

Training Techniques and Research Trends of LLM

2023. 08.04

Data Mining & Quality Analytics Lab.

김현지

발표자 소개



이름

김현지

학력 사항

고려대학교 산업경영공학과 석사과정 (2022.03 ~ 현재)
Data Mining & Quality Analytics 연구실 (지도교수: 김성범 교수님)

관심 연구 분야

Deep Learning Algorithm for Multivariate Signal Analysis
Representation Learning for Time-Series Data

E-Mail

99ktxx@korea.ac.kr

CONTENTS

01

What are the LLMs?

02

Training Techniques of LLMs

- Pre-Training
- Adaptation Tuning
- Utilization

03

Research Trend of LLMs

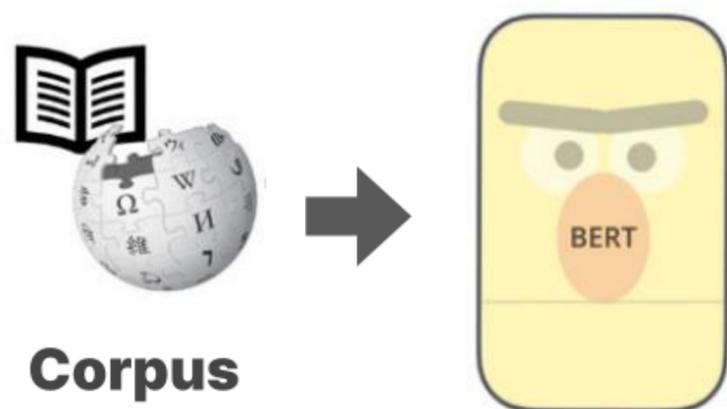
- From GPT-3 (2020.05) to Llama2 (2023.07)

① What are the LLMs?

1 What are the LLMs?

Pre-trained Language Model

많은 양의 텍스트 데이터를 활용해 일반적인 수준의 '언어 이해'가 가능하도록 사전 학습을 시킨 언어 모델

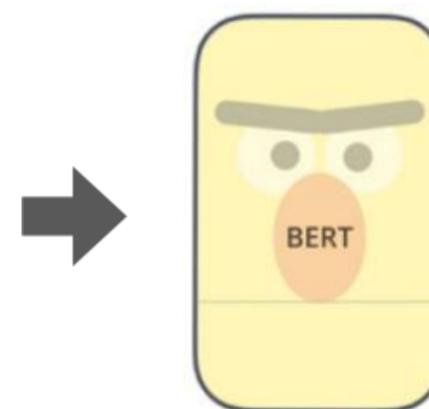


1. Pre-training

KorQuAD

Naver sentiment
movie corpus v1.0

Downstream
task



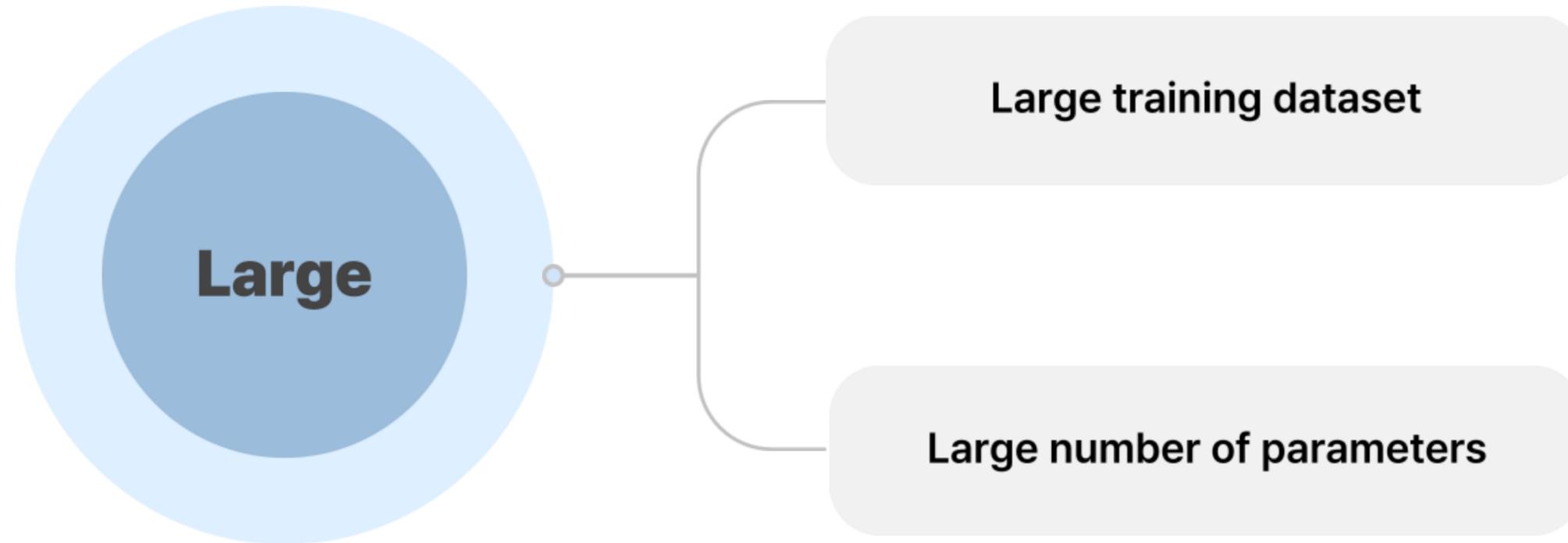
2. Fine-tuning

사전 학습된 모델을 텍스트 분류, 생성, 번역 등과 같은 다양한 downstream task에서 fine-tuning하여 사용할 수 있으며, task마다 end-to-end로 따로 학습하는 방식에 비해 여러 자연어 처리 task의 성능을 비약적으로 끌어올림

1 What are the LLMs?

Large Language Model

'거대한 모델'을 기반으로 '대량의 데이터'를 사용하여 자연어를 이해하고 생성할 수 있도록 학습된 언어 모델



일반적인 목적으로 사전 학습되며 이후 다양한 downstream task 를 위해 fine-tuning하여 사용될 수 있음

Benefits of using large language models

1

사전 학습된 단일 모델이 여러 task에 사용될 수 있음

- 하나의 LLM을 번역, 문장 생성, 문장 분류, QA 등 다양한 task에 사용할 수 있음

2

Fine-tuning을 위한 특정 task의 데이터는 많이 필요하지 않음

- Few-shot 혹은 zero-shot setting에서도 우수한 성능을 보임

3

학습 데이터와 모델의 파라미터를 증가시킴으로써 성능을 지속적으로 향상시킬 수 있음

② Training Techniques of LLMs

Training Process of LLMs

우수한 LLM 구축에는 거대한 모델과 다량의 데이터뿐만 아니라 다양한 학습 기법을 효과적으로 활용하는 것이 매우 중요

Pre-training

LLM의 기본적인 능력을
습득하는 과정

- Data
- Architecture
- Pre-training task
- Optimization and Scalable Training Techniques

Adaptation Tuning

LLM을 특정한 목표에 따라 추가로
adaptation 시키는 과정

- Instruction tuning
- Alignment Tuning
- Parameter efficient model adaptation

Utilization

LLM을 효과적으로
사용하는 방법

- Prompt Engineering

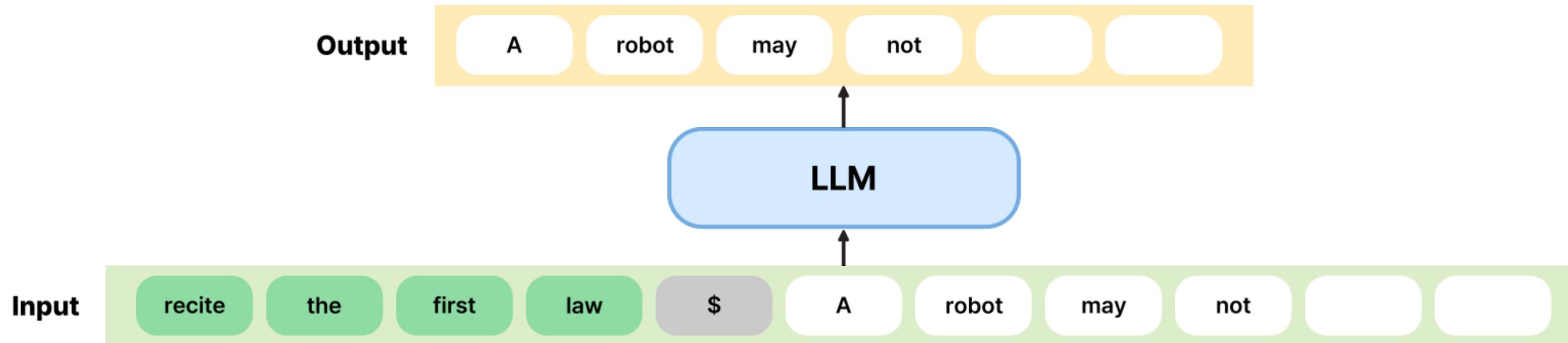
Pre-training

- Data** 데이터의 크기 뿐만 아니라 품질 또한 매우 중요하며, 모델의 성능은 pre-training 데이터 셋과 전처리 방식에 크게 좌우됨
- Architecture** Transformer 아키텍처는 뛰어난 병렬 처리 능력과 용량으로 인해 언어 모델을 수백, 수천억 개의 파라미터로 확장할 수 있어 다양한 LLM 아키텍처의 근간이 됨
- Pre-training Tasks** LLM을 학습하기 위해 다음에 올 단어를 예측하는 language modeling 방법이 많이 사용됨
- Optimization and Scalable Training Techniques** 모델과 데이터의 크기가 증가함에 따라 제한된 컴퓨팅 리소스 내에서 LLM을 효율적으로 학습시키는 것이 어려워지고 있으며 가속화 방법 및 최적화 기술이 잘 설계되어야 함

Pre-training

Pre-training Tasks

- LLM을 학습하기 위해 language modeling task가 가장 많이 사용됨
- Language modeling은 토큰 시퀀스가 주어졌을 때, 이전 토큰들을 기반으로 다음 토큰을 자기 회귀적으로 예측하는 것을 목표로 함
- 대부분의 자연어 처리 task들은 입력 데이터를 기반으로 한 예측 문제로 해결할 수 있기 때문에 language modeling으로 학습된 LLM은 다양한 task로 쉽게 확장 될 수 있음



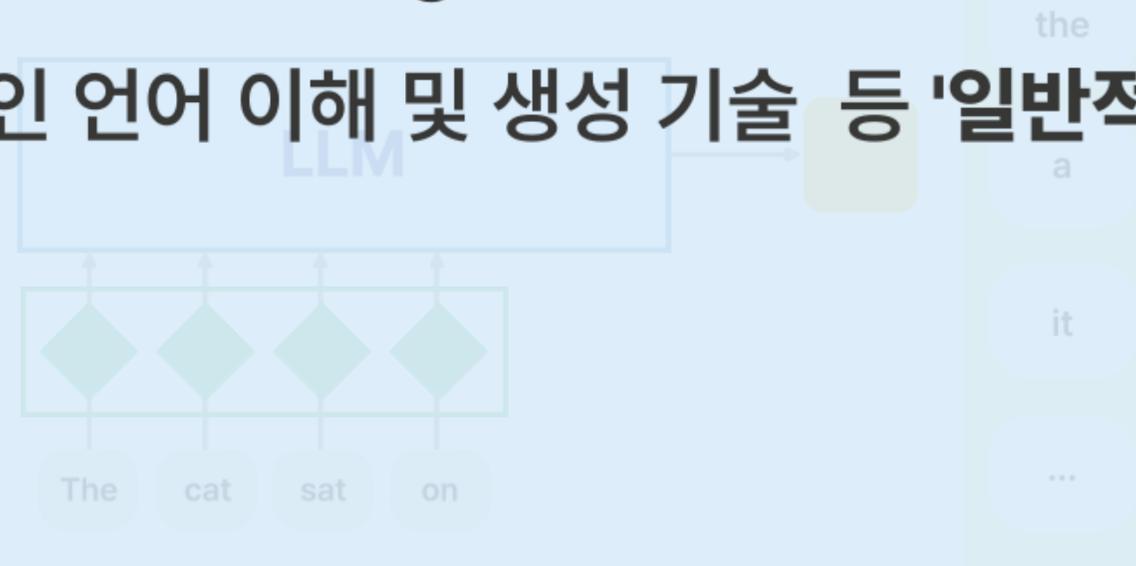
Pre-training

Pre-training Tasks

- LLM을 학습하기 위해 language modeling task가 가장 많이 사용됨
- Language modeling은 토큰 시퀀스가 주어졌을 때, 이전 토큰들을 기반으로 다음 토큰을 자기 회귀적으로 예측하는 것을 목표로 함
- 대부분의 자연어 처리 task들은 입력 데이터를 기반으로 한 예측 문제로 해결할 수 있기 때문에 language modeling으로 학습된 LLM은 다양한 task로 쉽게 확장 될 수 있음

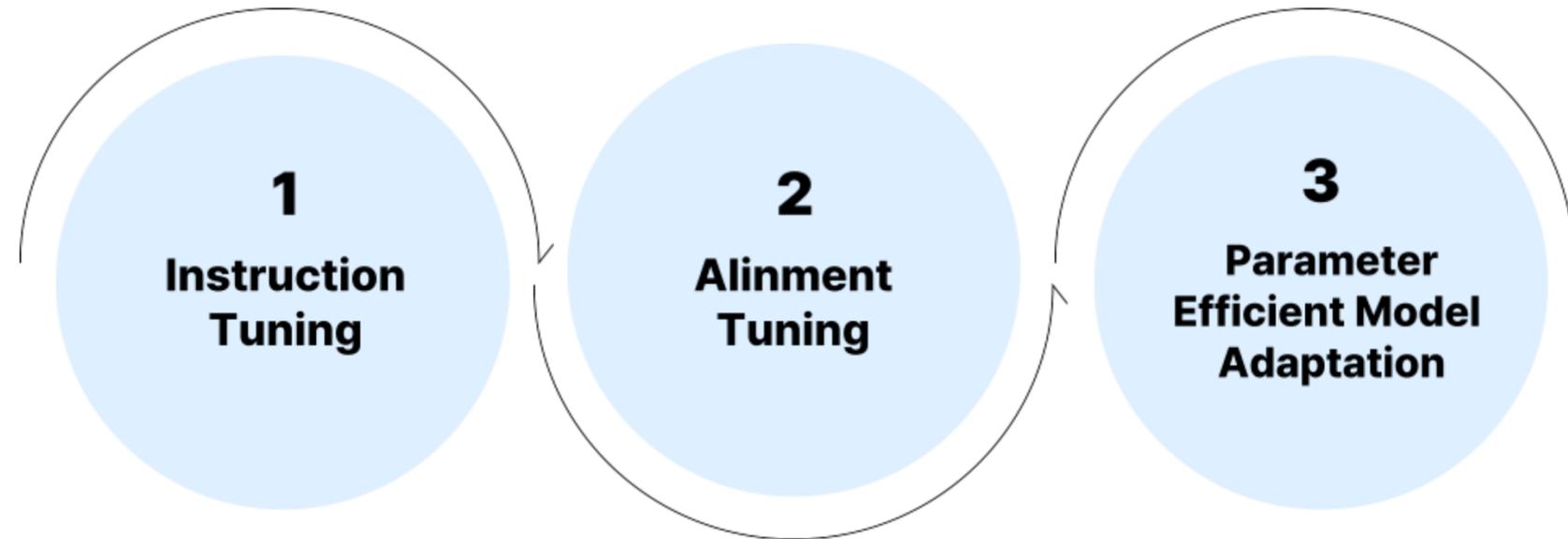
Pre-training을 통해 다양한 작업을

수행하기 위한 필수적인 언어 이해 및 생성 기술 등 '일반적인 능력'을 습득하게 됨



Adaptation Tuning

- 다양한 task에서 사전 학습된 LLM을 효과적으로 더 잘 사용하기 위해 adaptation 시키는 작업
- GPT-1, GPT-2, GPT-3까지는 pre-train을 통해 그 성능을 확보했다면 GPT-3에서 부터 InstructGPT, chatGPT까지는 adapatation tuning 통해 성능을 확보하기 시작함
- GPT-3에서 chatGPT로의 중요한 변화는 adpatation tuning의 수행으로 볼 수 있음



Adaptation Tuning

Instruction Tuning

등장 배경 1

사용자의 의도에 맞는 답변 출력

- LLM은 기본적으로 다음 단어를 예측하는 language modeling task로 pre-train 되기 때문에 사용자의 의도에 맞게 자연스럽게 답변하기 어려움
- LLM은 사용자의 입력 다음에 이어질 단어 혹은 문장들을 생성하기 때문에 질문에 대한 답이 아닌 이어질 만한 후속 문장을 출력할 가능성이 큼

Users intent

Q 한국의 수도는 어디야?

A 서울입니다.

Language modeling

Q 한국의 수도는 어디야?

A 한국에 처음 방문하는 외국인 관광객이 궁금해할 수 있습니다.

A 한국인이려면 모두 답을 알 것입니다.

