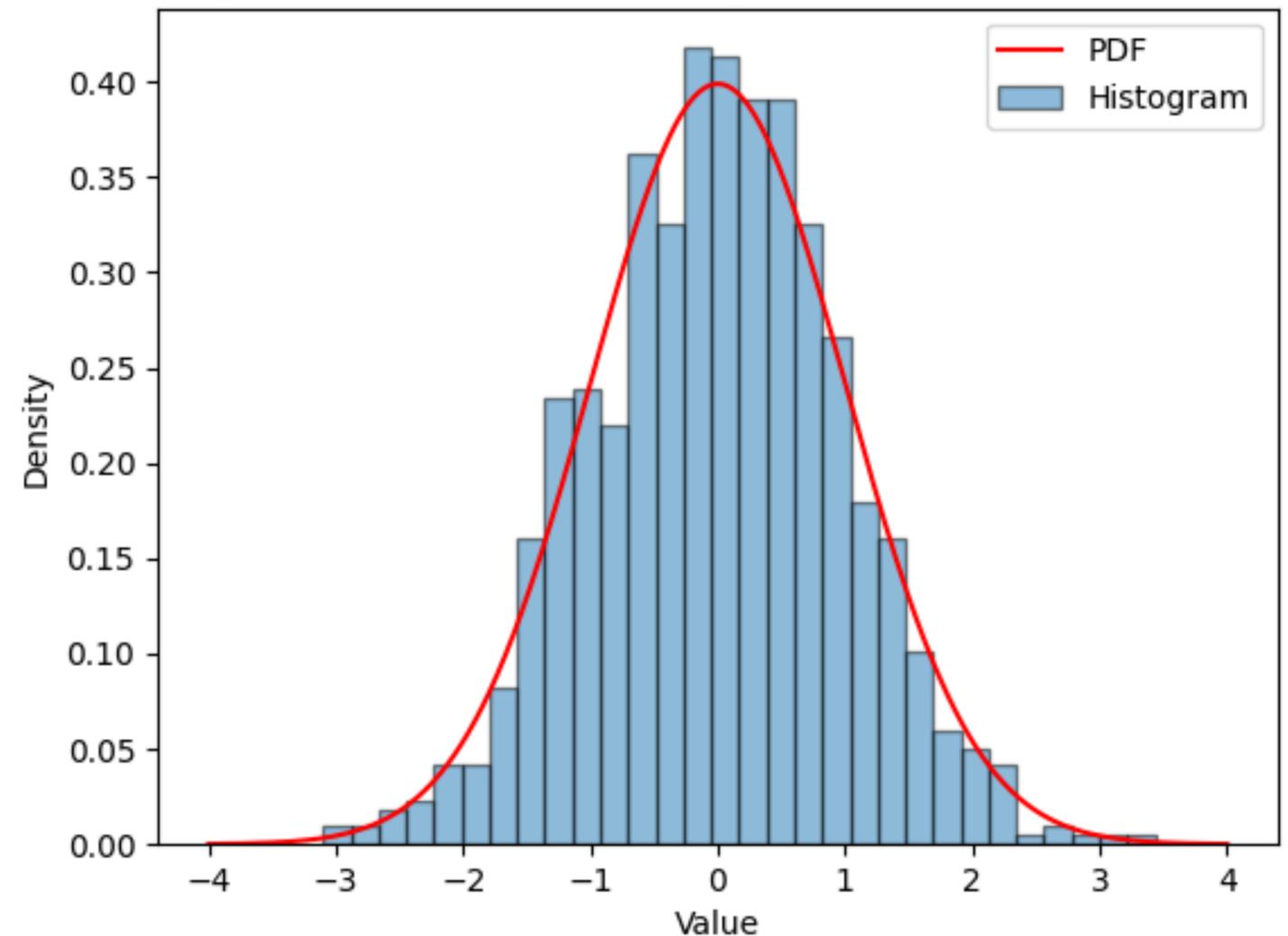
```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Generate 1000 samples from a 1-dimensional gaussian distribution
samples = np.random.normal(loc=0, scale=1, size=1000)

