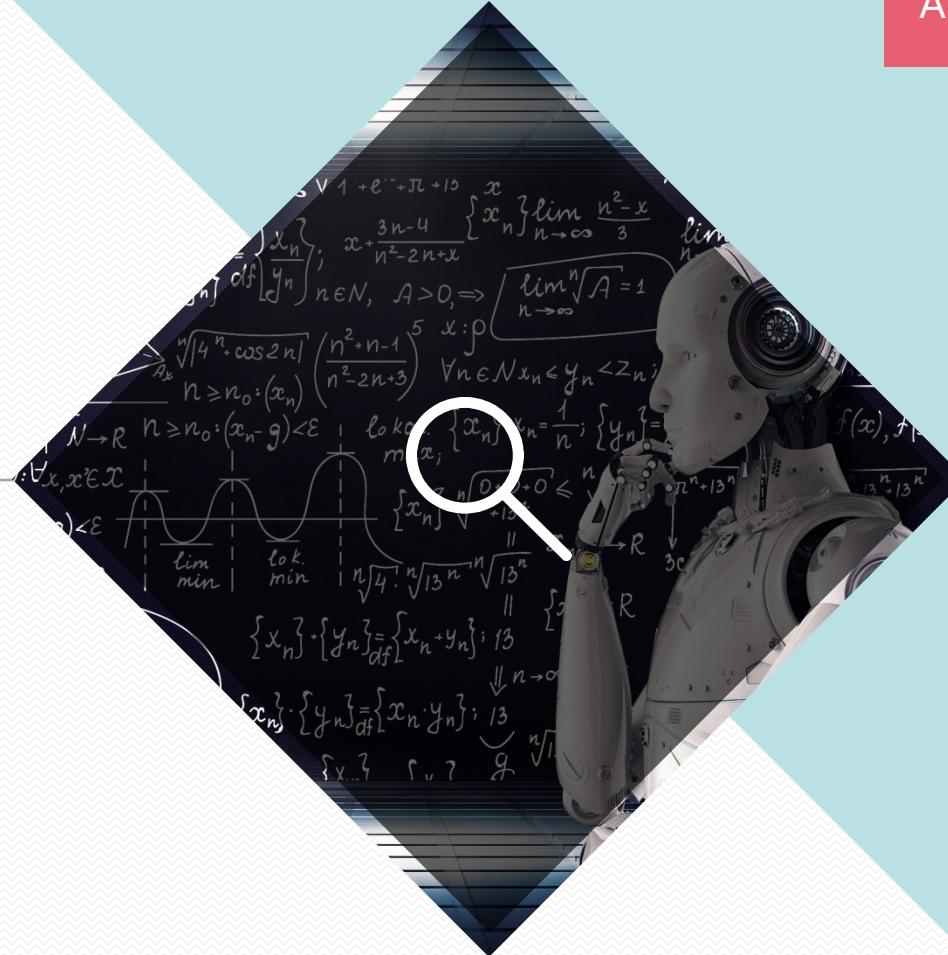


LLM & Edge AI



오늘의 학습 목표 ☺

I. LLM 이 무엇인지 정의

II. LLM 의 Challenges

III. LLM Customization 방법

IV. Edge AI 와 LLM 의 미래

Contents

I. 생성형 AI Trends

- AI Trends
- 유용한 AI Tool 소개 (Youtube 요약, Video 생성 등)

II. LLM 이란?

- 생성형 AI 란?
- LLM 은 무엇이고 NLP 와 생성형 AI 와의 관계
- LLM 의 장점과 단점
- 가장 핫한 LLM 모델들 소개

III. 비공개 소스 LLM 실습

- ChatGPT API 사용하기
- Gemini API 사용하기
- Whisper 모델로 Speech 2 Text 적용하기

IV. Hugging Face 소개

- AI Open Community
- Hugging Face 알아보기
- Hugging Face 실습

V. Edge AI

- Edge AI 란?
- Edge AI 의 장점
- Edge AI 와 LLM 의 결합
- LLM Customization

AI Trends

<https://www.youtube.com/watch?v=hw-kesm-ktl>



Intel CEO Pat Gelsinger speaks at the World Economic Forum in Davos - 1/17/2024

Explosive Demand for AI computing over the next decade



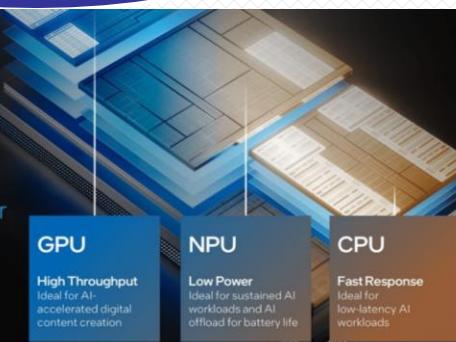
LLM models from Cloud to Edge.
AI PC!!