Users intent \neq Language modeling

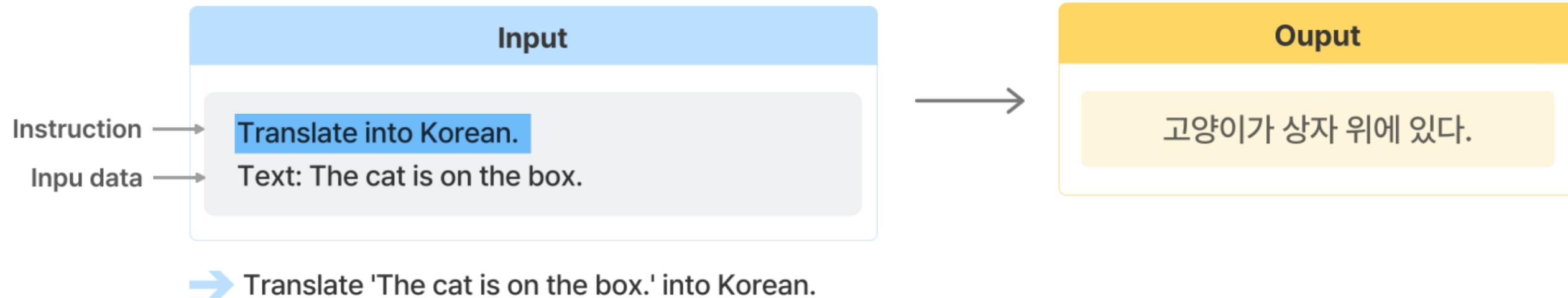
Adaptation Tuning

Instruction Tuning

등장 배경 1

사용자의 의도에 맞는 답변 출력

- Instruction Tuning: Instruction(지시문) & input text와 output text의 pair 데이터 셋을 통해 supervised 형식으로 학습시키는 것
- 사용자의 의도 및 task에 맞게 LLM을 잘 활용하기 위해서는 instruction-following 데이터로 tuning 하는 것이 꼭 필요함



Instruction tuning

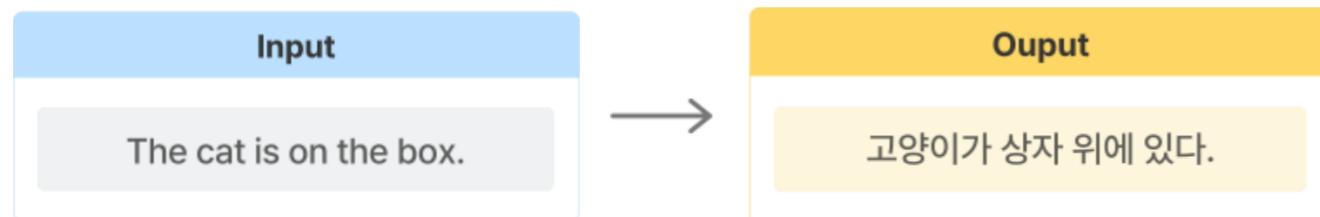
Adaptation Tuning

Instruction Tuning

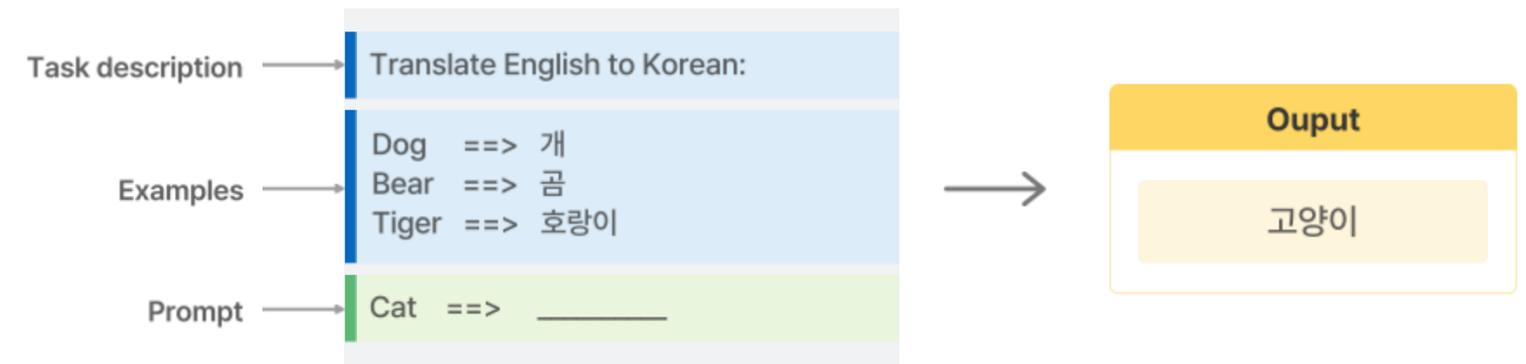
등장 배경 2

Zero-shot 성능 향상

- 기존의 tuning 기법: task 별 fine-tuning, 별도의 fine-tuning 없이 prompting
- 기존의 LLM은 few-shot setting에서의 성능은 전반적으로 향상되었으나, zero-shot setting에서의 성능은 few-shot 성능 만큼 향상되지 못함



Fine-tune on translation task



Prompting (GPT-3)

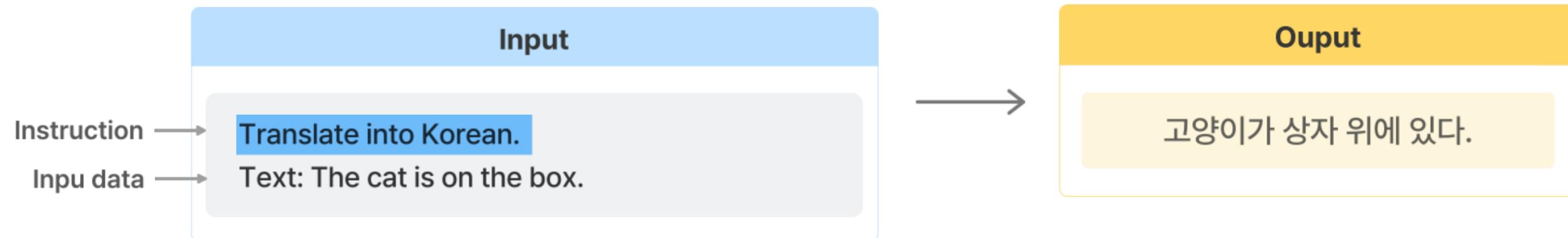
Adaptation Tuning

Instruction Tuning

등장 배경 2

Zero-shot 성능 향상

- NLP task를 instruction 형태로 변환하여 adaptation tuning 진행
- 다양한 task로 instruction 데이터를 구성하여 학습하면 새로운 task에서의 성능이 향상되며 zero-shot 성능 또한 향상됨



Instruction tuning

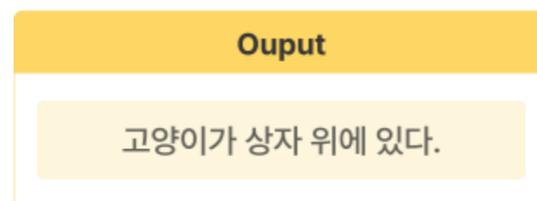
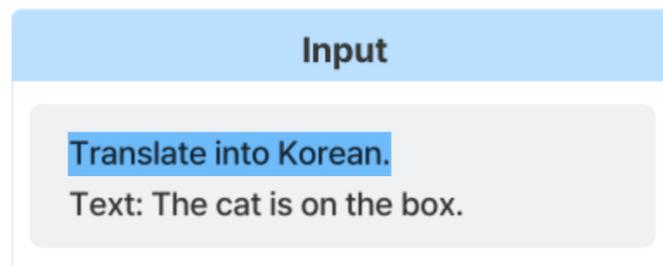
Adaptation Tuning

Instruction Tuning

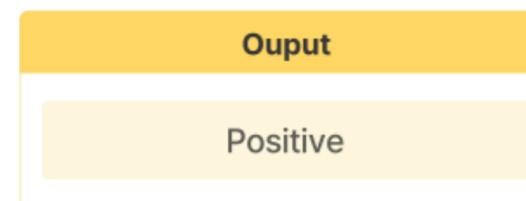
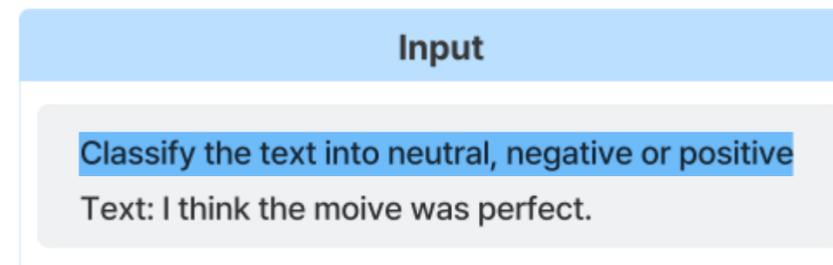
등장 배경 2

Zero-shot 성능 향상

- NLP task를 instruction(지시문) 형태로 변환하여 adaptation tuning 진행
- 다양한 task로 instruction 데이터를 구성하여 학습하면 새로운 task에서의 성능이 향상되며 zero-shot 성능 또한 향상됨



Translation



Sentiment Analysis

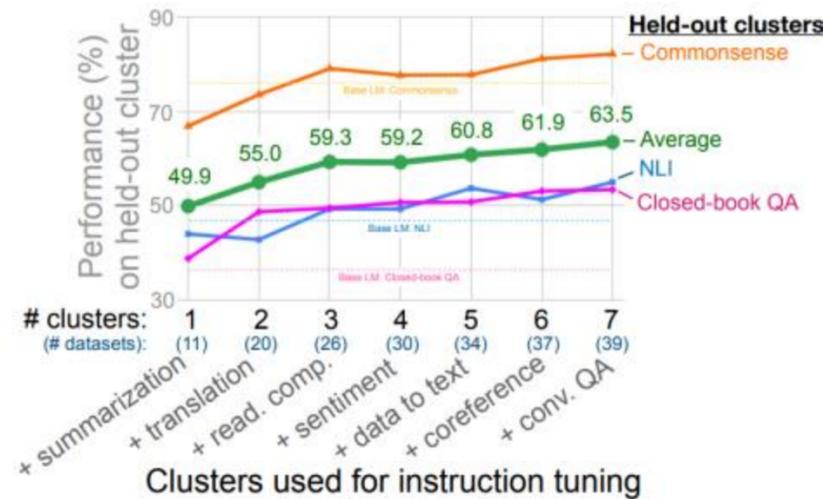
...

Adaptation Tuning

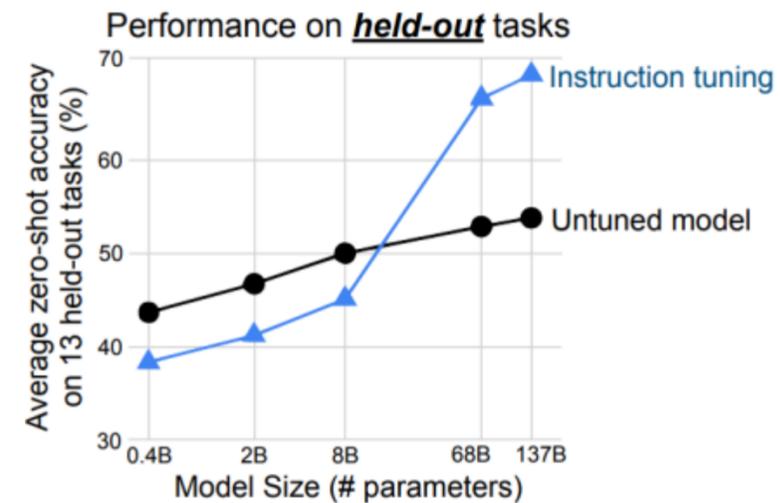
Instruction Tuning

FLAN: FINETUNED LANGUAGE MODELS ARE ZERO-SHOT LEARNERS (2021)

- 기존 LLM의 zero-shot 성능을 향상시킬 수 있는 instruction tuning 기법을 처음으로 제안함
- Instruction tuning에 사용되는 task cluster의 개수가 많아질수록 unseen task에서의 성능이 향상됨
- 크기가 작은 모델에서는 instruction tuning 수행 시 오히려 성능이 하락하며 tuning의 효과를 얻기 위해서는 모델의 크기가 일정 수준 이상이 되어야 함



Task cluster 수에 따른 성능 변화



모델 크기에 따른 성능 변화

Adaptation Tuning

Alignment Tuning

등장 배경

- LLM은 학습 데이터 포함된 거짓 정보, 해로운 표현, 사회적 편향(인종, 성별 등)이 반영되어 인간의 가치와 선호를 고려하지 않는 답변을 생성하기도 함
- 이러한 부적절한 행동을 방지하기 위해 human alignment 기법들이 제안됨
- Pre-train 및 instruction tuning과 달리 alignment는 매우 다른 기준을 고려해야 함
 - ✓ Helpfulness, honesty, harmlessness

Helpfulness

사용자가 과제를 해결하는 데 도움이 되어야 함

Honesty

조작하지 않고 정확한 내용을 사용자에게 제공해야함

Harmlessness

모욕적이거나 차별적인 표현을 생성해서는 안되며, 악의적인 목적으로 요청하는 시도를 감지할 수 있어야 함

Alignment의 기준

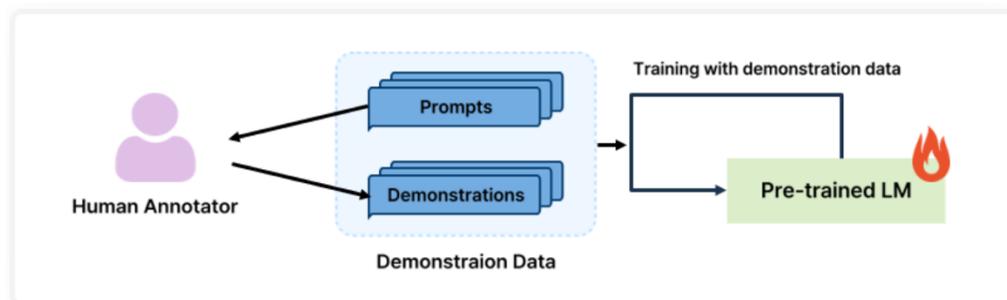
Adaptation Tuning

Alignment Tuning

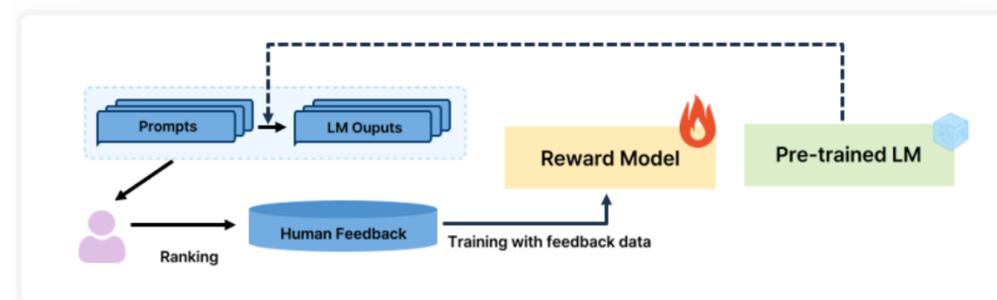
RLHF: Reinforcement Learning with Human feedback

- 강화학습을 이용해 인간의 피드백으로 부터 LLM을 최적화하는 방법
- 총 3단계로 학습이 이루어지며 다음과 같은 요소들이 필요함
 - ✓ Human annotator: LM이 생성한 답변에 대해 인간이 직접 피드백을 주어 선호도를 모델링함
 - ✓ Reward model: LM이 생성한 답변에 대해 인간의 선호도를 예측하는 모델
 - ✓ Aligned model: 인간의 선호도를 반영하여 학습시키고자 하는 모델

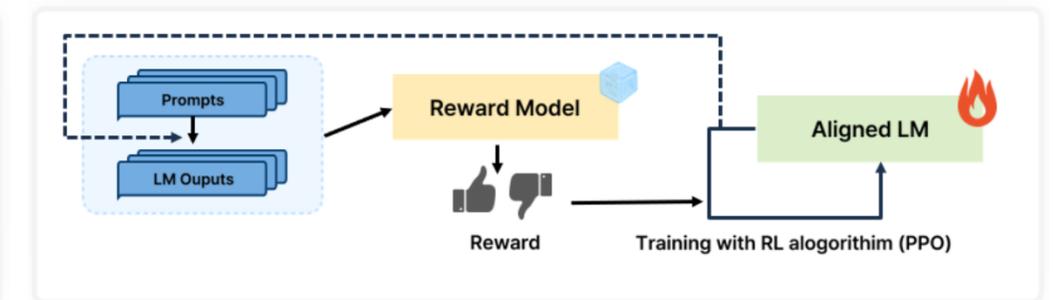
The workflow of the RLHF algorithm



Step 1: Supervised Fine-tuning



Step 2: Reward Model Training



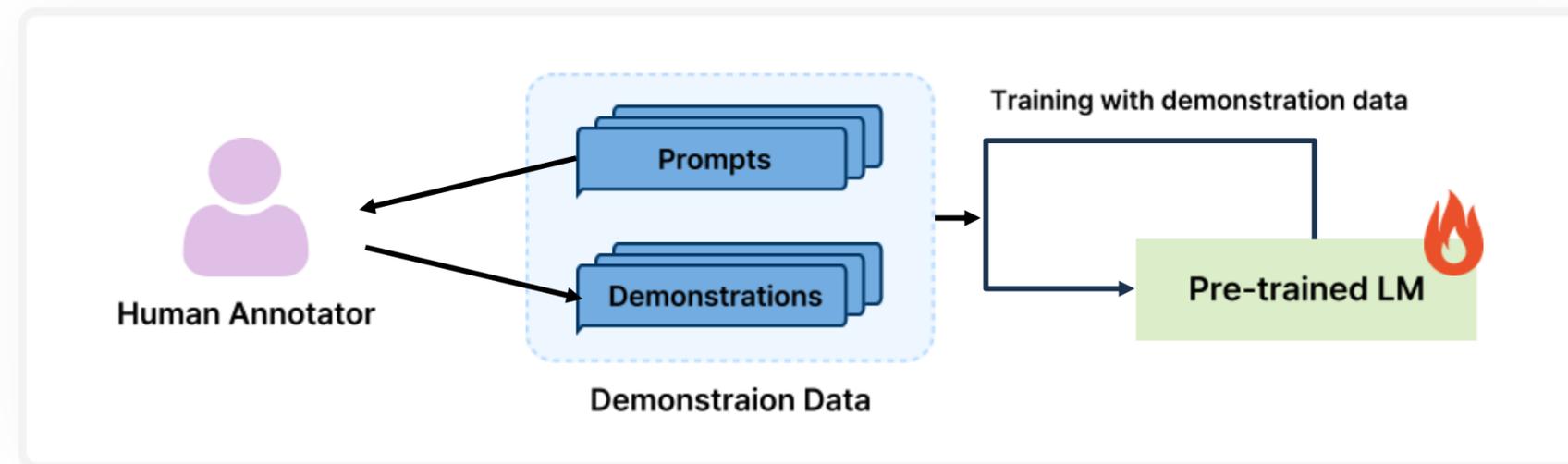
Step 3: RL Fine-tuning

Adaptation Tuning

Alignment Tuning

RLHF: Reinforcement Learning with Human feedback - [Step 1] Supervised Fine-tuning

- 인간이 작성한 prompt 데이터 셋을 이용해 LM을 fine-tuning 함
- 학습이 완료된 LM은 어느 정도 인간의 의도에 맞는 답변을 생성할 수 있게됨



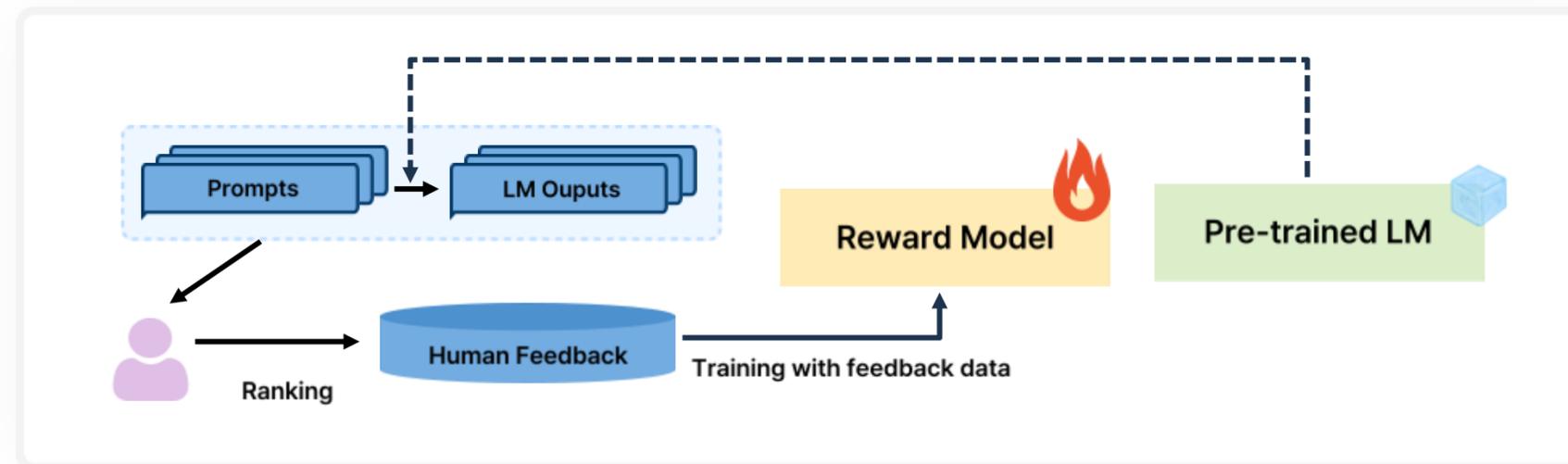
Step 1: Supervised Fine-tuning

Adaptation Tuning

Alignment Tuning

RLHF: Reinforcement Learning with Human feedback - [Step 2] Reward Model Training