# Plot histogram of the generated samples
plt.hist(samples, bins=30, density=True)

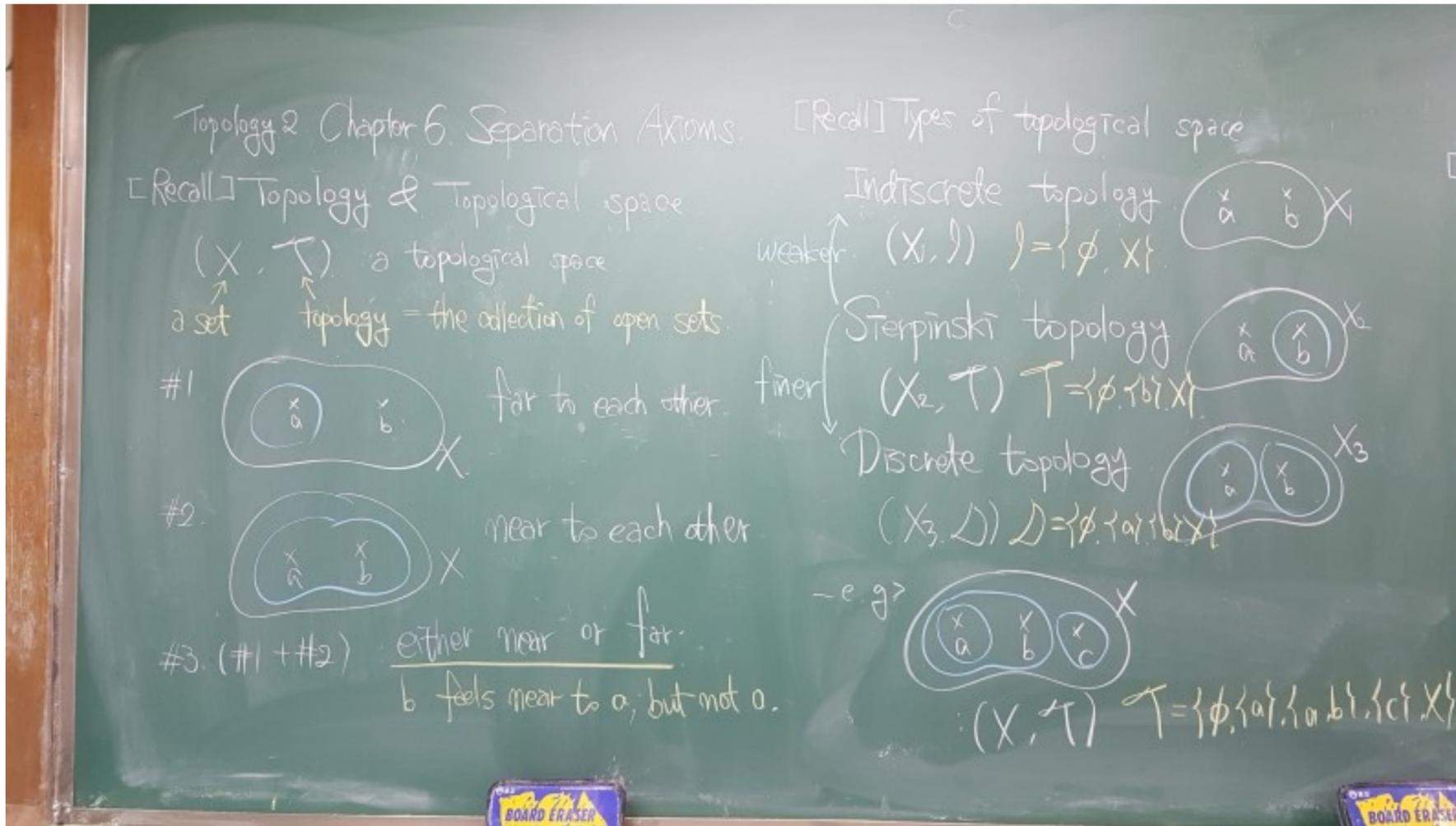
# Plot the real probability distribution function
x = np.linspace(-4, 4, 1000)
pdf = norm.pdf(x, loc=0, scale=1)
plt.plot(x, pdf, color='red', label='PDF')
```