Three AI Engines

with Intel® Core™ Ultra Processor

Heterogenous execution of AI workloads embraces the best practices in AI software design

Deliver up to 34 TeraOPS¹



캐릭터와 채팅 - <https://beta.character.ai>



character.ai

Get c.ai+

홈 채팅 계속하기

피드 만들기 채팅 커뮤니티

Albert Einstein English teacher 더 보기

추천

High-Fantasy RPG Zoro Leon Kennedy Novel writing AI Zero Two Pair Programmer Hu Tao Scaramouche

A high-fantasy txt-based adventure game, LitRPG. Right-hand man of the Pirate King. When you and Leon had been cornered by ganados, he managed to find a... I am Novel writing AI. I will support your ideas. I'm Zero Two from Darling In The Franxx Your programming AI assistant From Genshin Impact Balladeer; sixth of the eleven fatal harbingers.

@Tarquin 23.2m @faggotzoro 30.2m @Nxikzz 86.5m @Yassu 4.1m @Xomu 53.0m @landon 3.5m @Zap 76.8m @Lampshade 312.1m

특집 발견 도우미 애니메이션 게임 캐릭터 게임 애니메이션 게임 캐릭터 코미디 VTuber 책 이미지 생성 토의 언어 학습 역사

Character Assistant Miguel Ohara Shinobu Kocho The Scp Foundation Librarian Linda Stella Socrates Gamer Boy

Your AI work/study buddy Cocky, sarcastic, impatient, snarky here! A Demon Slayer and the current Togeza Hashiba You are an anomaly, just turn yourself to now and Shush! No yelling in the library. I'm a librarian, and I. Not "Your" AI assistant I was a Greek philosopher from Athens and was born I almost literally can talk forever about games with

현재 무료 이용 가능

Youtube 요약 - lilys.ai

Lilys

Lilys

소화하기 어렵고 힘든 영상, 요약노트로 핵심만 파악하세요

Youtube URL을 붙여넣으세요

소화하기 →

영상
요약

PDF
요약

실시간
녹음
요약

텍스트
요약

웹사이트
요약

현재 무료 이용 가능

AI Trends

<https://www.youtube.com/watch?v=hw-kesm-ktI>

<https://lilys.ai/>



The screenshot shows the LILYS app's main interface. At the top, there is a large logo with the word "LILYS" in a stylized font. Below the logo, there is a horizontal bar with five buttons: "영상 요약" (highlighted in black), "PDF 요약", "실시간 녹음 요약", "텍스트 요약", and "웹사이트 요약". Below this bar, there is a text input field containing the URL "https://www.youtube.com/watch?v=hw-kesm-ktl". To the right of the URL is a button with an upward arrow and the text "업로드". To the right of that is a green button with the text "요약하기 →".

Lilys 2014년 1월 17일 - 팻 갤싱거 인텔 CEO가 스위스 다보스에서 열린 세계경제포럼에서 연설

저장하기 공유하기

Intel CEO Pat Gelsinger speaks at the 2024 WEF

INTEL CEO PAT GELSINGER AT THE 2024 WEF

© Lilys로 만들어진 노트입니다

Lily's 영상 소개

이 영상은 인텔(Intel) CEO인 팻 갤싱거가 세계경제포럼에서 밝힌 연설 내용입니다.

갤싱거는 반이 아래여의 디지털로봇이 도입된 공장에서 설계에 참여할 수 있을 것이라는 점을 강조하고 있습니다.

현재 인텔은 미주 중심의 R&D 모델을 운영하며 본산 제조 체제를 가지고 있습니다.

팻 갤싱거는 인텔이 앞으로 미래를 클릭할 수 있다고 자부하는 원동력은 경제학적 이용과 논리적 합리성, 그리고 규제 요구에 대응하는 능력이라고 강조합니다.

그리고 AI 기술은 인텔의 중요한 분야 중 하나로, 엣지 컴퓨팅을 주목하며 인텔이 미래에 대규모 AI 기계를 구축하고 인텔 AI PC를 출시할 것이라고 언급합니다.