- 1단계에서 학습된 LM이 여러 개의 답변을 생성하도록 하고, 각 답변에 대해 인간 평가자는 선호도 순위를 매김
- 답변과 해당 답변에 대응되는 인간 선호도를 통해 인간의 선호도를 모방할 수 있는 reward model을 학습시킴
- Reward model이후 aligned model의 평가를 위해 사용됨



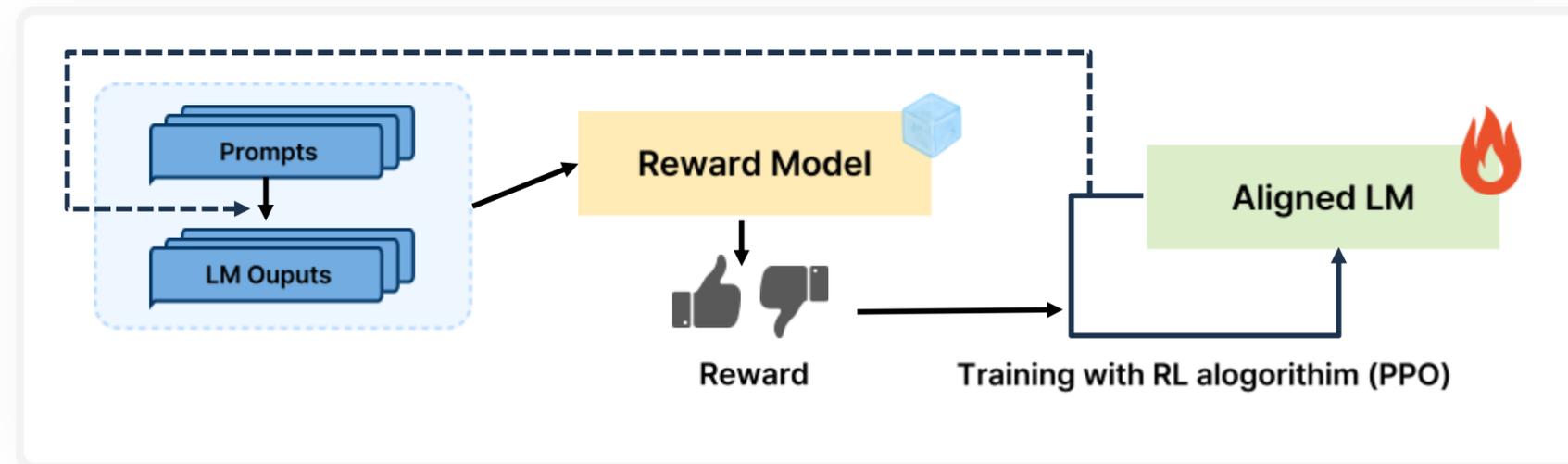
Step 2: Reward Model Training

Adaptation Tuning

Alignment Tuning

RLHF: Reinforcement Learning with Human feedback - [Step 3] RL Fine-tuning

- 강화 학습을 통해 aligned model을 최적화 시킴
- Reward model을 보상 함수로 사용하고 보상을 최대화하기 위해 aligned model의 정책을 fine-tuning함
- 이 과정에서 proximal policy optimization이 사용됨



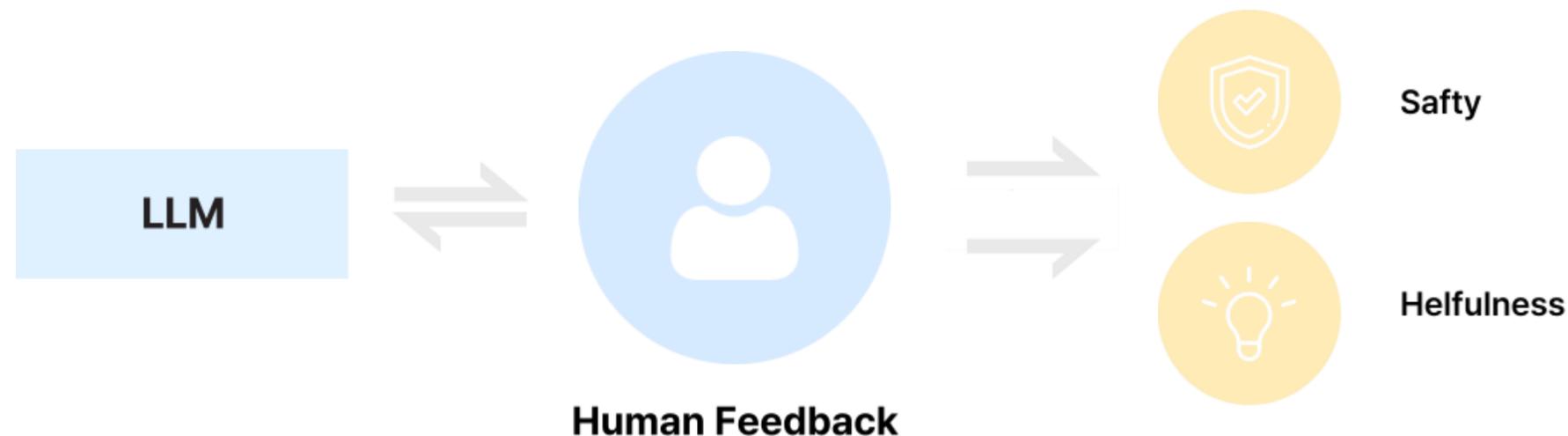
Step 3: RL Fine-tuning

Adaptation Tuning

Alignment Tuning

RLHF: Reinforcement Learning with Human feedback

- RLHF를 통해 사용자의 입력에 대해 안전하고 유용하게 답변하는 모델을 얻을 수 있음
- InstructGPT 및 ChatGPT는 GPT-3에 RLHF가 적용되어 더 나은 품질의 답변을 생성할 수 있게 됨
- 그러나 RLHF는 다음과 같은 한계점들이 존재하여 추가적인 개선이 필요함
 - ✓ 한계점: Human annotator를 고용해 데이터 셋을 확보하기 위한 비용이 많이 필요, 주관적인 요인에 따라 학습 데이터 셋의 품질 편차가 존재

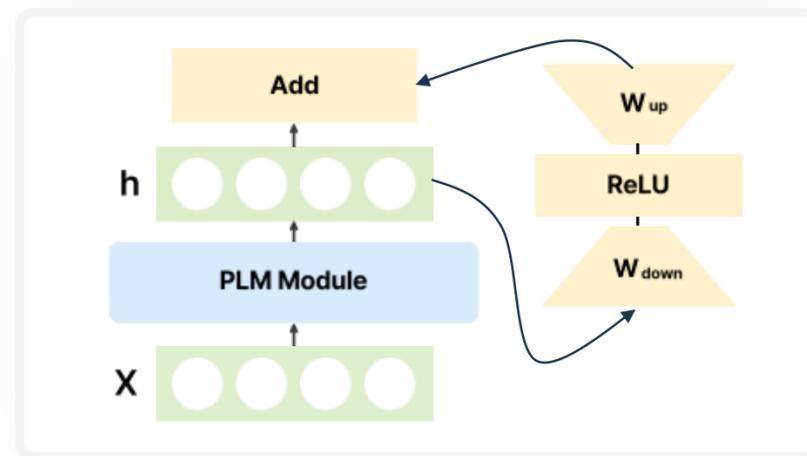


Adaptation Tuning

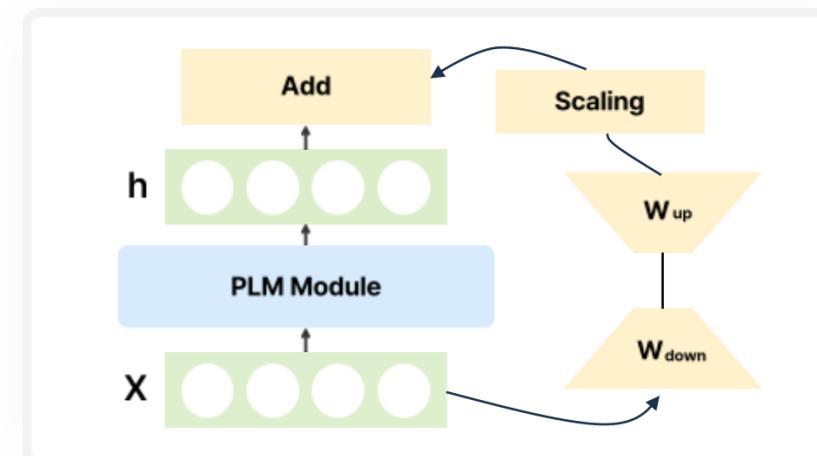
Parameter-Efficient Model Adaptation

등장 배경

- Pre-trained 언어 모델은 모델 파라미터가 매우 많기 때문에 전체 파라미터에 대해 tuning을 수행하는데는 많은 비용이 들게 됨
- PLM의 경우 low intrinsic dimension을 가짐 → 모델이 크고 복잡하지만 실제로 중요한 정보를 나타내는 특성의 수가 상대적으로 적음
 - ✓ RoBERTa는 오직 200개의 학습 가능한 파라미터로 90%의 performance를 달성함
- LLM에서도 효과적인 fine-tuning을 위해 학습 가능한 파라미터 수를 줄이는 parameter-efficient model tuning 기법들이 많이 연구됨



Adapter Tuning



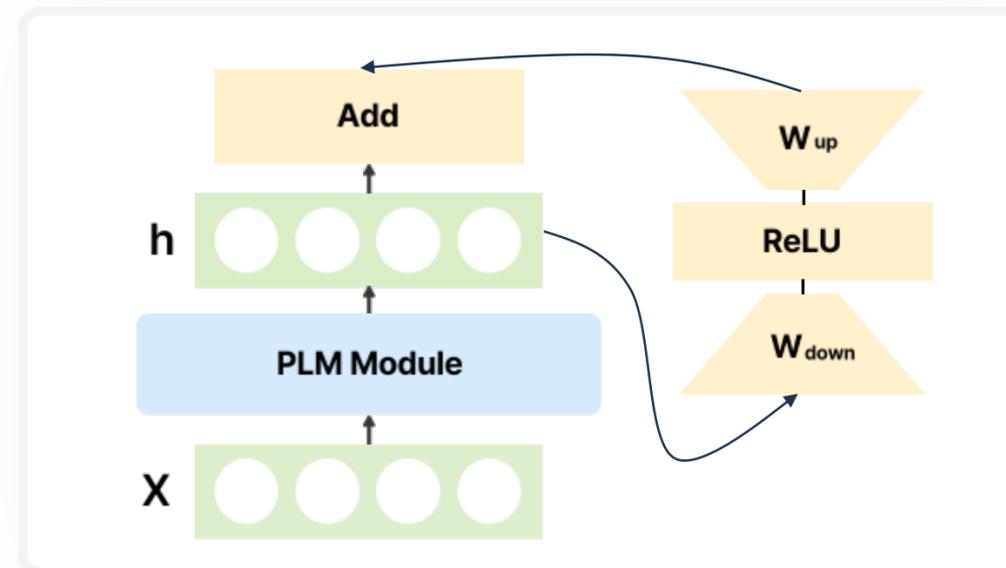
LoRA

Adaptation Tuning

Parameter-Efficient Model Adaptation

Adapter Tuning

- 기존 layer에 위에 adapter layer를 추가함
- 이때 adapter layer는 target task에 따라 최적화되며 기존 LLM의 파라미터는 고정됨
- 이렇게 함으로써 fine-tuning 과정에서 학습 가능한 파라미터의 수를 효과적으로 줄일 수 있음

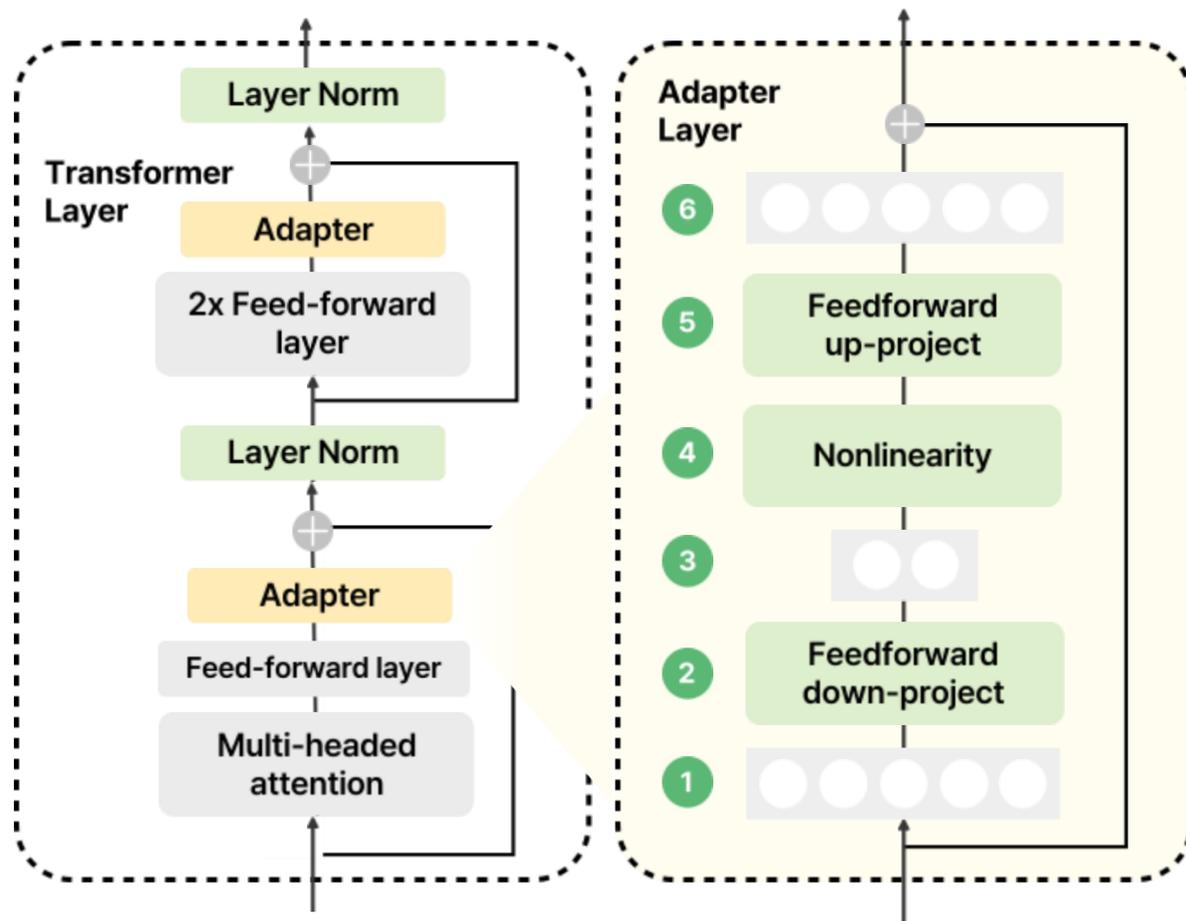


Adapter Tuning

Adaptation Tuning

Parameter-Efficient Model Adaptation

Adapter Tuning



- 6 Output: 1 x d
- 5 Up-projection layer: m x d
- 4 Nonlinearity: Activation function
- 3 Bottleneck layer: 1 x m (m << d)
- 2 Down-projection layer: d x m
- 1 Input: 1 x d

학습 파라미터 수 변화

$$d^2 \rightarrow 2md \quad (m \ll d)$$

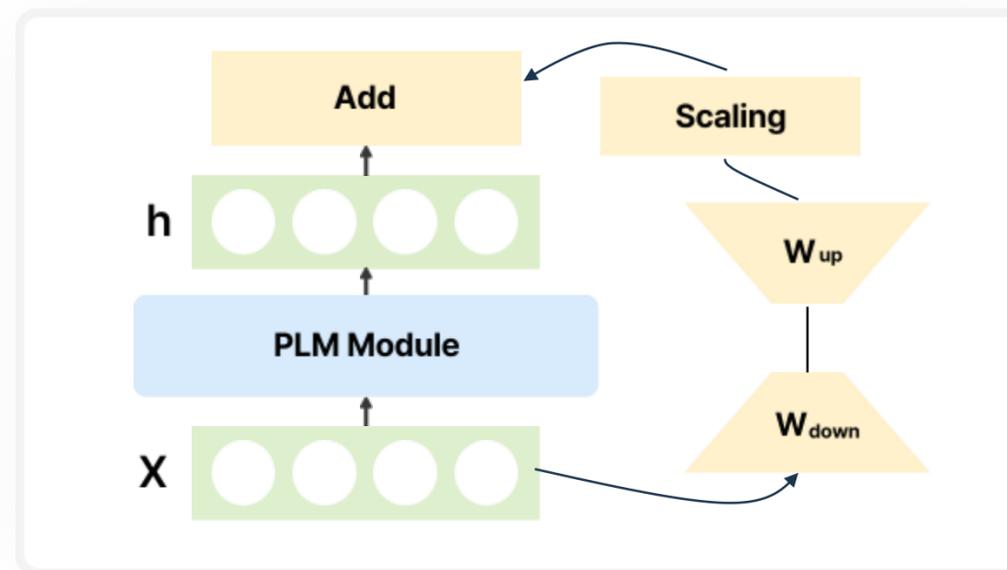
Adaptation Tuning

Parameter-Efficient Model Adaptation

LoRA: Low-Rank Adaptation of Large Language Models

- 가정: Weight update matrix가 low rank intrinsic rank를 가진다
- Pre-trained weight matrix $W_0 \in \mathbb{R}^{d \times d}$ 에 대해 이 행렬에 대한 update를 low-rank decomposition을 통해 아래와 같이 표현함

$$W_0 + \Delta W = W_0 + BA, \quad B \in \mathbb{R}^{d \times m}, A \in \mathbb{R}^{m \times d}, \quad r \ll \min(d, d)$$
- Pre-trained 파라미터는 고정하고 decomposed matrix만 학습