Histogram and PDF of 1000 Samples from Gaussian Distribution



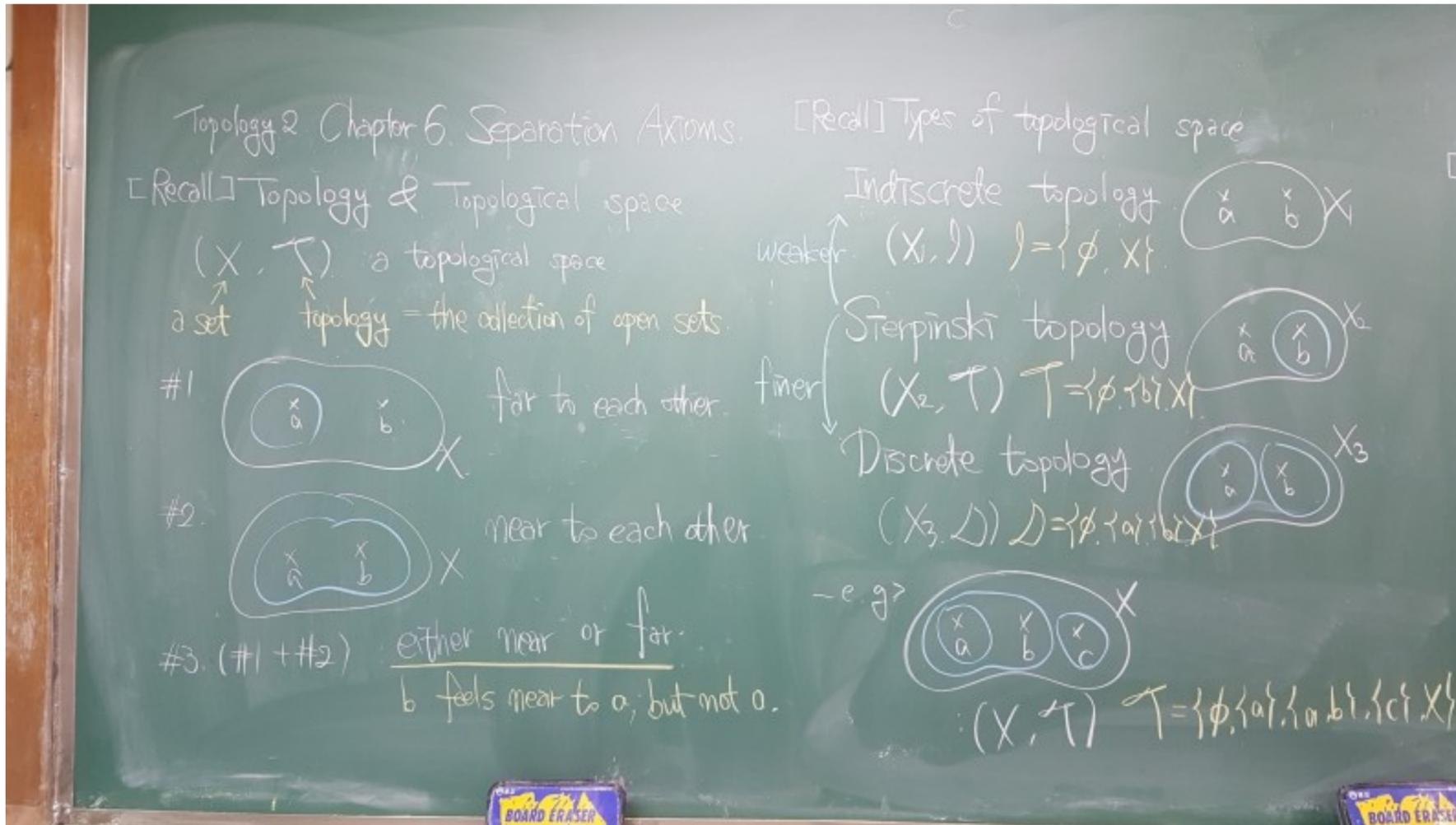
# Example: Make Lecture Note

Input:



# Example: Make Lecture Note

Input:



Output:

## Definition 1.1.1: Limit of Sequence in $\mathbb{R}$

Let  $\{s_n\}$  be a sequence in  $\mathbb{R}$ . We say

$$\lim_{n \rightarrow \infty} s_n = s$$

where  $s \in \mathbb{R}$  if  $\forall$  real numbers  $\epsilon > 0 \exists$  natural number  $N$  such that for  $n > N$

$$s - \epsilon < s_n < s + \epsilon \text{ i.e. } |s - s_n| < \epsilon$$

## Question 1

Is the set  $x\text{-axis} \setminus \{\text{Origin}\}$  a closed set

**Solution:** We have to take its complement and check whether that set is a open set i.e. if it is a union of open balls

## Note:-

We will do topology in Normed Linear Space (Mainly  $\mathbb{R}^n$  and occasionally  $\mathbb{C}^n$ ) using the language of Metric Space

## Claim 1.1.1 Topology

Topology is cool

## Example 1.1.1 (Open Set and Close Set)

- Open Set:
- $\emptyset$
  - $\bigcup_{x \in X} B_r(x)$  (Any  $r > 0$  will do)
  - $B_r(x)$  is open
- Closed Set:
- $X, \emptyset$
  - $\overline{B_r(x)}$
  - $x\text{-axis} \cup y\text{-axis}$

## Theorem 1.1.1

If  $x \in$  open set  $V$  then  $\exists \delta > 0$  such that  $B_\delta(x) \subset V$

**Proof:** By openness of  $V$ ,  $x \in B_r(u) \subset V$

## 2. LLMs for Medical Domain



# My Current Topics in LLM for Medical Domain

1. Using LLMs for Diagnosing Mental Disorder [**JMIR, under review**]
2. LLMs for Cognitive Distortion Classification [**NAACL 2024 Clinical Workshop, accepted**]

# 1. Using LLMs for Diagnosing Mental Disorder

## Aligning Large Language Models for Enhancing Psychiatric Interviews through Symptom Delineation and Summarization

Jae-hee So<sup>1†</sup>, Joonhwan Chang<sup>1†</sup>, Eunji Kim<sup>2,3</sup>, Junho Na<sup>1</sup>,  
JiYeon Choi<sup>5,6</sup>, Jy-yong Sohn<sup>‡1\*</sup>, Byung-Hoon Kim<sup>‡2,3,4\*</sup>,  
Sang Hui Chu<sup>‡5\*</sup>

<sup>1</sup>Department of Applied Statistics, Yonsei University.