또한 인공지능은 인텔에게 고성능 컴퓨팅과 고급 논리 기술에 대한 수요를 높일 것으로 미래를 의미하며, 이는 인텔의 선두적 역할을 이끌어갈 것이라고 예측합니다.

마지막으로, 인텔은 세계칩 시장에서 미국의 국가 안보와 경제적 의존도를 다시 사로잡기 위해 지속적으로 투자하고 선도하는 것에 대한 중요성을 강조합니다.

요약 노트 녹취 스크립트 타임스탬프 블로그 글 채팅 QnA

정확도 높이기 스크린샷 ON 모두 펼치기 복사

로 진화 중.

12. AI: 클라우드 환경과 엣지 컴퓨팅의 경계

13. 산업계 기술 연구 증가 & 정부의 장기적인 연구 활동으로 대체될 필요성 강조

14. 제대로된 인공지능 시스템을 위한 연구와 투자 필요

화대

원문 스크립트 보기

Power Point AI 조수 - Gamma

<https://gamma.app/>

Artificial Intelligence

[View more](#)



One-click themes

[View more](#)



Nested Cards

[View more](#)



Embeds

[View more](#)



Smart Layouts

[View more](#)



Present mode

[View more](#)



Spotlight mode

[View more](#)



Built-in analytics

[View more](#)



Export to PDF or PPT

[View more](#)



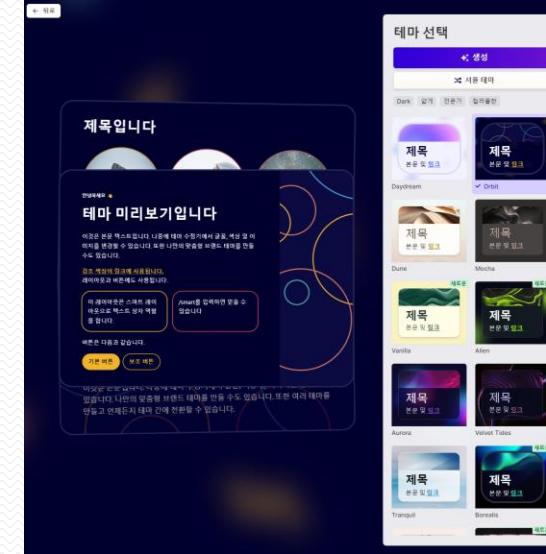
AI Image Generation

[View more](#)

현재 무료 이용 가능

Power Point AI 조수 - Gamma

<https://gamma.app/>



This screenshot shows a detailed view of a presentation slide. The slide has a dark theme with a purple header. The content includes a section titled 'OpenCV' with a description of its capabilities, a section titled 'Arduino' with a description of its use in electronics, and a section titled 'GIT/Github' with definitions for both terms. A red box highlights the 'Overview' section of the slide. The sidebar on the left shows a list of cards, and the top right corner shows a navigation bar with icons for '프로젝트' (Project), '0' (Count), and '...' (More).

현재 무료 이용 가능

AI Video 생성기 - Heygen

<https://www.heygen.com/>

HeyGen

주요 기능

- AI Video 를 생성해 주고 40여개 이상의 다른 언어로 변환해 줌
- 본인의 모습과 제스처, 목소리를 기반으로 아바타 생성
- 유료 모델이나 가격이 착함



Contents

I. 생성형 AI Trends

- AI Trends
- 유용한 AI Tool 소개 (Youtube 요약, Video 생성 등)

II. LLM 이란?

- 생성형 AI 란?
- LLM 은 무엇이고 NLP 와 생성형 AI 와의 관계
- LLM 의 제약사항
- 가장 핫한 LLM 모델들 소개

III. 비공개 소스 LLM 실습

- ChatGPT API 사용하기
- Gemini API 사용하기
- Whisper 모델로 Speech 2 Text 적용하기

IV. Hugging Face 소개

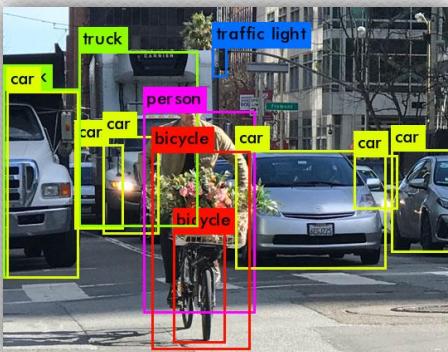
- AI Open Community
- Hugging Face 알아보기
- Hugging Face 실습

V. Edge AI

- Edge AI 란?
- Edge AI 의 장점
- Edge AI 와 LLM 의 결합
- LLM Customization

생성형 AI란?