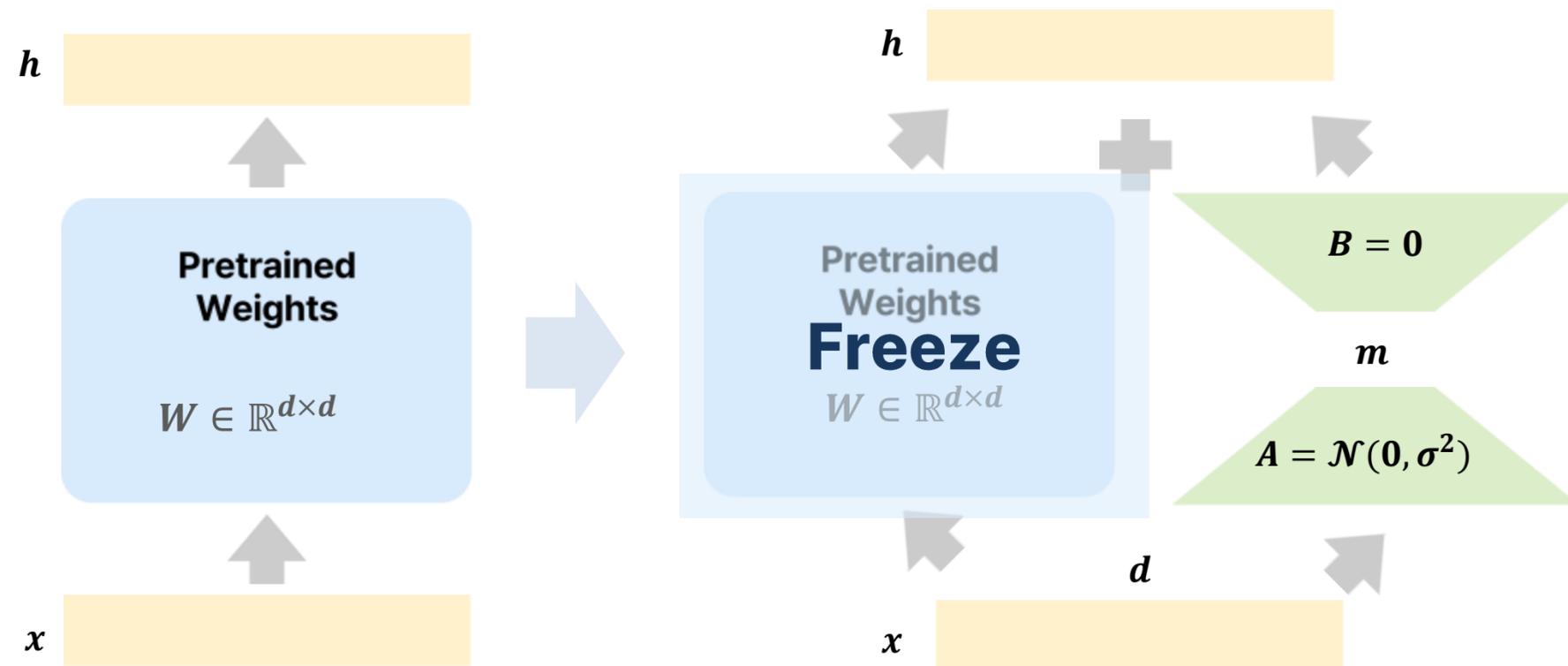


LoRA

Adaptation Tuning

Parameter-Efficient Model Adaptation

LoRA: Low-Rank Adaptation of Large Language Models



$$W_0 + \Delta W = W_0 + BA,$$

$$B \in \mathbb{R}^{d \times m}, A \in \mathbb{R}^{m \times d}, \quad r \ll \min(d, d)$$

학습 파라미터 수 변화

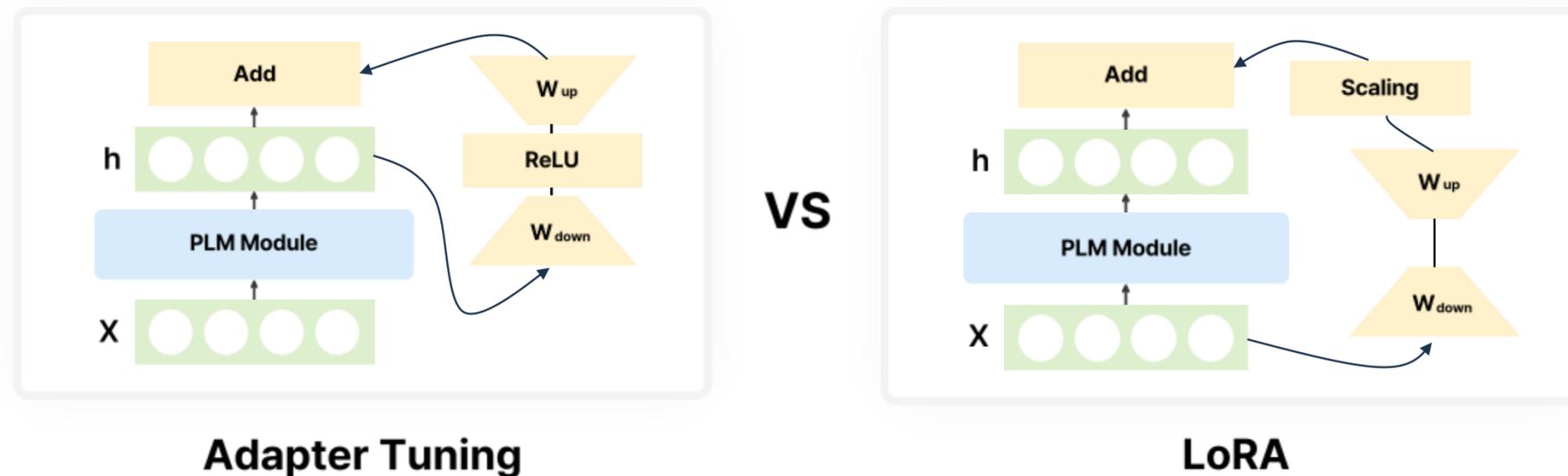
$$d^2 \rightarrow 2md \quad (m \ll d)$$

Adaptation Tuning

Parameter-Efficient Model Adaptation

LoRA: Low-Rank Adaptation of Large Language Models

- 순차적인 방식으로 진행되는 adapter와 달리 병렬적으로 진행이 되어 추가적인 inference latency가 발생하지 않음
- 입력 시퀀스 길이를 줄이지 않고 훈련이 가능함
- 최근 open-source LLM에 많이 적용되어 활용 가능성을 높여주고 있음 (ex. Alpaca-LoRA)



Adaptation Tuning

Parameter-Efficient Model Adaptation

LoRA: Low-Rank Adaptation of Large Language Models

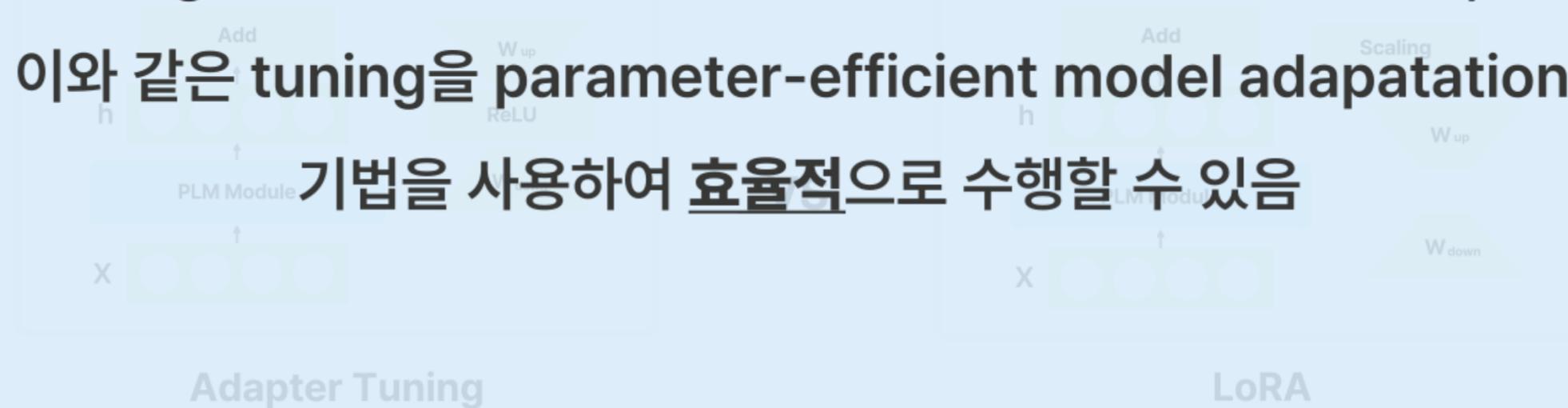
- 순차적인 방식으로 진행되는 adapter와 달리 병렬적으로 진행이 되어 추가적인 inference latency가 발생하지 않음
- 입력 시퀀스 길이를 줄이지 않고 훈련이 가능함
- 최근 open-source LLM에 많이 적용되고 있음 (ex. Alpaca, GPT4o)

Instruction tuning을 통해 LLM의 능력은 강화되고,

Alignment를 통해 인간의 가치와 선호에 맞게 학습이 되며,

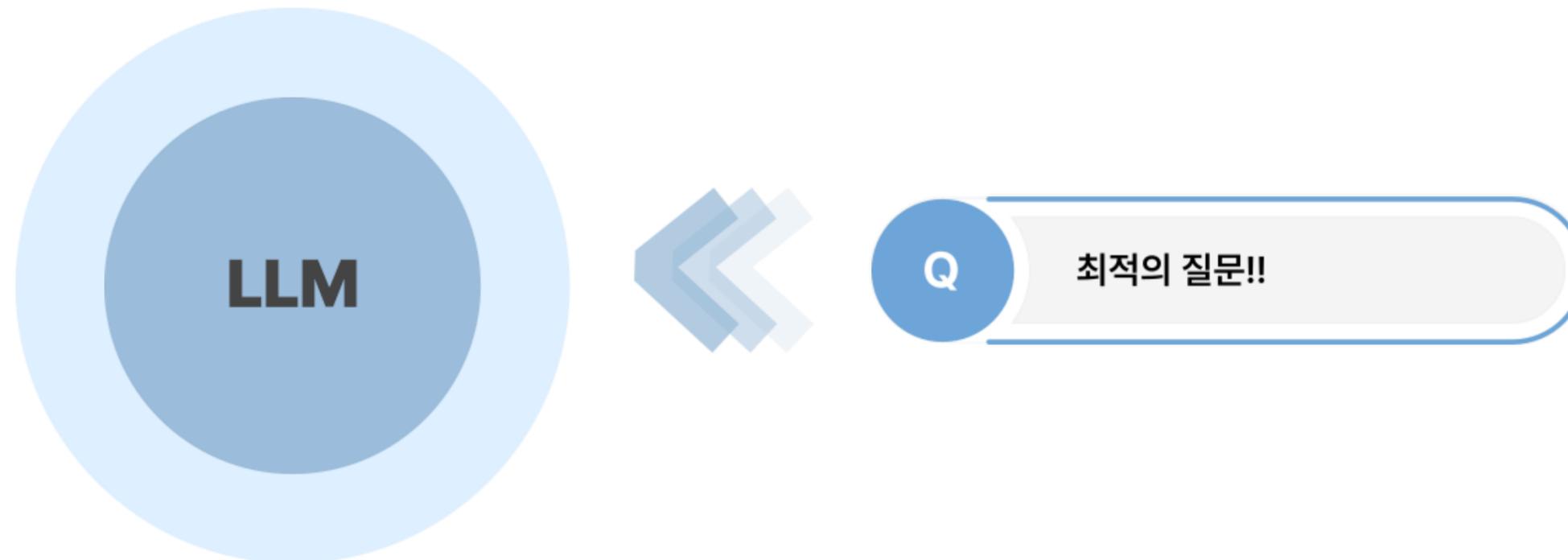
이와 같은 tuning을 parameter-efficient model adaptation

기법을 사용하여 효율적으로 수행할 수 있음



Utilization

Pre-trained 및 adaptation tuning 후의 LLM을 잘 사용하기 위한 주요 방법



Prompt Engineering

질문을 잘하여 LLM을 효과적으로 쓰는 방법

Utilization

Prompt Engineering

- Prompt: LLM으로부터 응답을 생성하기 위한 입력값
- Prompt engineering: LLM이 생성하는 결과물의 품질을 높일 수 있는 입력값들의 조합을 찾는 작업
- LLM의 성능은 prompt의 퀄리티에 의해 크게 좌우되며, 입력값이 조금만 달라져도 완전히 다른 결과물을 생성할 수 있음

01	Instruction	모델이 수행하기를 원하는 특정 task 혹은 지시 사항
02	Context	모델이 보다 더 나은 답변을 하도록 유도하는 외부 정보 또는 추가 내용
03	Input Data	답을 구하고자 하는 것에 대한 질문
04	Output Indicator	결과물의 유형 또는 형식을 나타내는 요소

Prompt의 구성 요소

Utilization

Prompt Engineering

Zero-shot Prompting

- 추가 학습 또는 예제 없이 완수할 task를 주는 것
- 다양하고 다이나믹한 텍스트 생성을 할 수 있게 함

Classify the text into neutral, negative or positive
Text: I think the moive was perfect.



✓ Positive.

Utilization

Prompt Engineering

One/Few-shot Prompting

- 하나 혹은 몇 개의 예제나 템플릿을 기반으로 답변을 생성하게 하는 기법으로, 몇 개의 예제를 보여주는 것만으로도 성능이 향상됨
- GPT-3는 fine-tuning 없이 one/few-shot prompting 만으로도 놀라운 성능을 보여줌
- 그러나 몇 단계의 추가적인 추론 과정을 거쳐야하는 보다 복잡한 문제의 경우 풀 수 없다는 한계점을 지님

Translate English to Korean:

Dog ⇒ 개

Bear ⇒ 곰

Tiger ⇒ 호랑이

Cat ⇒ _____



고양이

Utilization

Prompt Engineering

Chain-of-Thought Prompting

- 답변에 도달하는 과정을 학습시키기 위해 목적으로 본 질문 전에 미리 태스크와 추론 과정을 포함한 답변 예제를 제공하는 것
- PaLM을 학습할 때 reasoning 관련 데이터 셋을 확장시켜 중간 논리를 설명한 부분을 추가한 결과 성능이 향상됨
- LLM의 추론 능력을 향상시켜 복잡한 산술, 상식 및 기호 추론 작업에서도 좋은 성능을 보임

Standard Prompting

Example Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

Example Output

A: The answer is 11.

Chain of thought prompting

Example Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

Example Output

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5+6 = 11$. The answer is 11.

Utilization

Prompt Engineering

Zero-shot Chain-of-Though Prompting

- 트리거 문장 "Let's think step by step (단계별로 생각해보자)"을 프롬프트에 추가해 LLM이 단계에 따라 결과에 도달하게 하는 프레임워크
- 트리거 문장 하나만 추가해도 상세한 답변 도출 및 추론 과정을 출력함

A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

Let's think step by step!



There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls

Training Process of LLMs

Pre-training

LLM의 기본적인 능력을
습득하는 과정

- Data
- Architecture
- Pre-training task
- Optimization and Scalable Training Techniques

Adaptation Tuning

LLM을 특정한 목표에 따라 추가로
adaptation 시키는 과정

- Instruction tuning
- Alinment tuning
- Parameter efficient model adaptation

Utilization

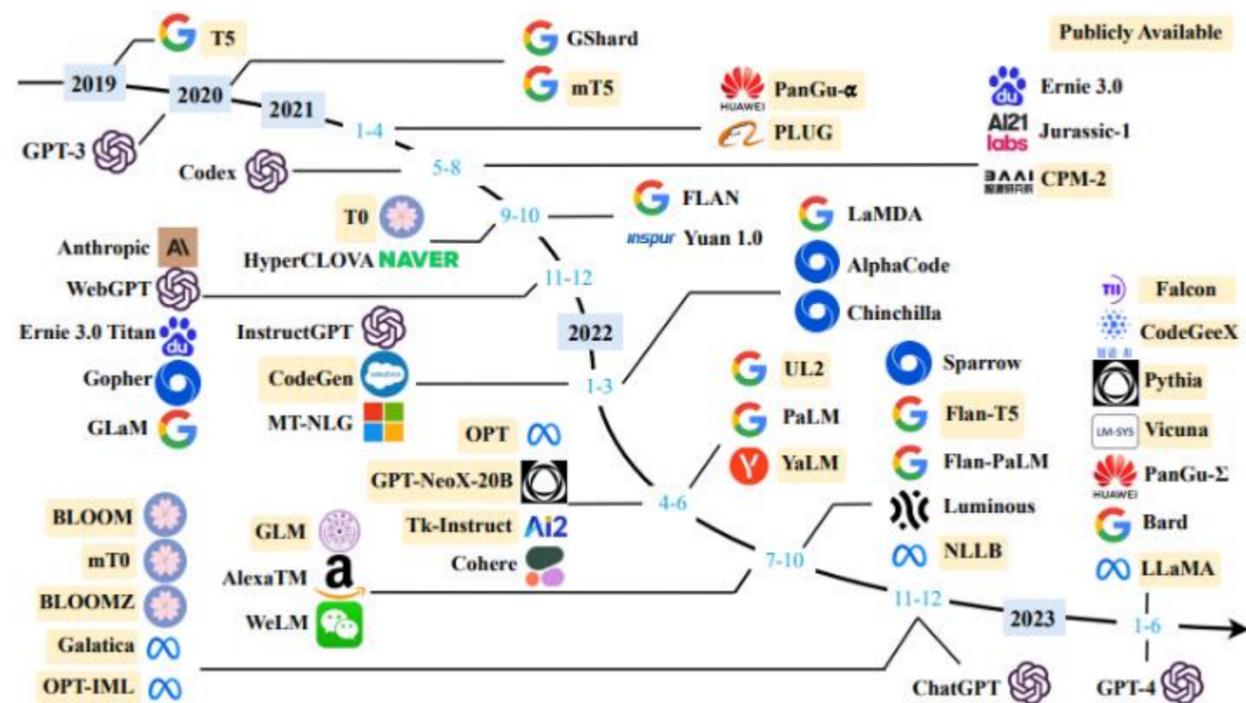
LLM을 효과적으로
사용하는 방법

- Prompt Engineering

③ Research Trend of LLMs

3 Research Trend of LLMs

Research Trend of LLMs



Chapter 1

구조는 단순하게, 스케일은 크게, 데이터는 많이

Chapter 2

게임 체인저 ChatGPT의 등장

Chapter 3

Llama가 이끈 LLM의 민주화

Chapter 4

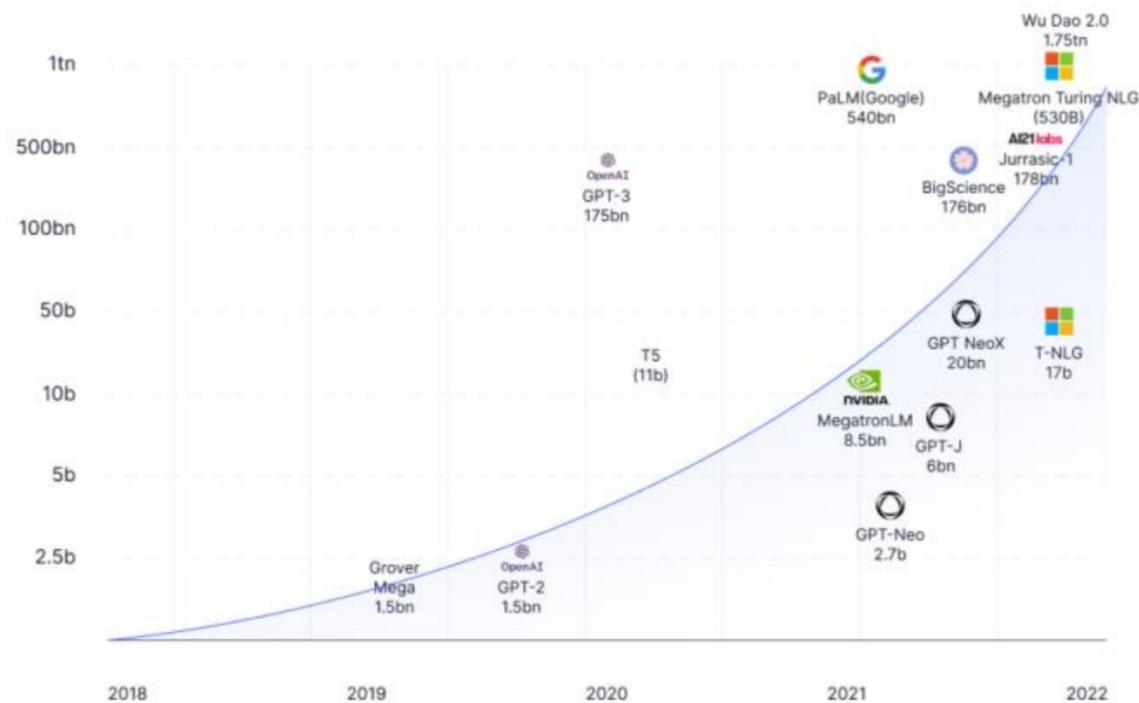
We are LLaMA's Family!