<sup>2</sup>Department of Psychiatry, Yonsei University College of Medicine.

<sup>3</sup>Department of Biomedicine Systems Informatics, Yonsei University  
College of Medicine.

<sup>4</sup>Institute of Behavioral Sciences in Medicine, Yonsei University College  
of Medicine.

<sup>5</sup>Department of Nursing, Mo-Im Kim Nursing Research Institute, Yonsei  
University College of Nursing.

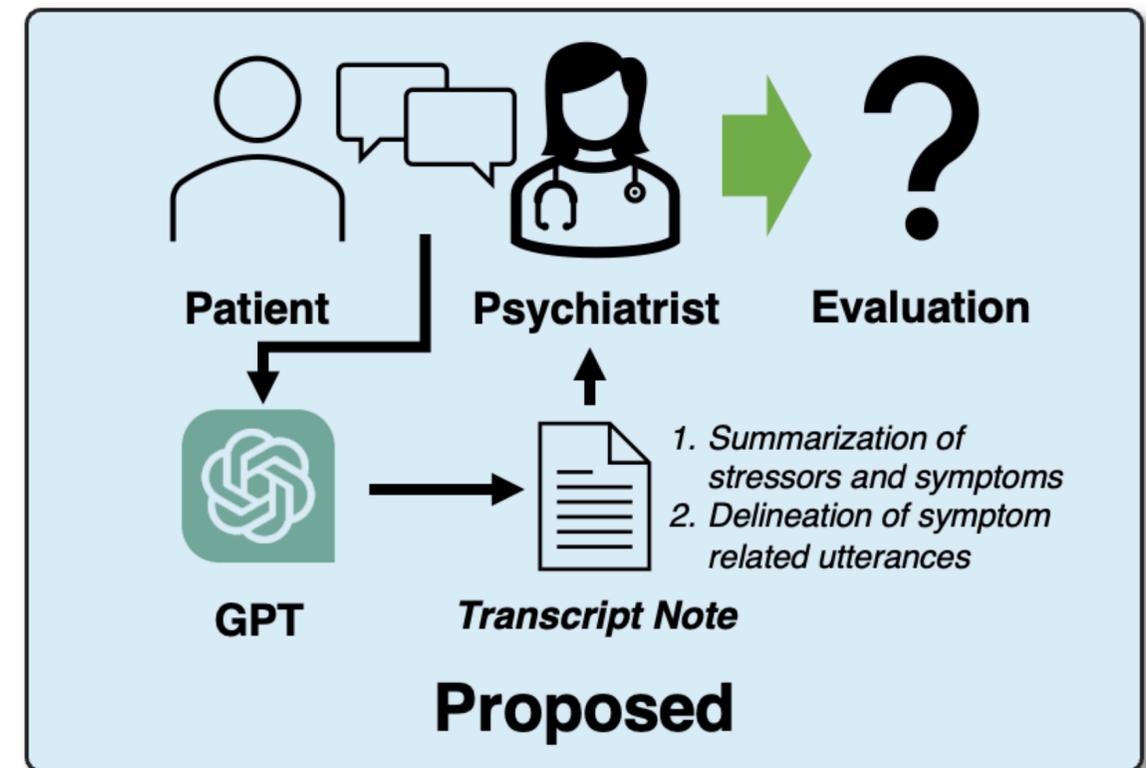
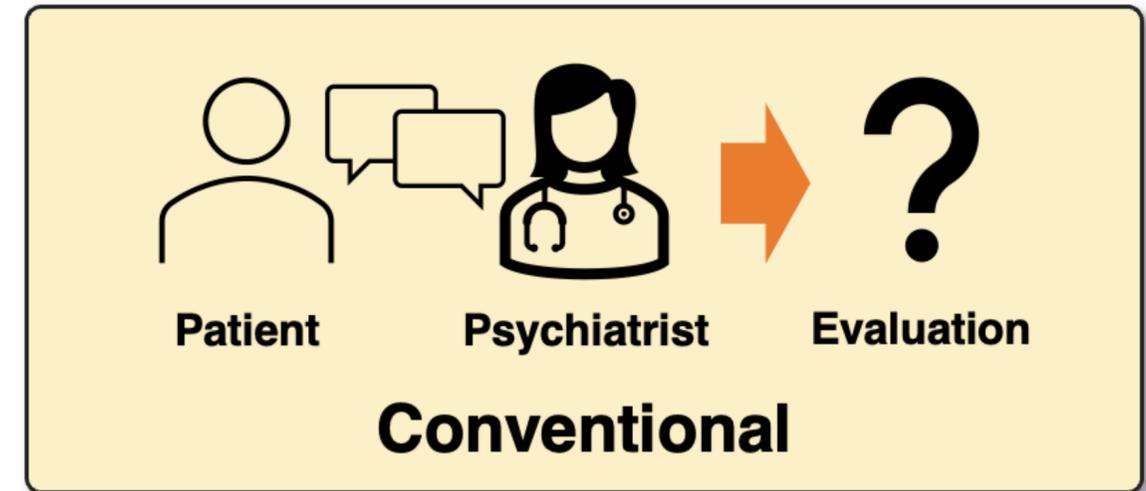
<sup>6</sup>Institute for Innovation in Digital Healthcare, Yonsei University.