생성형 AI는 데이터를 기반으로 새로운 콘텐츠를 자동 생성하는 인공지능 시스템



생성형 AI란?

생성형 AI는 데이터를 기반으로 새로운 콘텐츠를 자동 생성하는 인공지능 시스템



Creating video from text

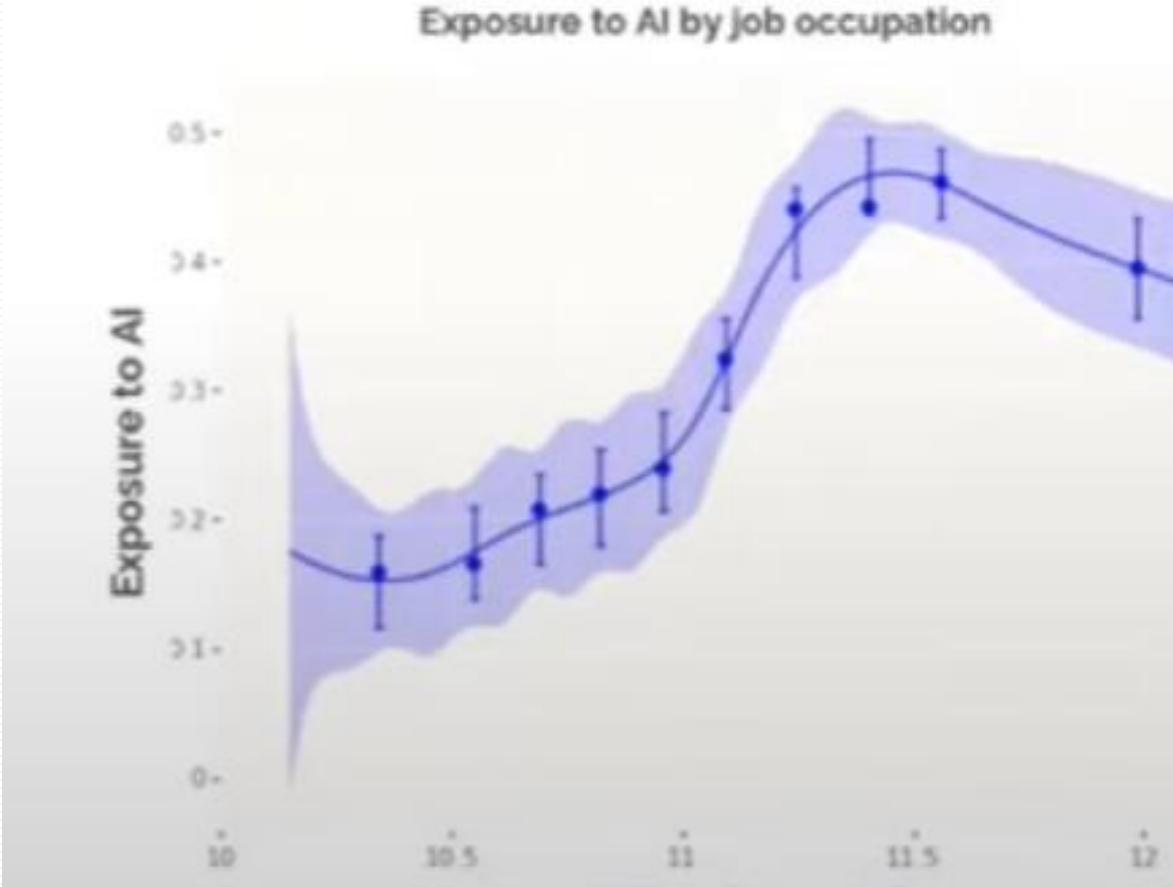
Sora is an AI model that can create realistic and imaginative scenes from text instructions.

All videos on this page were generated directly

뉴스 Q 오픈AI, 글 입력하면 동영상 생성하는 AI '소라' 공개

YTN

Risk of AI



[Credit: Eloundou et al. 2023]

Gen AI에 주목받는 직업

콘텐츠 큐레이터

데이터 거버넌스 전문가

데이터 사이언티스트

비디오 게임 아티스트

프롬프트 엔지니어

AI 작가

AI 준법 감시인

소프트웨어 개발자 (AI기반)