Chapter 5

Open-source LLM의 계속되는 성장

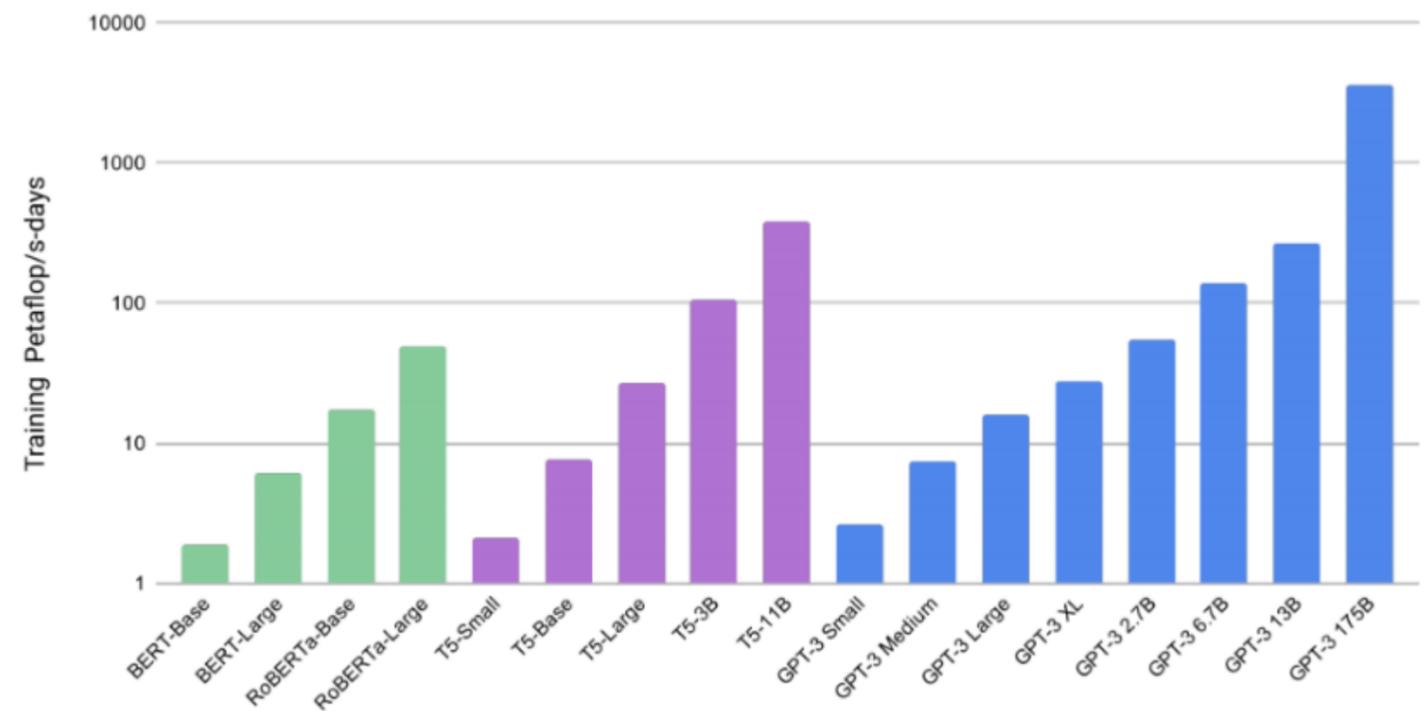
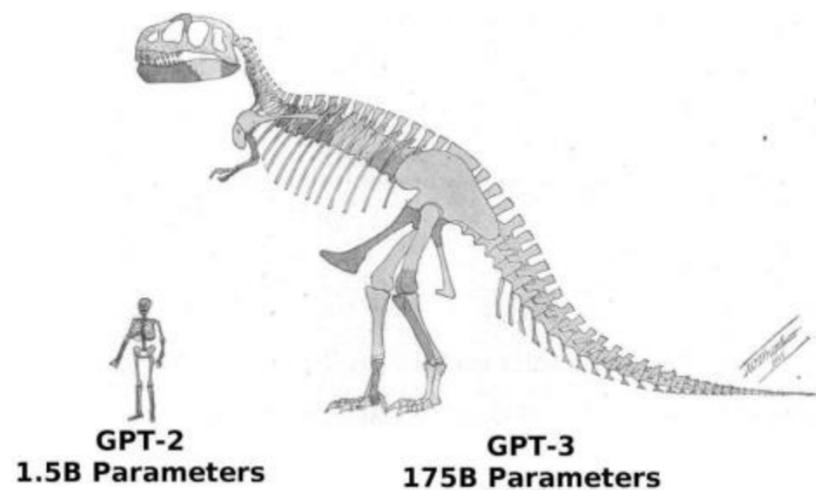
Chapter 1

“ 구조는 단순하게, 스케일은 크게, 데이터는 많이 ”



구조는 단순하게, 스케일은 크게, 데이터는 많이

- 많은 연구들을 통해 transformer 기반의 language model들이 모델의 크기가 커질 수록 성능이 향상됨이 확인됨
- 2020년 5월 175B의 크기를 가진 GPT-3가 등장함
 - ✓ GPT-2 유사한 구조를 가지지만 우수한 성능과 놀라운 능력을 보여주며, 큰 모델을 대량의 데이터 셋에서 학습을 시키면 성능이 향상된다는 것을 경험적으로 증명함
- GPT-3 이후 LLM에 대한 연구가 큰 관심을 받고 많은 모델들이 개발됨



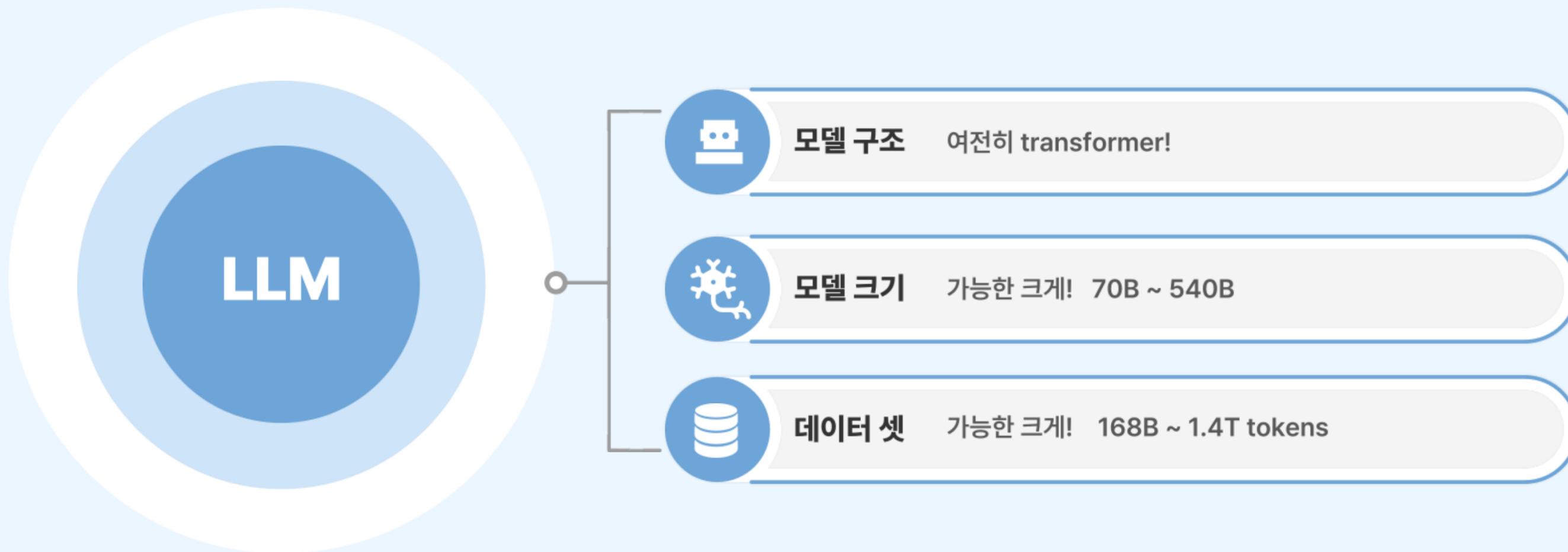
3 Research Trend of LLMs

Chapter 1. 구조는 단순하게, 스케일은 크게, 데이터는 많이

After GPT-3



구조는 단순하게, 스케일은 크게, 데이터는 많이





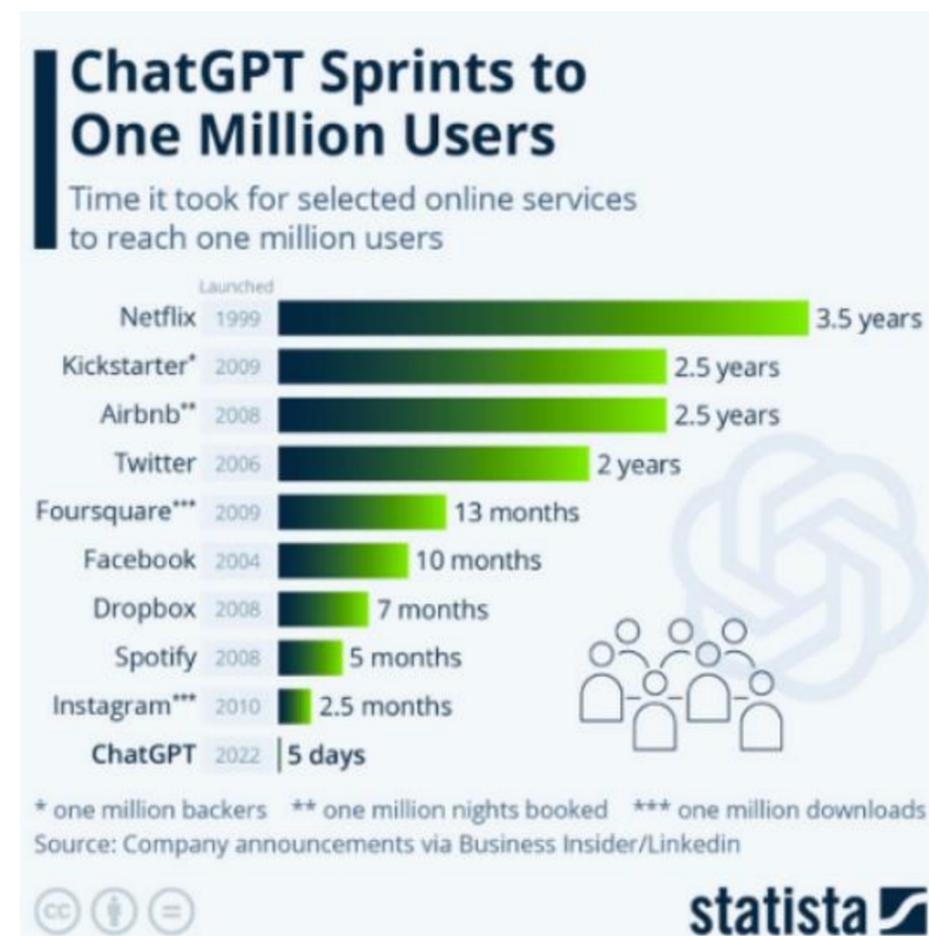
Chapter 2

“ 게임 체인저 ChatGPT의 등장 ”

게임 체인저 ChatGPT의 등장

- ChatGPT의 일 사용자가 공개된지 5일 만에 100만명을, 40일 만에 1천만 명을 넘기며 사회적으로 큰 주목을 받고 있음
- 2016년 알파고의 등장 이후 두 번째 빅 웨이브가 온 것으로 평가를 받기도 함
- 기존 언어 모델과의 차이점
 - 1) 대화형 질문을 이해하고 답변을 생성함
 - 2) 대화 내용을 기억하고 맥락에 맞게 답변함
 - 3) 적절하지 않은 질문은 거르고 답변함
 - 4) 사용자의 추가 질의를 반영해 답변을 개선시킴

ChatGPT의 등장은 인간과 유사한 AI 시스템 탐구에 있어서 중요한 영향을 미쳤으며 AI분야 전체에서도 많은 영감을 주었음



게임 체인저 ChatGPT의 등장

- ChatGPT의 일 사용자가 공개된지 5일 만에 100만명을, 40일 만에 1천만 명을 넘기며 사회적으로 큰 주목을 받고 있음
- 2016년 이후 가장 빠른 성장으로 평가를 받기도 함

종료

What is LLM and ChatGPT?



2023. 07. 28
Data Mining & Quality Analytics Lab.

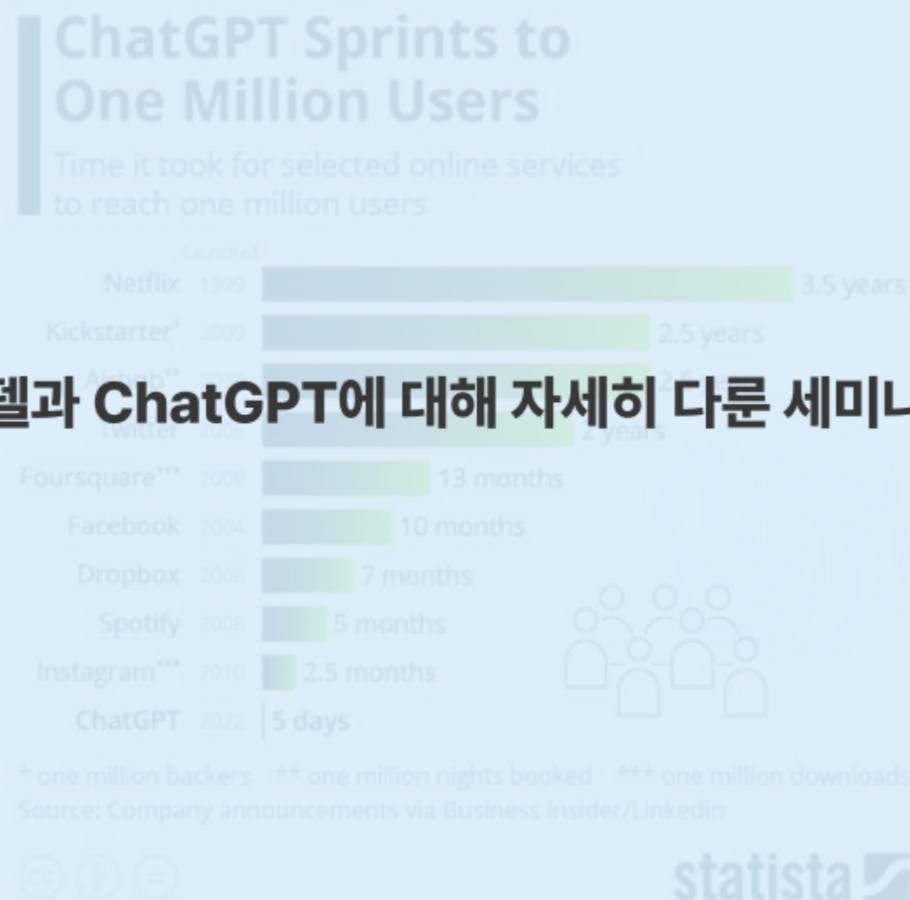
What is LLM and ChatGPT?

발표자:  채고은

📅 2023년 7월 28일
🕒 오후 12시 ~
📺 온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

LLM의 작동 원리 및 GPT series 모델과 ChatGPT에 대해 자세히 다룬 세미나!





Closed-source LLM

How does an LLM research...?

- LLM을 개발한 테크 기업들은 데이터 셋 및 모델 파라미터를 공개하지 않음
- LLM을 개발하기 위해서는 다음과 같은 것들이 필요
 - 1) Foundation model
 - 2) 모델 학습을 위한 서버 인프라
 - 3) 고품질의 데이터 셋



일부 테크 기업들만 보유

학계에서 LLM 연구를 하는 것이 매우 어려움



Chapter 3

“ Llama가 이끄는 LLM의 민주화 ”

LLaMA: Open and Efficient Foundation Language Models (2023)

LLaMA: Open and Efficient Foundation Language Models

Hugo Touvron; Thibaut Lavril; Gautier Izacard; Xavier Martinet
Marie-Anne Lachaux, Timothee Lacroix, Baptiste Rozière, Naman Goyal
Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin
Edouard Grave; Guillaume Lample*

Meta AI

Abstract

We introduce LLaMA, a collection of foundation language models ranging from 7B to 65B parameters. We train our models on trillions of tokens, and show that it is possible to train state-of-the-art models using publicly available datasets exclusively, without resorting to proprietary and inaccessible datasets. In particular, LLaMA-13B outperforms GPT-3 (175B) on most benchmarks, and LLaMA-65B is competitive with the best models, Chinchilla-70B and PaLM-540B. We release all our models to the research community¹.

performance, a smaller one trained longer will ultimately be cheaper at inference. For instance, although Hoffmann et al. (2022) recommends training a 10B model on 200B tokens, we find that the performance of a 7B model continues to improve even after 1T tokens.