‡The corresponding authors contributed equally to this work.

# 1. Using LLMs for Diagnosing Mental Disorder

- Joint work with Profs @ Yonsei
  - 간호학과 추상희, 최지연 교수
  - 의생명시스템정보학교실 김병훈 연구조교수

+ 소재희, 장준환, 김은지, 나준호 학생



# 1. Using LLMs for Diagnosing Mental Disorder

- Research Questions
  - RQ1: Can LLMs (1) delineate which part of the patients' utterances are related to psychiatric symptoms and (2) name the corresponding symptoms?

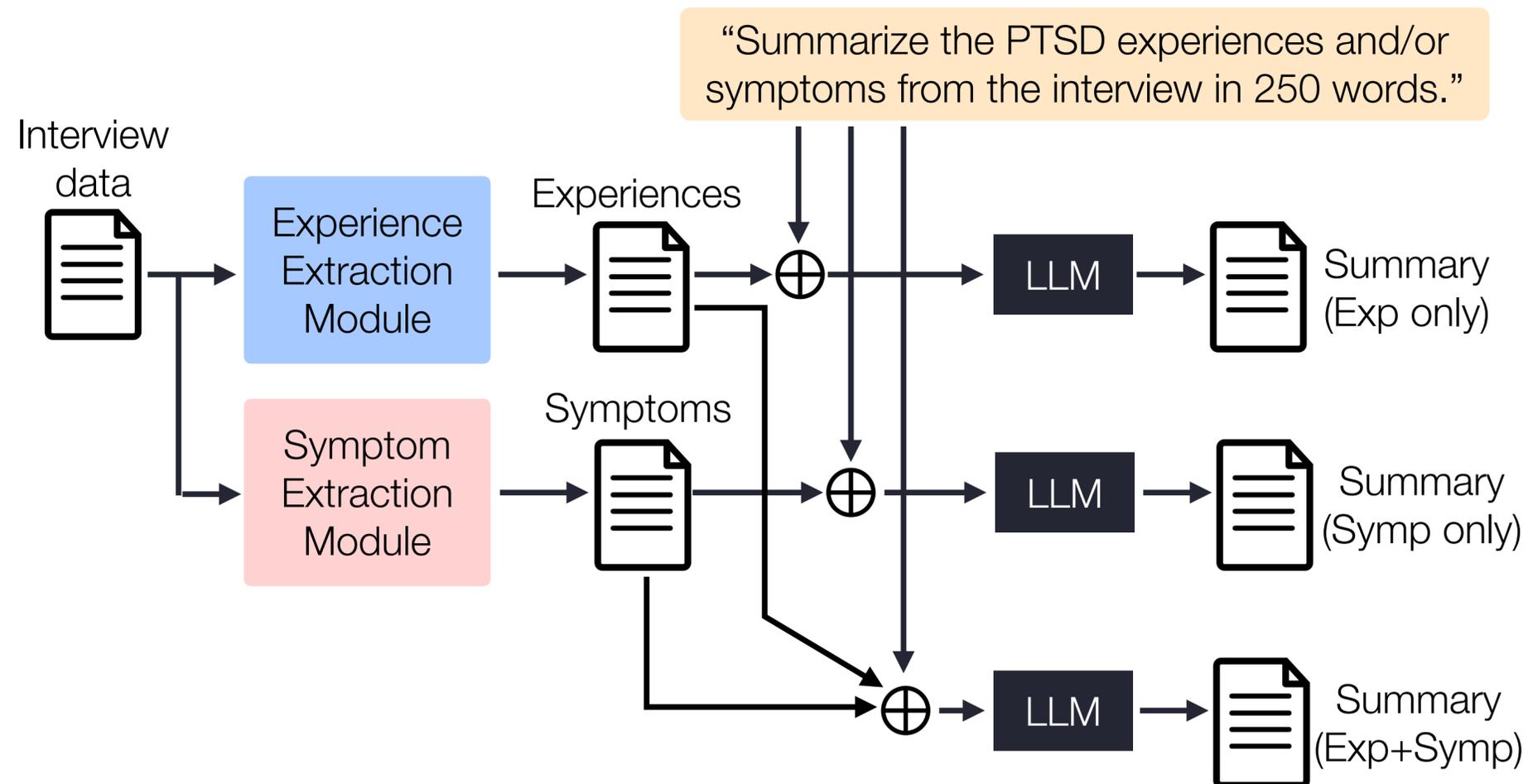


# 1. Using LLMs for Diagnosing Mental Disorder

- Research Questions
  - RQ2: Can LLMs summarize stressors and symptoms from an interview between a Post-traumatic stress disorder (PTSD) patient and a trained interviewer?

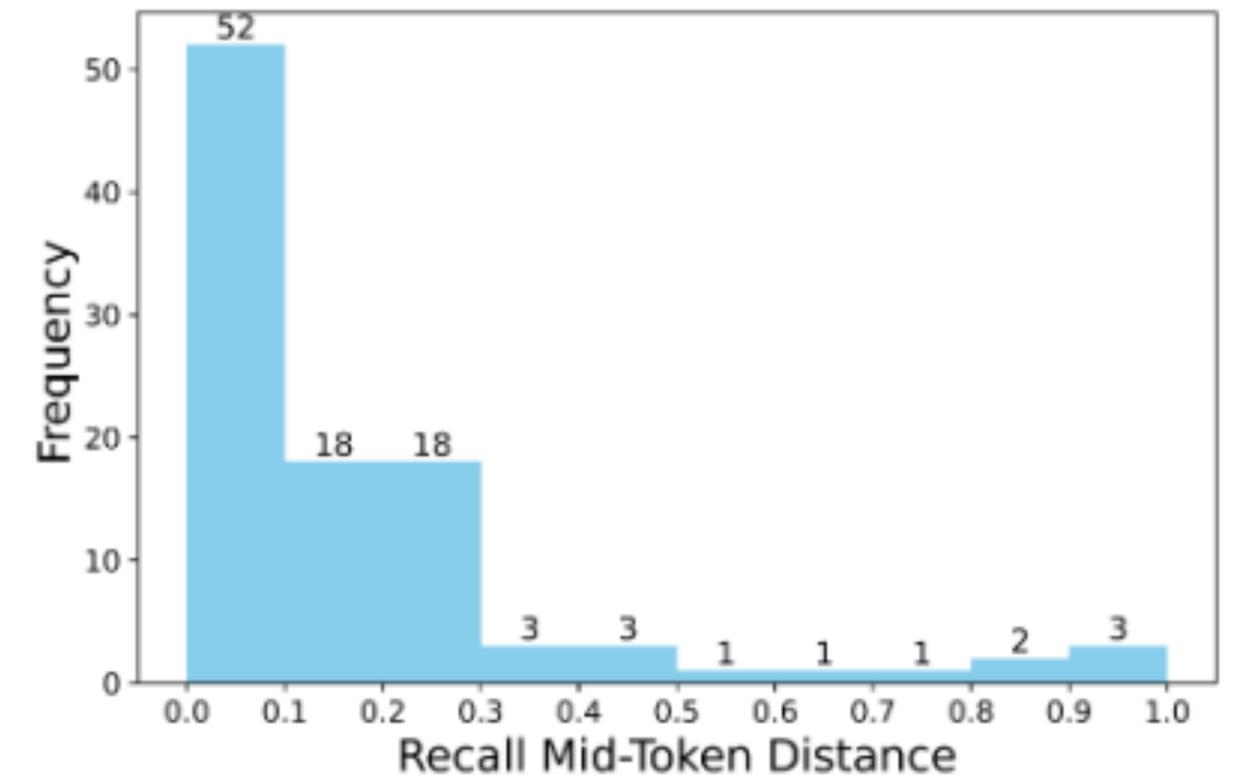
# 1. Using LLMs for Diagnosing Mental Disorder

- Research Questions
  - RQ2: Can LLMs summarize stressors and symptoms from an interview between a Post-traumatic stress disorder (PTSD) patient and a trained interviewer?



# Result: Section Estimation

Recall mid-token distance	Ground-truth Section	Estimated Section
0	But when I dream about it, I dream about the scene of my escape, the scene of my escape from North Korea, the scene of my escape from the police, and I still dream about it.	But when I dream about it, I dream about the scene of my escape, the scene of my escape from North Korea, the scene of my escape from the police, and I still dream about it.
0.11	Memory I don't really want to think about	I don't know, I haven't pulled it out in a long time, and it's actually a memory I don't really want to think about. Yeah.
0.27	That's what I still think about now, why did I say that, when he's gone, why did I say that, and that's what I regret.	Never the things of my heart. I am unjust. My heart is broken. I'm hurting. I'm just not expressing it.
0.4	Yes. That's hard and scary too.	It sounds like it's hard for you to be intimate with guys and have new relationships and stuff like that. P3: Yeah. That's hard and scary too.
1	It's because we're conditioned to think that anyone in black is someone who's out to get us.	None



Estimated section is quite similar to the ground-truth section (in many cases with small recall mid-token distance)