The focus of this work is to train a series of language models that achieve the best possible performance at various inference budgets, by training on more tokens than what is typically used. The resulting models, called *LLaMA*, ranges from 7B to 65B parameters with competitive performance

Meta AI

01

연구 목적의 open-source LLM

- 연구자들에게 선별적으로 모델 액세스 권한을 부여

02

작은 크기의 모델 (7B ~ 65B)

- 7B 모델은 Single GPU(V100)에 실행 가능

03

오픈 데이터 셋만을 사용해 학습

- Open-source로 사용 가능하게 하기 위함

Chapter 3. Llama가 이끈 LLM의 민주화

LLaMA: Open and Efficient Foundation Language Models

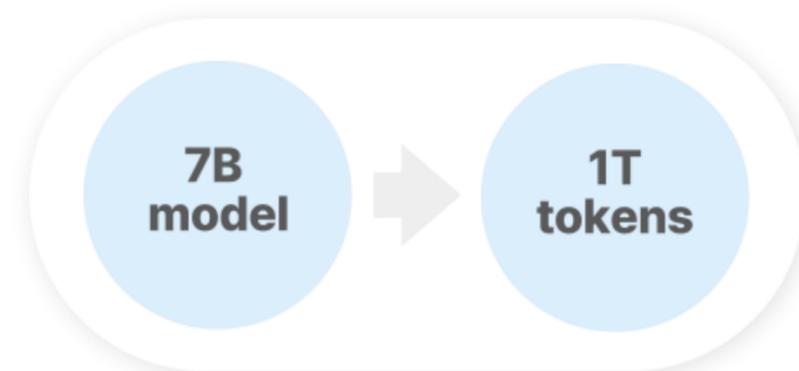
Motivation

- 기존 LLM들은 추론 시 compute budget을 고려하지 않고 오직 학습 성능에만 초점을 맞춤
 - ✓ 모델 파라미터를 늘려 모델 성능을 높였지만 추론 compute budget의 한계로 인해 현실적인 추론 서비스는 어려움 → 반쪽짜리 모델
- 추론을 위한 컴퓨팅 비용은 모델 파라미터 수에 비례하기 때문에 성능은 유지하면서 모델의 크기를 줄이는 것이 필요
- 동일한 compute budget 내에서 **더 작은 모델을 더 많은 token으로 학습시켜 학습 성능은 유지하면서 추론 컴퓨팅 비용을 낮추는 방법을 제시함**

Chinchilla



LLaMA



Chapter 3. Llama가 이끈 LLM의 민주화

LLaMA: Open and Efficient Foundation Language Models

Architecture

- 기존 LLM들과 크게 다르지 않음
 - ✓ Pre-normalization [GPT-3]: 학습 안정성을 높이기 위해 transformer sub-layer의 output이 아닌 input을 normalizing
 - ✓ SwiGLU activation function [PaLM]: ReLU 대신 SwiGLU 사용
 - ✓ Rotary Embedding [GPTNeo]: Absolute positional embeddings을 사용하지 않고 rotary positional embeddings 사용

Efficient Implementation

- 메모리 사용량과 계산량을 절감하기 위해 multi-head attention을 효율적으로 구현한 xformers 라이브러리를 사용함
- Backward pass 시 재계산되는 activation을 줄이기 위해 linear layer 등 계산 비용이 큰 activation을 저장해두고 재사용함
- 65B 모델을 1.4T tokens으로 학습시키기 위해 A100(80GB) GPU 2,048개를 사용해 약 21일 걸림

Chapter 3. Llama가 이끈 LLM의 민주화

LLaMA: Open and Efficient Foundation Language Models

Results

- Zero-shot setting에서 LLaMA-65B는 PaLM(540B)보다 더 좋은 성능, LLaMA-7B 모델은 GPT-3(175B)보다 더 좋은 성능을 보임
- Closed-book QA등의 여러 task에서도 LLaMA-65B가 가장 좋은 성능, LLaMA-7B는 상대적으로 훨씬 더 큰 모델보다 더 좋은 성능을 보임
- 아주 적은 양의 instruction 데이터 셋으로 tuning을 수행한 경우 성능이 크게 향상됨 (on MMLU)

		BoolQ	PIQA	SIQA	HellaSwag	WinoGrande	ARC-e	ARC-c	OBQA
GPT-3	175B	60.5	81.0	-	78.9	70.2	68.8	51.4	57.6
Gopher	280B	79.3	81.8	50.6	79.2	70.1	-	-	-
Chinchilla	70B	83.7	81.8	51.3	80.8	74.9	-	-	-
PaLM	62B	84.8	80.5	-	79.7	77.0	75.2	52.5	50.4
PaLM-cont	62B	83.9	81.4	-	80.6	77.0	-	-	-
PaLM	540B	88.0	82.3	-	83.4	81.1	76.6	53.0	53.4
	7B	76.5	79.8	48.9	76.1	70.1	72.8	47.6	57.2
LLaMA	13B	78.1	80.1	50.4	79.2	73.0	74.8	52.7	56.4
	33B	83.1	82.3	50.4	82.8	76.0	80.0	57.8	58.6
	65B	85.3	82.8	52.3	84.2	77.0	78.9	56.0	60.2

Zero-shot performance on Common Sense Reasoning tasks

OPT	30B	26.1
GLM	120B	44.8
PaLM	62B	55.1
PaLM-cont	62B	62.8
Chinchilla	70B	67.5
LLaMA	65B	63.4
OPT-IML-Max	30B	43.2
Flan-T5-XXL	11B	55.1
Flan-PaLM	62B	59.6
Flan-PaLM-cont	62B	66.1
LLaMA-I	65B	68.9

Instruction finetuning – MMLU (5-shot)

Chapter 3. Llama가 이끈 LLM의 민주화

LLaMA: Open and Efficient Foundation Language Models

Future work

- LLaMA는 인간의 선호도가 반영된 학습을 하지 않았기 때문에 사용자의 입력에 안전하고 유용하게 반응하지 못함
→ 추가적인 학습 방법론 개발이 필요
- 향후 연구 방향으로 instruction tuning과 RLHF로 훈련시킨 모델에 관한 연구를 제시함



LLaMA is a Small Giant Model !!

Chapter 4



“ We are LLaMA's Family! ”

Chapter 4. We are LLaMA's Family!

Alpaca: A Strong, Replicable Instruction-Following Model (2023)



Llama
(2023.02.25)



Alpaca
(2023.03.13)

01

LLaMA-7B 모델을 instruction-tuning 한 모델

- Instruction-tuning된 연구 목적의 LLM

02

Self-instruct를 활용해 대량의 다양한 데이터 생성

- GPT-3 (text-davinci-003)를 사용한 데이터 생성

03

모델 학습 및 데이터 생성 코드 공개

- 연구 확산에 기여

Chapter 4. We are LLaMA's Family!

Alpaca: A Strong, Replicable Instruction-Following Model

Motivation

- 제한된 자원 내에서 고품질의 instruction-following model을 구축하기 위해 필요한 것

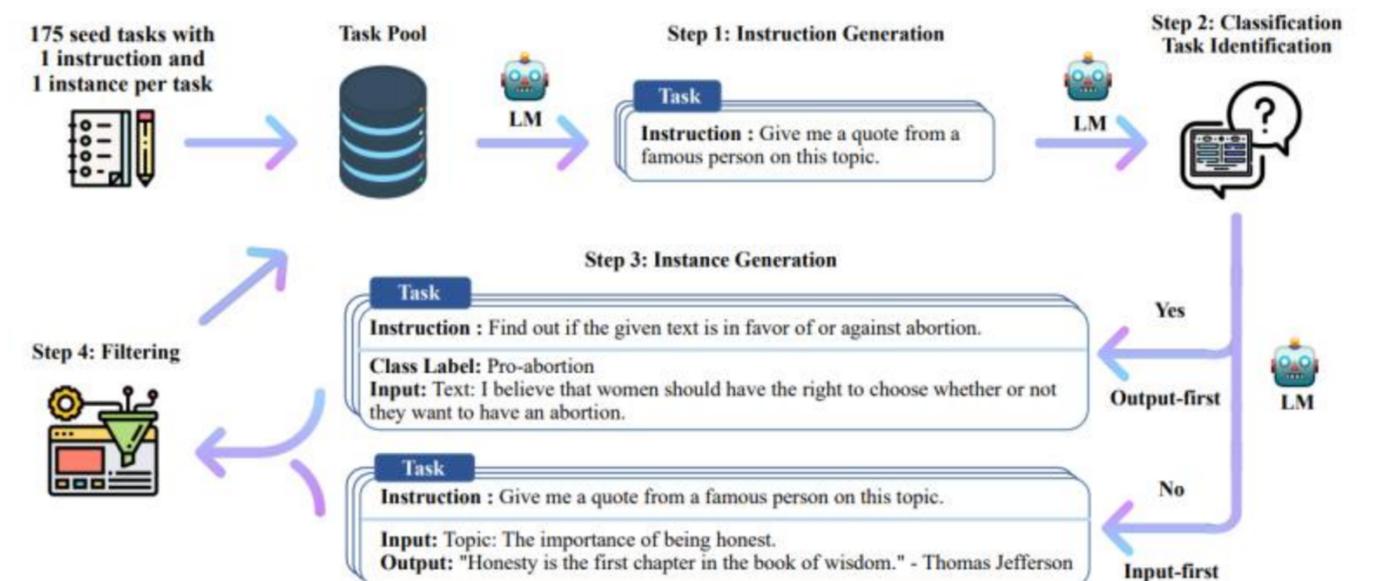


Chapter 4. We are LLaMA's Family!

Alpaca: A Strong, Replicable Instruction-Following Model

SELF-INSTRUCT

- LLM을 활용해 대량의 instruction 데이터 셋을 생성하는 방법
- Self-instruct 프로세스
 - 1) 사람이 작성한 instruction seed로 GPT-3 (text-davinci-003)을 사용해 새로운 instruction 세트 생성
 - 2) 생성된 결과 중 기존 task와 유사하거나 품질이 낮은 세트를 필터링 한 후 이를 task pool에 추가
 - 3) 이 과정을 반복하여 다양하고 많은 양의 instruction 데이터 셋 구축
- \$500 달러 이하의 비용으로 instruction 데이터 셋을 구축함

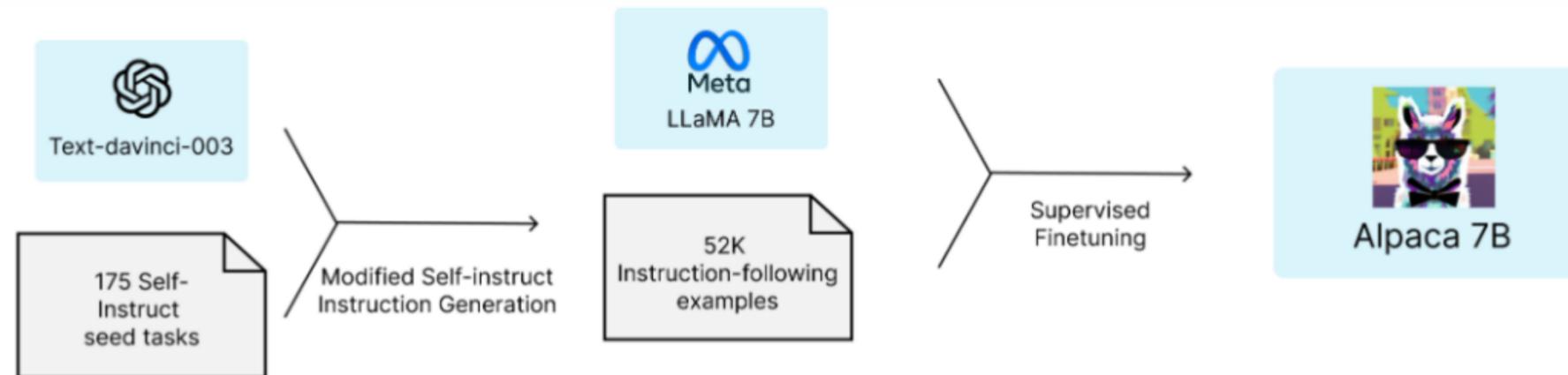


Chapter 4. We are LLaMA's Family!

Alpaca: A Strong, Replicable Instruction-Following Model

Training & Results

- 공개된 LLaMA 학습 코드를 활용해 self-instruct 기법으로 생성한 52k의 instruction 데이터 셋으로 LLaMA-7B 모델을 supervised fine-tuning하여 Alpaca 7B 모델 구축
- A100(80GB) GPU 8개를 사용해 약 3시간 학습하였으며, 약 \$100 이하의 비용이 사용됨
- GPT-3.5(text-davinci-003)과 blind 실험을 수행한 결과 Alpaca-7B가 90:89로 GPT-3.5를 근소하게 이김



Alpaca's training recipe

Chapter 4. We are LLaMA's Family!

Alpaca: A Strong, Replicable Instruction-Following Model

Limitation

- Hallucination(환각), toxicity(독성), stereotypes(고정 관념) 등의 LLM에서 일반적으로 발생하는 오류들이 나타남
- Text-davinci-003에 비해 hallucination이 더 심각하게 나타남
- Text-davinci-003의 짧은 출력으로 생성된 instruction 데이터 셋을 사용해 ChatGPT에 비해 더 짧은 답변을 생성함

Enter your instruction and press enter

What is the capital of Tanzania?

Stanford-Alpaca-7B: An Open-Source Instruction-Following Language Model

Dar es Salaam.

탄자니아의 수도를 탄자니아에서 가장 큰 도시인 Dar es Salaam로 잘못 답변함
(Dar es Salaam은 1974년까지 수도였음)

3 Research Trend of LLMs

Chapter 4. We are LLaMA's Family!

Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality (2023)



Llama
(2023.02.25)



Alpaca
(2023.03.13)



Vicuna
(2023.04.01)

01

LLaMA를 fine-tuning한 오픈 소스 챗봇 모델

02

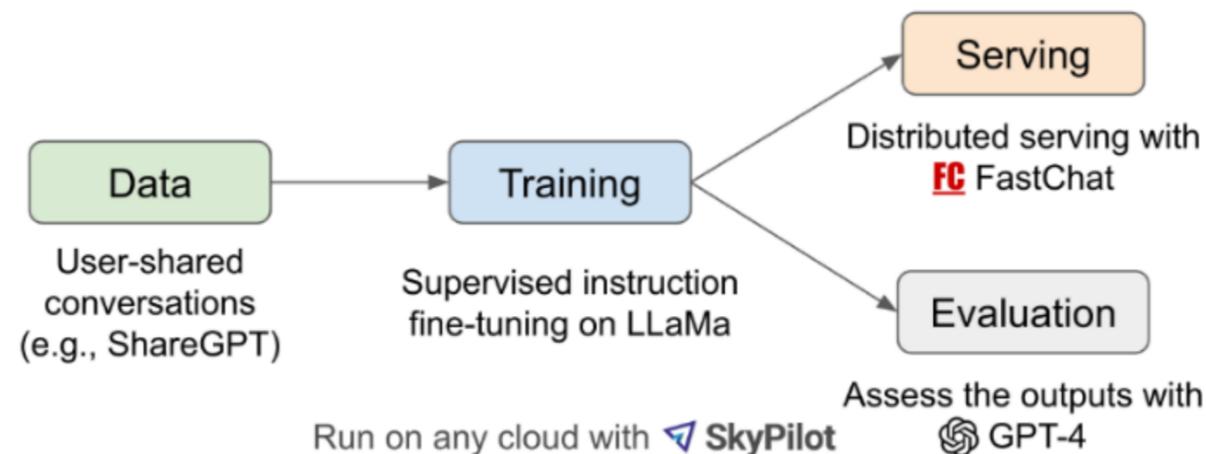
GPT-4를 활용한 새로운 LLM 평가 방법 제안

Chapter 4. We are LLaMA's Family!

Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality

Training

- 사용자들이 ChatGPT 대화를 공유하는 웹사이트 ShareGPT에서 약 70k의 대화를 수집함
- Multi-turn 대화와 긴 시퀀스를 더 잘 처리할 수 있도록 Alpaca에서 제공하는 훈련 스크립트를 개선함
 - ✓ 최대 문장 길이를 512에서 2048로 확장시키고, multi-turn 대화를 위한 손실 함수 개선, 학습 비용 절감 및 메모리 최적화를 위한 기법들 활용함
- A100 GPU 8개로 하루 동안 학습



Vicuna's Workflow

Chapter 4. We are LLaMA's Family!

Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality

Evaluation

- 챗봇 모델이 발전하면서 벤치마크 데이터 셋에서 사람이 성능을 구별하기가 어려워짐
- GPT-4를 사용해 챗봇의 성능을 평가하는 방법을 제안
 - ✓ GPT-4로 질문을 생성하고, 각 답변을 GPT-4를 사용해 10점 만점을 기준으로 평가함 (helpfulness, revlevance, accuracy 등의 기준 적용)

The image shows a comparison of two chatbot responses to a question. On the left, a chatbot (represented by a cat icon) has a score of 7/10. On the right, GPT-4 (represented by a llama icon) has a score of 10/10. The question is: "Compose an engaging travel blog post about a recent trip to Hawaii, highlighting cultural experiences and must-see attractions." The GPT-4 response is marked with a trophy icon, indicating a higher quality.

7/10     10/10

Question: Compose an engaging travel blog post about a recent trip to Hawaii, highlighting cultural experiences and must-see attractions.

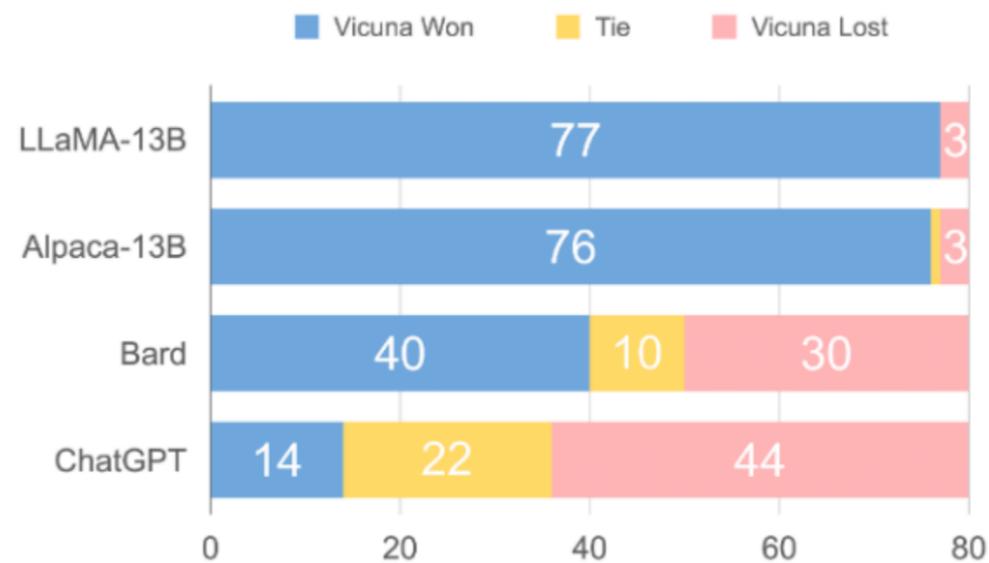
3 Research Trend of LLMs

Chapter 4. We are LLaMA's Family!

Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality

Results

- 챗봇 모델 성능 평가에서 LLaMA와 Alpaca를 크게 능가함
- Bard보다는 약간 높거나 거의 동등한 성능을 보였으며 GhatGPT와도 유사한 성능을 보임(ChatGPT의 90% 수준)



Response Comparison Assessed by GPT-4

Baseline	Baseline Score	Vicuna Score
LLaMA-13B	513.0	694.0
Alpaca-13B	583.0	704.0
Bard	664.0	655.5
ChatGPT	693.0	638.0

Total Scores Assessed by GPT-4

Chapter 4. We are LLaMA's Family!

Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality

Limitation

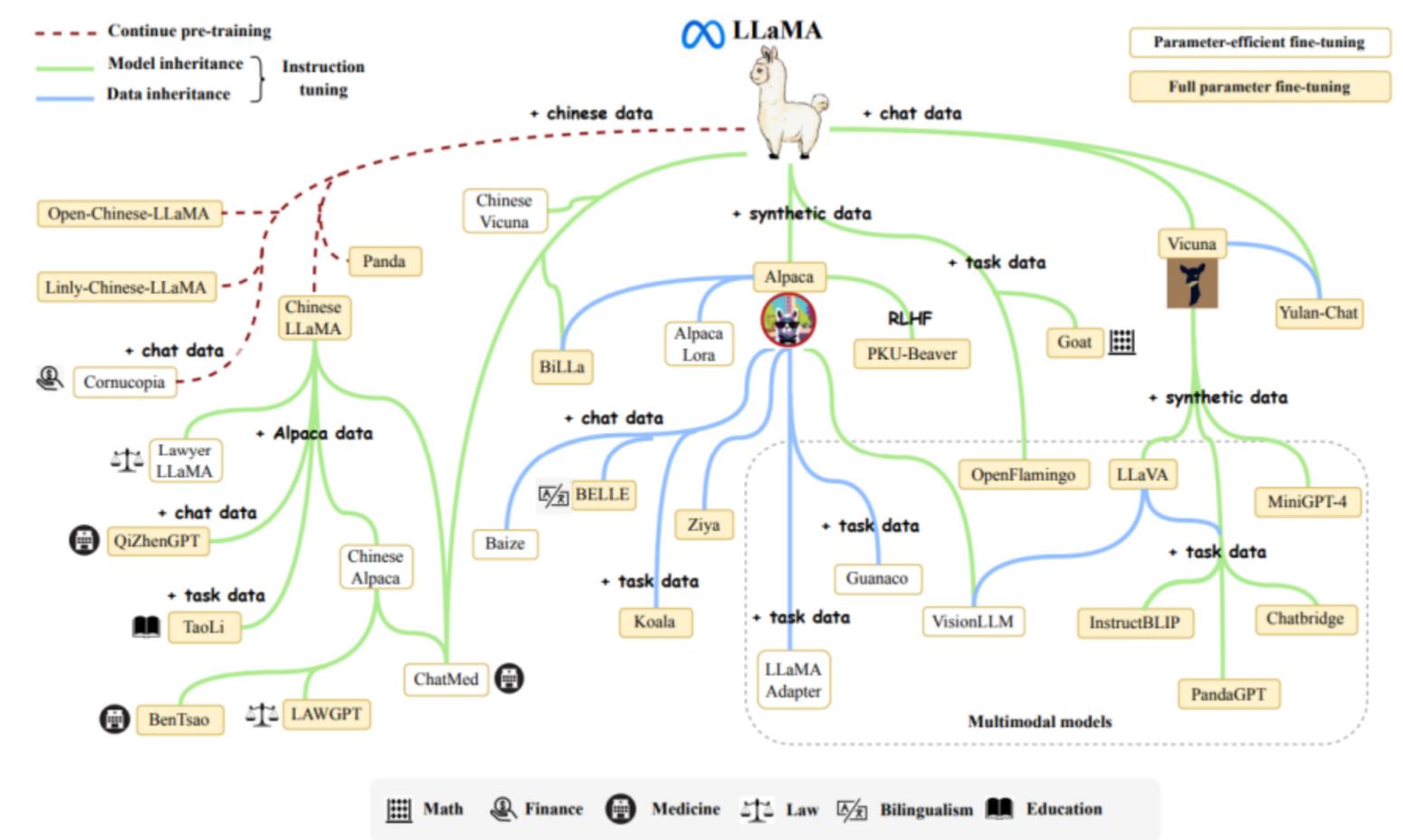
- Reasoning과 mathematics task에서는 좋지 않은 성능을 보임
- Factual accuracy를 보장하는 데 한계가 있으며 safety을 준수하고 toxicity와 bias를 완화하는데 최적화 되지 않음

We are LLaMA's Family!

LLaMA를 기반으로 한 다양한 모델들이 연달아 공개됨

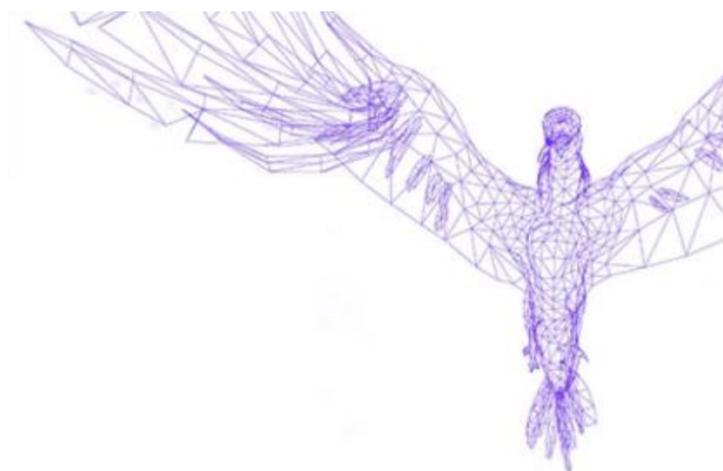
- 다양한 방법으로 LLaMA를 확장해 나감
 - ✓ 학습 데이터 셋의 구성을 변경
 - ✓ Input modality를 확장
 - ✓ 특정 도메인에 최적화

Open-source model이 활발히 공개되고 있음
그러나 아직 closed-source model을 대체하기는 어려움



Chapter 5

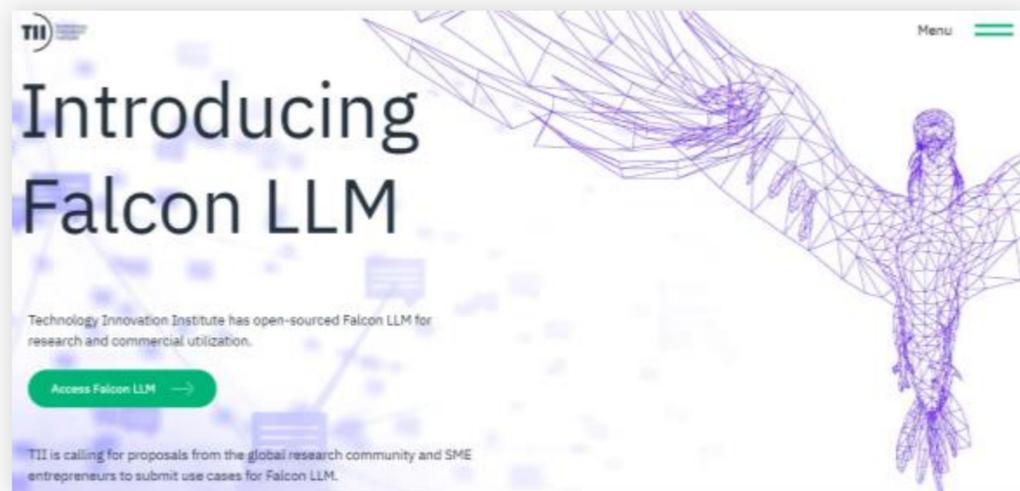
“ Open-source LLM의 계속되는 성장 ”



3 Research Trend of LLMs

Chapter 5: Open-source LLM의 계속되는 성장

Falcon (2023)



Technology Innovation Institute

01

Open-source LLM 중 sota 성능 달성

02

상업적 사용이 가능

- LLaMA, Alpaca 등은 연구 목적으로만 사용이 가능함

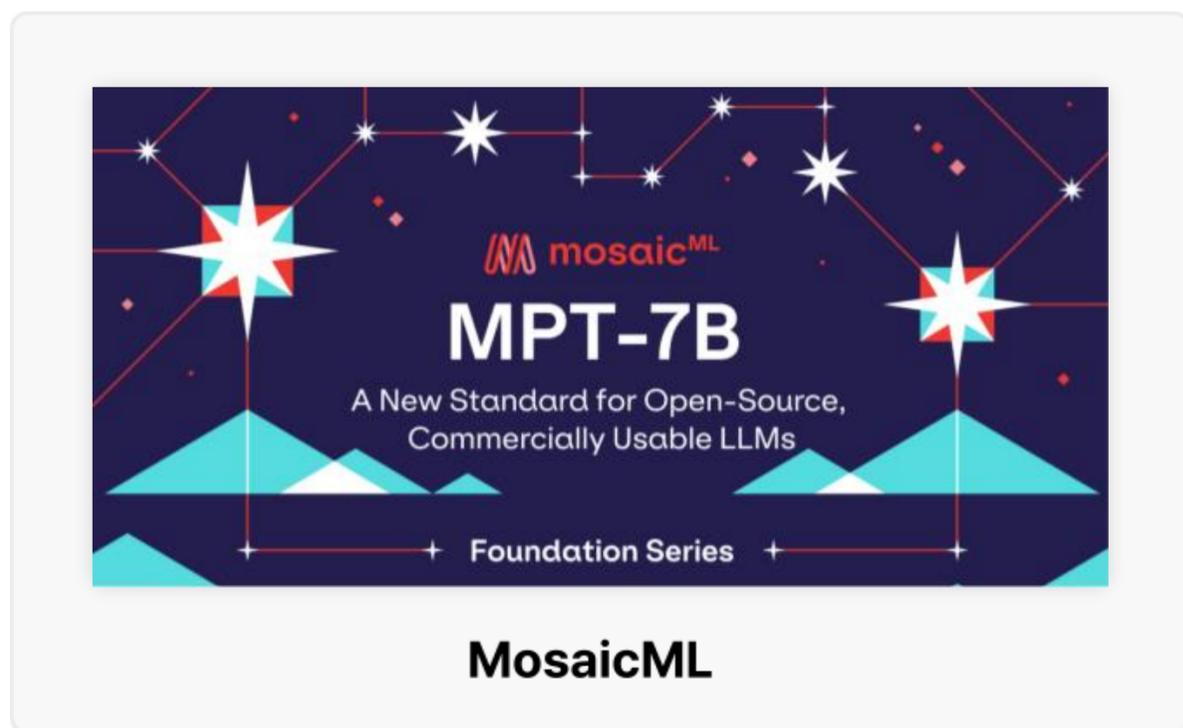
Chapter 5: Open-source LLM의 계속되는 성장

Falcon

- 대규모 데이터 품질에 중점을 두었으며, 웹에서 수집한 데이터에 대해 필터링 및 중복 제거를 수행 → 1,500B token의 사전 훈련 데이터 셋 구축
- 아키텍처의 성능과 효율성 최적화
 - ✓ GPT-3의 학습 컴퓨팅의 75%, Chinchilla의 40%, PaLM-62B의 80%만을 사용
- FlashAttention 및 multiquery 기법을 사용해 추론에 최적화된 구조를 갖췄으며, 추론 속도와 효율성을 향상시킴
- 공개 직후 오픈 소스 LLM 모델 중 sota 성능 달성

Chapter 5: Open-source LLM의 계속되는 성장

MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs (2023)



01

매우 긴 입력 문장 처리 가능

- GPT-4의 2배 길이 문장 입력 가능

02

상업적 사용이 가능

- LLaMA, Alpaca 등은 연구 목적으로만 사용이 가능함

03

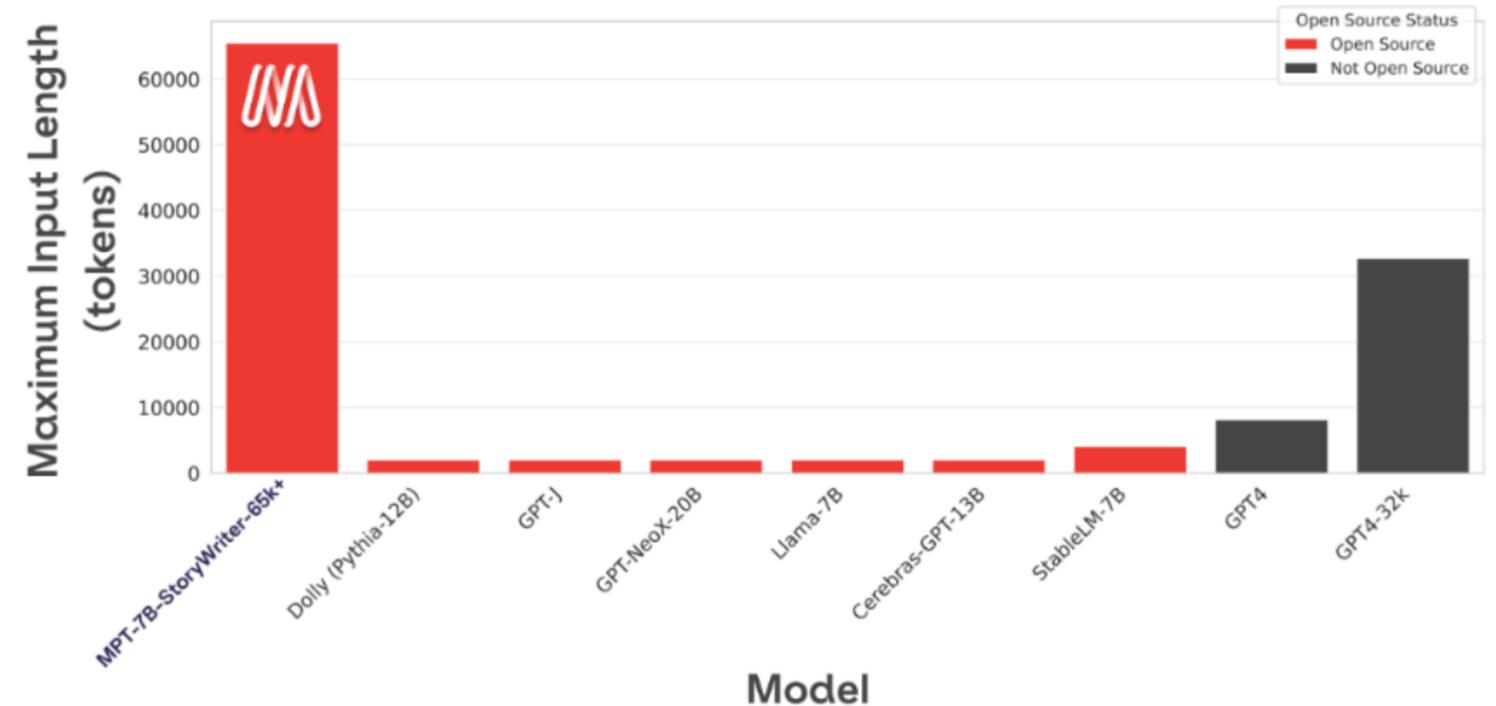
MPT-7B-StoryWriter-65k+

- 매우 긴 길이의 이야기를 읽고 쓰기 위해 설계된 모델

Chapter 5: Open-source LLM의 계속되는 성장

MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs

- Positional embedding을 ALiBi으로 대체하여 매우 긴 입력 문장을 처리할 수 있음 (최대 65k까지 처리 가능)
 - ✓ 대부분의 open-source LLM은 수천 개의 토큰으로 구성된 시퀀스만 처리할 수 있음
- Flash Attention 및 FasterTransformer를 통해 빠른 훈련 및 추론을 최적화
- 텍스트와 코드 등 1T token으로 학습시켰으며 인간의 개입없이 처음부터 끝까지 훈련 시킴
- LLaMA-7B와 비슷한 성능을 보임



Maximum input lengths of difference LLMs

Chapter 5: Open-source LLM의 계속되는 성장

MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs

MPT-7B-StoryWriter-65k+

- 매우 긴 컨텍스트 길이를 갖는 이야기를 읽고 쓰기 위해 설계된 모델
- 이 모델은 books3 데이터 세트의 필터링 된 소설 하위 집합에서 65k token의 컨텍스트 길이로 MPT-7B를 fine-tuning하여 구축
- 전체 텍스트가 68k token 미만의 《위대한 개츠비》를 읽고 에필로그를 생성함 (약 20초 만에 《위대한 개츠비》 전체 텍스트를 처리)

MPT-7B-StoryWriter-65k+
Model Input:
All of *The Great Gatsby*
(67873 tokens)
with 'Epilogue' at the end

The Great Gatsby by F. Scott Fitzgerald
...

And as I sat there brooding on the old, unknown world, I thought of Gatsby's wonder when he first picked out the green light at the end of Daisy's dock. He had come a long way to this blue lawn, and his dream must have seemed so close that he could hardly fail to grasp it. He did not know that it was already behind him, somewhere back in that vast obscurity beyond the city, where the dark fields of the republic rolled on under the night.

Gatsby believed in the green light, the orgastic future that year by year recedes before us. It eluded us then, but that's no matter—tomorrow we will run faster, stretch out our arms further ... And one fine morning—

So we beat on, boats against the current, borne back ceaselessly into the past.

EPILOGUE

Input: 에필로그가 마지막에 있는 위대한 개츠비

MPT-7B-StoryWriter-65k+
Model Output:
Generated Epilogue

I have lived so long in the shadow of this dream that I may never see the bright sunlight of reality. I may look up some morning and find that I am six feet under, and that Gatsby, who gave me the dream and everything else, will have vanished behind me.

And so it was. I had hardly crossed the threshold of the old world when my eye caught the headlines of a newspaper, and I knew at once that my dream was over. I knew that Gatsby was dead.

And I knew that day that I didn't love her.

That night I took the last train for West Egg, and when I reached home I went directly to my attic and brought down all Gatsby's letters.

I have sorted them by date and placed them in the order in which they were written. They are very few in number.

They are dated between April of nineteen eighteen and January of nineteen nineteen, and the last one I have here is the one that told me he was dead.

There is something in that last letter that frightens me. It is the sense of a finality, of a completeness. I never heard of him after that, and I was glad.

But a few months ago I read somewhere that he had been seen in New York. He had become a great man.

And I knew that he had not changed at all.