# Result: Symptom Estimation

Category of mental disorders	Symptom labels
PTSD	re-experiencing, avoidance, negative alterations in cognition, negative alterations in emotion, exaggerated arousal and reactivity, dissociation
C-PTSD	negative self-concept, difficulty in maintaining interpersonal relationships, emotional dysregulation
Major depressive episodes	depressed mood, loss of interest, decrease in appetite, increase in appetite, insomnia, hypersomnia, psychomotor agitation, psychomotor retardation, fatigue, feelings of worthlessness, excessive guilt, difficulty in concentration, impaired memory, impaired judgment, suicidal ideation, suicide planning, and suicide attempt.
Panic attack	general anxiety, heart palpitations, sweating, trembling, shortness of breath, choking, chest pain, nausea, dizziness, chills or heat sensations, paresthesia, dissociation, loss of control, and fear of dying
Alcohol use disorder	alcohol dependence, alcohol tolerance, alcohol withdrawal

Model	Method	Accuracy	Precision	Recall	F1-Measure
GPT-3.5 Turbo	Fine-Tuning	<b>0.817</b> $\pm$ 0.002	<b>0.828</b> $\pm$ 0.002	<b>0.818</b> $\pm$ 0.001	<b>0.821</b> $\pm$ 0.002
GPT-4 Turbo	ICL	0.537 $\pm$ 0.008	0.551 $\pm$ 0.009	0.550 $\pm$ 0.007	0.546 $\pm$ 0.008
GPT-4 Turbo	Zero-Shot	0.644 $\pm$ 0.004	0.649 $\pm$ 0.003	0.681 $\pm$ 0.002	0.657 $\pm$ 0.003
GPT-4 Turbo	Zero-Shot (w/ RAG)	0.708 $\pm$ 0.005	0.715 $\pm$ 0.007	0.745 $\pm$ 0.005	0.722 $\pm$ 0.005

# Result: Summary

- Summary of GPT on Patient P9

P9 experienced **extreme mental and physical stress** in North Korea and China. She faced natural disasters, extreme hunger, arrest of her family, and constant threats from public security, which led to fears of deportation and loss of her family. Experiential trauma manifested as regular nightmares, restless sleep, anxiety, avoidance, arousal, re-experiencing symptoms, and extreme stress reactions. She experienced severe hunger during the "hardship march" in North Korea, and legal insecurity and fear of family separation in China. These events led to her escape to South Korea and continue to impact her life today. The trauma recurs during the day and at night, and she continues to be extremely worried about her and her family's safety. Overt symptoms of mental health include anxiety, depression, suicidal thoughts, death-related dreams, overreaction in everyday life, and physical symptoms such as increased heart rate and sweating. **These symptoms meet the criteria for PTSD**, and Patient P9 is highly traumatized and in need of professional mental health care.

# Result: Summary

- Quantitative Evaluation of the summary generated by GPT

	Coherence	Consistency	G-Eval Fluency	Relevance	Overall
<b>Strs</b>	4.22 ± 0.19	4.02 ± 0.33	1.55 ± 0.60	4.21 ± 0.38	3.50 ± 0.26
<b>Symp</b>	4.43 ± 0.21	4.34 ± 0.71	1.15 ± 0.12	4.42 ± 0.17	3.59 ± 0.28
<b>Strs+Symp</b>	<b>4.66 ± 0.08</b>	<b>4.73 ± 0.07</b>	<b>2.16 ± 0.71</b>	<b>4.67 ± 0.13</b>	<b>4.01 ± 0.17</b>
<b>Strs (w/ RAG)</b>	4.31 ± 0.28	3.75 ± 0.85	1.45 ± 0.36	4.30 ± 0.28	3.41 ± 0.28
<b>Symp (w/ RAG)</b>	4.09 ± 0.41	3.92 ± 0.87	1.53 ± 0.69	4.09 ± 0.57	3.40 ± 0.48
<b>Strs+Symp (w/ RAG)</b>	4.51 ± 0.08	4.69 ± 0.09	2.11 ± 0.49	4.51 ± 0.17	3.96 ± 0.17