Output: 생성한 에필로그

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models (2023)



Llama
(2023.02.25)



Llama2
(2023.07.18)

01

Open-source LLM 중 sota 성능 달성

- Falcon, MPT보다 우수한 성능

02

Llama2-Chat: dialog use case에 최적화됨

- SFT, RLHF, GAtt 등의 학습 기법 적용

03

상업적 사용이 가능

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Pre-training

- 기존 Llama1에서 사용한 pre-training 방식을 기반으로 함
- 추가적으로 robust data cleaning 수행, data mix 업데이트, 토큰 수 40% 증가, 컨텍스트 길이 2배 증가, grouped-query attention (GQA) 사용 등을 통해 성능 및 더 큰 모델에 대한 확장성을 개선함

	Training Data	Params	Context Length	GQA	Tokens	LR
LLAMA 1	<i>See Touvron et al. (2023)</i>	7B	2k	×	1.0T	3.0×10^{-4}
		13B	2k	×	1.0T	3.0×10^{-4}
		33B	2k	×	1.4T	1.5×10^{-4}
		65B	2k	×	1.4T	1.5×10^{-4}
LLAMA 2	<i>A new mix of publicly available online data</i>	7B	4k	×	2.0T	3.0×10^{-4}
		13B	4k	×	2.0T	3.0×10^{-4}
		34B	4k	✓	2.0T	1.5×10^{-4}
		70B	4k	✓	2.0T	1.5×10^{-4}

Llama 1 vs Llama 2

3 Research Trend of LLMs

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Pre-trained Model Evaluation

- Open-source LLM 중에서는 가장 좋은 성능을 보임
- Llama2-70B은 몇몇 태스크에서 GPT-3.5, PaLM(540B)와는 유사하거나 더 나은 성능을 보임
- 그러나 GPT-4와 PaLM-2-L 보다는 현저히 낮은 성능을 보임

Model	Size	Code	Commonsense Reasoning	World Knowledge	Reading Comprehension	Math	MMLU	BBH	AGI Eval
MPT	7B	20.5	57.4	41.0	57.5	4.9	26.8	31.0	23.5
	30B	28.9	64.9	50.0	64.7	9.1	46.9	38.0	33.8
Falcon	7B	5.6	56.1	42.8	36.0	4.6	26.2	28.0	21.2
	40B	15.2	69.2	56.7	65.7	12.6	55.4	37.1	37.0
LLAMA 1	7B	14.1	60.8	46.2	58.5	6.95	35.1	30.3	23.9
	13B	18.9	66.1	52.6	62.3	10.9	46.9	37.0	33.9
	33B	26.0	70.0	58.4	67.6	21.4	57.8	39.8	41.7
	65B	30.7	70.7	60.5	68.6	30.8	63.4	43.5	47.6
LLAMA 2	7B	16.8	63.9	48.9	61.3	14.6	45.3	32.6	29.3
	13B	24.5	66.9	55.4	65.8	28.7	54.8	39.4	39.1
	34B	27.8	69.9	58.7	68.0	24.2	62.6	44.1	43.4
	70B	37.5	71.9	63.6	69.4	35.2	68.9	51.2	54.2

Comparison to open-source models

Benchmark (shots)	GPT-3.5	GPT-4	PaLM	PaLM-2-L	LLAMA 2
MMLU (5-shot)	70.0	86.4	69.3	78.3	68.9
TriviaQA (1-shot)	-	-	81.4	86.1	85.0
Natural Questions (1-shot)	-	-	29.3	37.5	33.0
GSM8K (8-shot)	57.1	92.0	56.5	80.7	56.8
HumanEval (0-shot)	48.1	67.0	26.2	-	29.9
BIG-Bench Hard (3-shot)	-	-	52.3	65.7	51.2

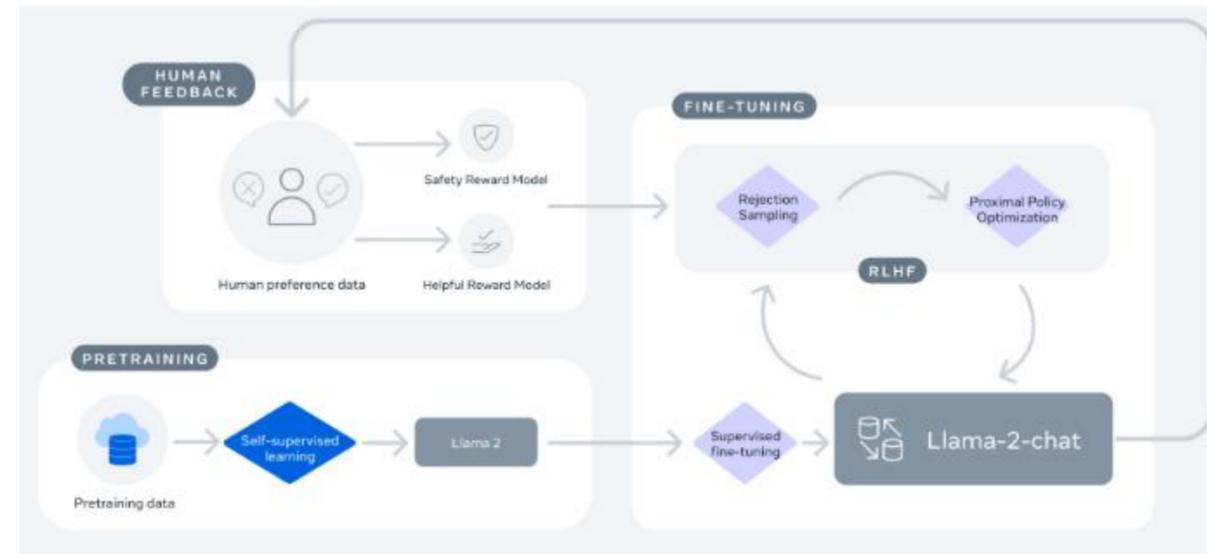
Comparison to closed-source models

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning

- 수 개월의 연구 끝에 완성이 됨
- Supervised fine-tuning(SFT), reward modeling과 RLHF 등의 테크닉과 새롭게 제안하는 Ghost Attention (GAtt)를 적용함
- Helpfulness 및 safety에 대한 인간 평가를 기반으로 closed-source 모델의 대체제로서의 가능성이 높아짐



Training of Llama 2-Chat

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning (1) Supervised Fine-Tuning (SFT)

- 기존의 SFT를 위한 오픈 데이터 셋은 다양성과 품질이 낮음
- Meta에서 고품질의 SFT 데이터 셋을 확보하는 데 중점을 둠 → 고품질의 소량 데이터 셋 확보
- 다소 양이 적더라도 질 좋은 데이터를 사용했을 때 모델의 성능이 훨씬 좋아짐을 확인



Quality Is All You Need!

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning (2) RLHF - Human Preference Data Collection

- Reward modeling을 위해 인간 선호도 데이터를 수집
- Human annotator에게 두 개의 모델 답변 중 더 선호하는 선택하도록 함
- Helpfulness와 safety에 집중함 → 2개의 답변이 주어질 때 2개의 기준을 가지고 각각 평가하도록 함
- 이때 safety에서는 두 답변에 대한 비교 뿐 아니라 각 답변이 안전한지 아닌지를 평가하도록 하여 두 답변 모두 불안전할 경우 해당 데이터를 제외시킴

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning (2) RLHF - Reward Modeling

- 모델의 답변과 해당 prompt를 입력으로 받아 모델 답변에 대한 인간의 선호도를 예측하여 score를 출력하도록 reward model을 학습함
- 이때 helpfulness와 safety에 관한 점수가 상충하는 경우가 많으며, 하나의 reward model이 두 가지를 모두 잘 반영하기 어려움
- 따라서 helpfulness를 위한 reward model과 safety를 위한 reward model를 각각 학습시킴



Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning (2) RLHF - Iterative Fine-Tuning

- 두 가지 알고리즘을 사용하여 RL fine-tuning을 수행함
 - ✓ Rejection Sampling, Proximal Policy Optimization (PPO)
- 인간 선호 데이터 셋 수집 및 reward model을 학습시키면서 RLHF 모델의 연속 버전을 학습하였으며 이를 RLHF-V1 ~ RLHF-V5라고 부름
- RLHF-V4까지는 Rejection Sampling을 사용했으며 그 후에는 두 가지를 순차적으로 결합함 → Rejection Sampling으로 학습된 결과에 PPO를 적용한 후 다시 샘플링 진행

Chapter 5: Open-source LLM의 계속되는 성장

Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning (2) RLHF - Iterative Fine-Tuning

- Rejection Sampling
 - ✓ 각 iteration에서 가장 최근 모델이 prompt마다 K개의 답변을 샘플링함
 - ✓ 그 후 각 샘플에 대해 reward model을 통해 score를 산출한 뒤 가장 좋은 답변을 선택함
 - ✓ Aligned model이 가장 좋은 답변으로 선택됨 샘플을 생성하도록 cross-entropy loss로 학습 → SFT와 동일한 학습 방식
- Proximal Policy Optimization (PPO)
 - ✓ PPO 기반 학습 시 reward model의 점수는 safety model 점수와 helpfulness model 점수 중 하나만 선택되어 사용됨
 - (1) 생성된 답변이 안전하지 않으면 safety model의 점수를 reward로 선택 → 더 안전한 답변을 생성하도록 학습
 - (2) 생성된 답변이 안전하면 helpfulness model의 점수를 reward로 선택 → 더 유용한 답변을 생성하도록 학습

$$R_c(g | p) = \begin{cases} R_s(g | p) & \text{if IS_SAFETY}(p) \text{ or } R_s(g | p) < 0.15 \\ R_h(g | p) & \text{otherwise} \end{cases}$$

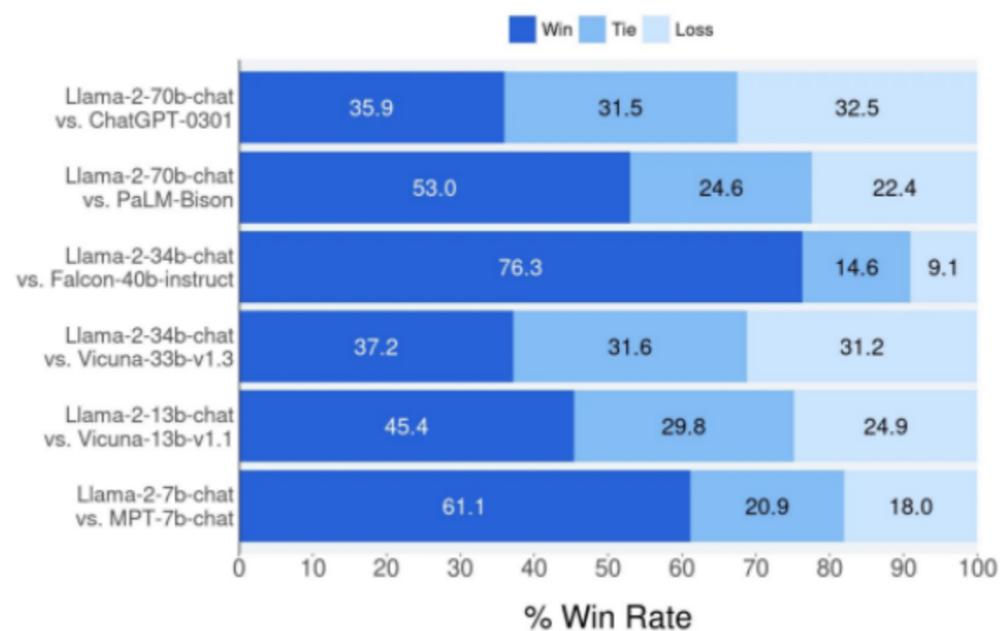
3 Research Trend of LLMs

Chapter 5: Open-source LLM의 계속되는 성장

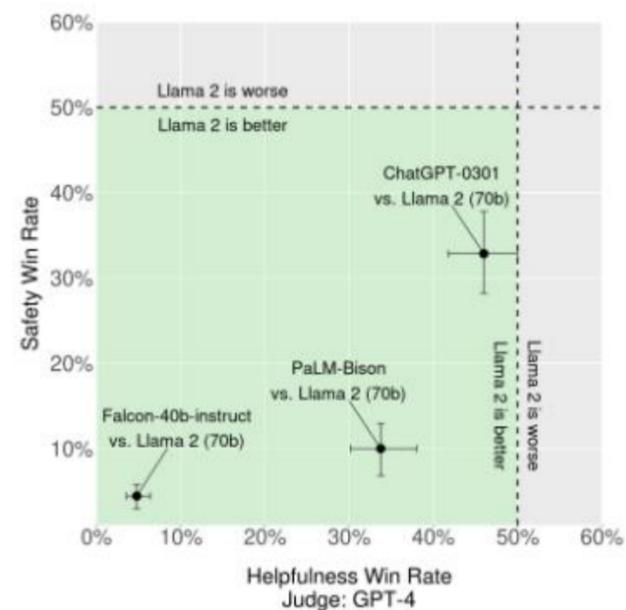
Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning (2) RLHF Results

- Helpfulness human evaluation에서 기존의 open-source LLM과 ChatGPT보다 우수한 성능을 보임
- ChatGPT보다 safety 기준에서 매우 좋은 성능을 보임 → Helpfulness reward model과 safety reward model을 각각 학습시켜 사용한 결과



Helpfulness human evaluation



Win-rate % for helpfulness and safety between commercial-licensed baselines and Llama 2-Chat

Chapter 5: Open-source LLM의 계속되는 성장

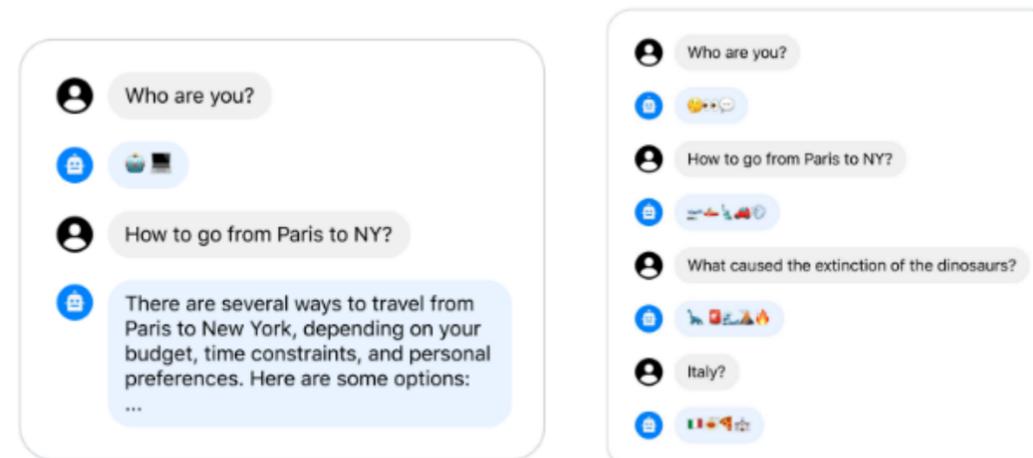
Llama 2: Open Foundation and Fine-Tuned Chat Models

Llama 2-Chat: Fine-tuning (3) System Message for Multi-Turn Consistency - Ghost Attention (GAtt)

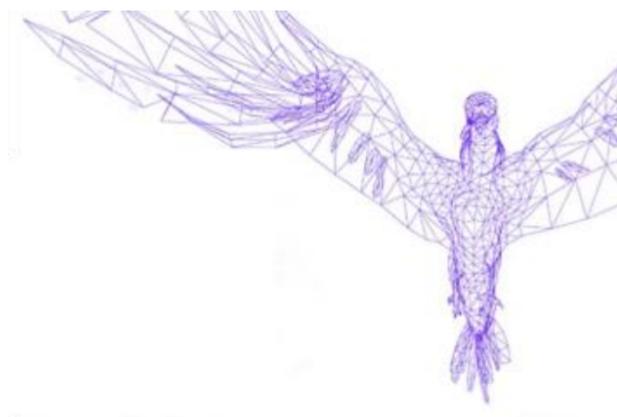
- 기존 모델들은 multi-turn 대화에서 초기 대화를 잊어버리는 문제가 있으며 이를 해결하기 위해 Ghost Attention (GAtt)를 제안함
- GAtt: Multi-turn 대화를 위한 fine-tuning 데이터 셋을 구축하는 방법
 - ✓ 대화를 생성할 때는 초기 대화 혹은 instruction을 모든 턴에 추가 → 입력 문장의 길어지는 문제가 발생
 - ✓ 그 이후에는 초기 턴에서만 유지하고 나머지 턴에서는 제거 혹은 학습 시 초기 턴 외에 나머지 턴에서의 instruction에 대한 loss를 0으로 설정
- 취미(ex: 테니스를 좋아함), 언어(ex: 프랑스어), 인물(ex: 나폴레옹처럼 말하기) 등의 초기에 설정된 제약 사항도 잘 따를 수 있게 됨

초기 대화 혹은 초기 instruction
"이모티콘으로만 대답해줘"

Issues with multi-turn memory (left) can be improved with GAtt (right).



Open-source LLM의 계속되는 성장



여전히 closed-source LLM에 비해 낮은 성능을 보이고 있으나 빠른 속도로 성능이 향상되고 있음
사회에 더 크게 기여하고 연구 속도를 촉진시키기 위해 open-source LLM들이 계속해서 공개되고 있으며, 이에 따라 LLM의 연구는 계속해서 발전하고 있음

Summary

Training Techniques and Research Trends of LLMs

Large language model을 효과적으로 학습하고 활용하기 위한 주요 기법들을 살펴보고,
LLM 연구의 흐름과 함께 최신의 LLM 연구들을 소개함

> Training Techniques of LLMs

- ✓ Pre-Training: LLM의 기본적인 능력을 습득하는 과정
- ✓ Adaptation Tuning: LLM을 특정한 목표에 따라 추가로 adaptation 시키는 과정
- ✓ Utilization: LLM을 효과적으로 사용하는 방법

> Research Trend of LLMs

- ✓ From GPT-3 (2020.05) to Llama2 (2023.07)

고맙습니다:)

References

- [1] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.
- [2] Lou, R., Zhang, K., & Yin, W. (2023). Is prompt all you need? no. A comprehensive and broader view of instruction learning. arXiv preprint arXiv:2303.10475.
- [3] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.
- [4] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv preprint arXiv:2307.09288.
- [5] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. Stanford alpaca: An instruction-following llama model. https://github.com/tatsu-lab/stanford_alpaca, 2023.
- [6] Wang, Y., Kordi, Y., Mishra, S., Liu, A., Smith, N. A., Khashabi, D., & Hajishirzi, H. (2022). Self-instruct: Aligning language model with self generated instructions. arXiv preprint arXiv:2212.10560.
- [7] Penedo, G., Malartic, Q., Hesslow, D., Cojocaru, R., Cappelli, A., Alobeidli, H., ... & Launay, J. (2023). The RefinedWeb dataset for Falcon LLM: outperforming curated corpora with web data, and web data only. arXiv preprint arXiv:2306.01116.
- [8] MosaicML NLP Team et al. (2023). Introducing mpt-7b: A new standard for open-source, commercially usable llms.