G-Eval score above 3.8 (out of 4.5): quite well done

# 2. LLMs for Cognitive Distortion Classification

## ERD: A Framework for Improving LLM Reasoning for Cognitive Distortion Classification

**Sehee Lim\***

Yonsei University

sehee0706@yonsei.ac.kr

**Yejin Kim\***

Yonsei University

yjkim.stat@yonsei.ac.kr

**Chi-Hyun Choi\***

EverEx

leo@everex.co.kr

**Jy-yong Sohn<sup>†</sup>**

Yonsei University

EverEx

jysohn1108@yonsei.ac.kr

**Byung-Hoon Kim<sup>†</sup>**

Yonsei University

EverEx

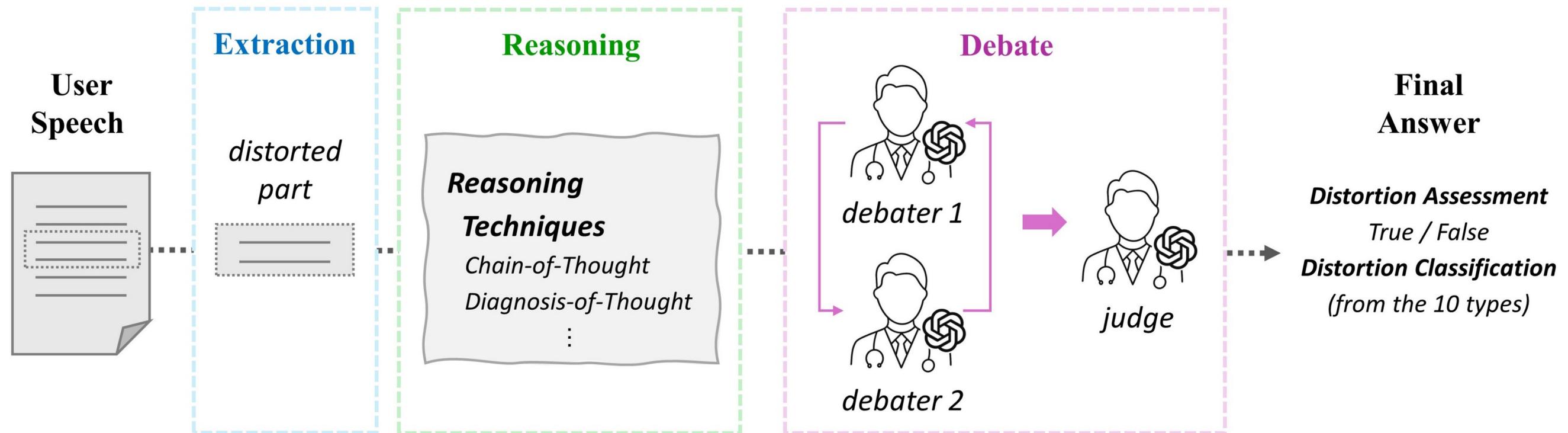
egyptdj@yonsei.ac.kr

# 2. LLMs for Cognitive Distortion Classification

- Joint work with EverEx
  - 보라매병원 정신건강의학과 최치현 교수
  - 의생명시스템정보학교실 김병훈 연구조교수



+ 임세희, 김예진 학생



# 2. LLMs for Cognitive Distortion Classification

- Input: User Speech
- Target Tasks
  - Distortion Assessment (True/False)
  - Distortion Classification (out of 10 types + No distortion)

<b>Distortion Type</b>	<b>Count</b>
All-or-nothing thinking	100
Emotional Reasoning	134
Fortune-telling	143
Labeling	165
Magnification	195
Mental filter	122
Mind Reading	239
Overgeneralization	239
Personalization	153
Should statements	107
No Distortion	933
<b>In Total</b>	<b>2530</b>

# 2. LLMs for Cognitive Distortion Classification

- LLMs can classify the cognitive distortion in the user speech
  - “Extraction” step improves the distortion classification performance
  - “Debate” step reduces the cases of false positives

Method	Distortion Assessment (True/False)			Distortion Classification (out of 10 types)
	Sensitivity	Specificity	F1 Score	Weighted F1 Score
Reasoning	<u>99.29</u> <sub>0.19</sub>	6.79 <sub>0.34</sub>	<b>78.26</b> <sub>0.16</sub>	15.28 <sub>0.65</sub>
+Extraction	<b>99.83</b> <sub>0.03</sub>	0.93 <sub>0.22</sub>	<u>77.48</u> <sub>0.04</sub>	<b>24.40</b> <sub>0.69</sub>
+Debate	73.10 <sub>0.26</sub>	<b>33.05</b> <sub>0.58</sub>	68.89 <sub>0.24</sub>	22.18 <sub>0.99</sub>
+Extraction+Debate	74.89 <sub>2.31</sub>	<u>30.74</u> <sub>3.92</sub>	69.49 <sub>0.62</sub>	<u>24.27</u> <sub>1.14</sub>

## 2. LLMs for Cognitive Distortion Classification

- Multiple rounds of “Debate” step improves the performance

<b>Metric</b>	<b>Round 1</b>	<b>Round 2</b>	<b>Round 3</b>
Binary F1	52.13 <sub>1.25</sub>	69.49 <sub>0.62</sub>	<b>70.74</b> <sub>0.44</sub>
Multi-class F1	22.79 <sub>1.62</sub>	24.27 <sub>1.14</sub>	<b>24.83</b> <sub>0.81</sub>

# Summary

- LLMs are Dominating the ML/AI Field
- LLMs for Medical Domain
  - My recent papers show that LLMs can assist psychiatric evaluations
  - What about other tasks?  
Probably better to rely on the success of multi-modal LLMs

# Thanks!

<https://itml.yonsei.ac.kr>  
[jysohn1108@gmail.com](mailto:jysohn1108@gmail.com)