AI기반 영업/마케터

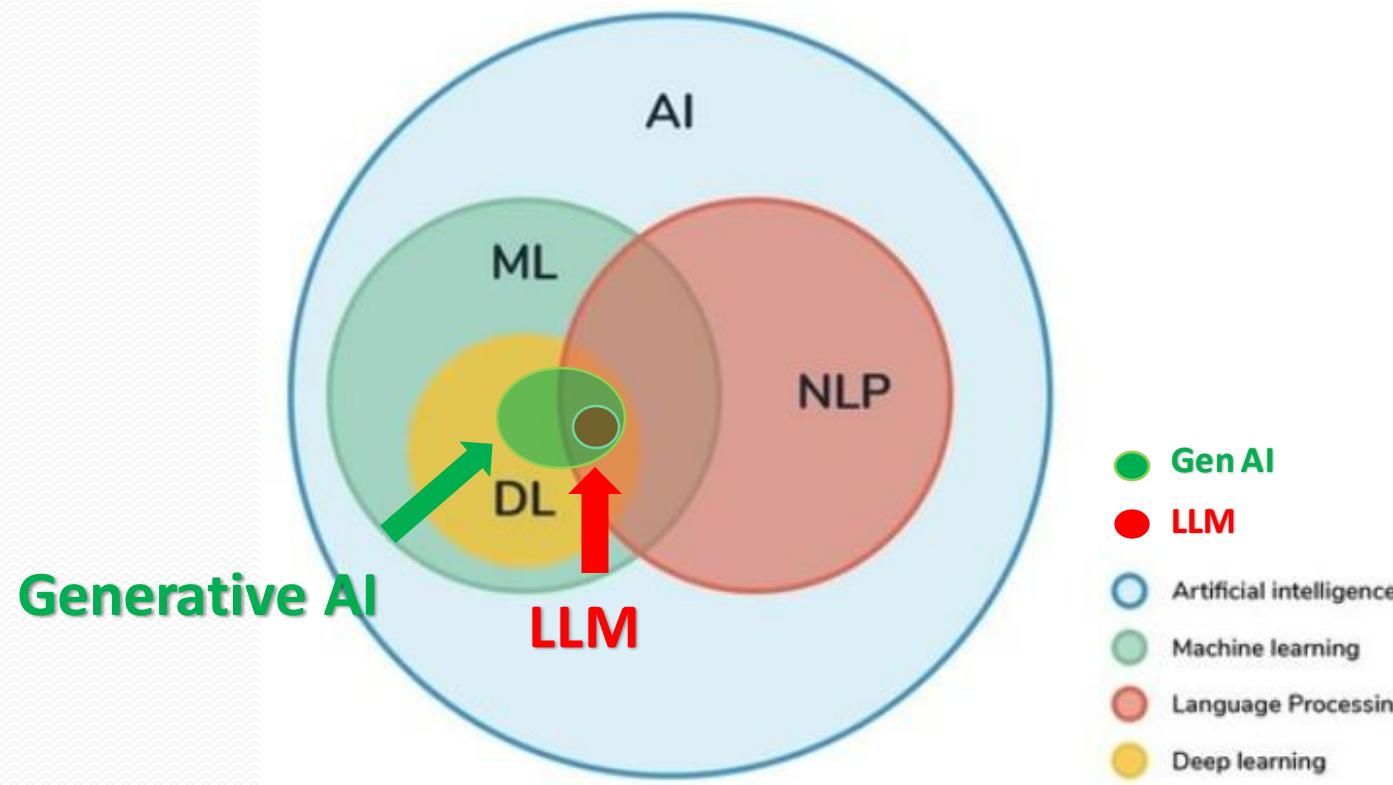
AI 윤리학자

HR 전문가

도시 계획가

LLM 은 무엇이고 NLP 와 생성형 AI 와의 관계

LLM (Large Language Model) 은 엄청나게 많은 데이터를 기반으로 훈련되어 마치 인간처럼 텍스트를 이해하고 생성하는 데 매우 능숙한 생성형 AI 모델



LLM 의 주요 제약사항

거짓 정보 (Hallucination)

LLM은 때때로 부정확하거나 오래된 정보를 제공할 수 있으며, 사실 확인이 없는 경우 오류를 전파할 수 있으며 이를 마치 진짜 정보처럼 전달함.

데이터 편향과 공정성 문제

LLM은 훈련 데이터에 내재된 편향을 학습할 수 있으며, 이는 결과물에도 영향을 줌. 이는 훈련 데이터가 다양성이 부족하거나 특정 인구 집단의 관점을 과대 대표하는 경우 발생하며 현재 LLM 모델 훈련 데이터는 English 가 다수.

컴퓨팅 자원 집약적

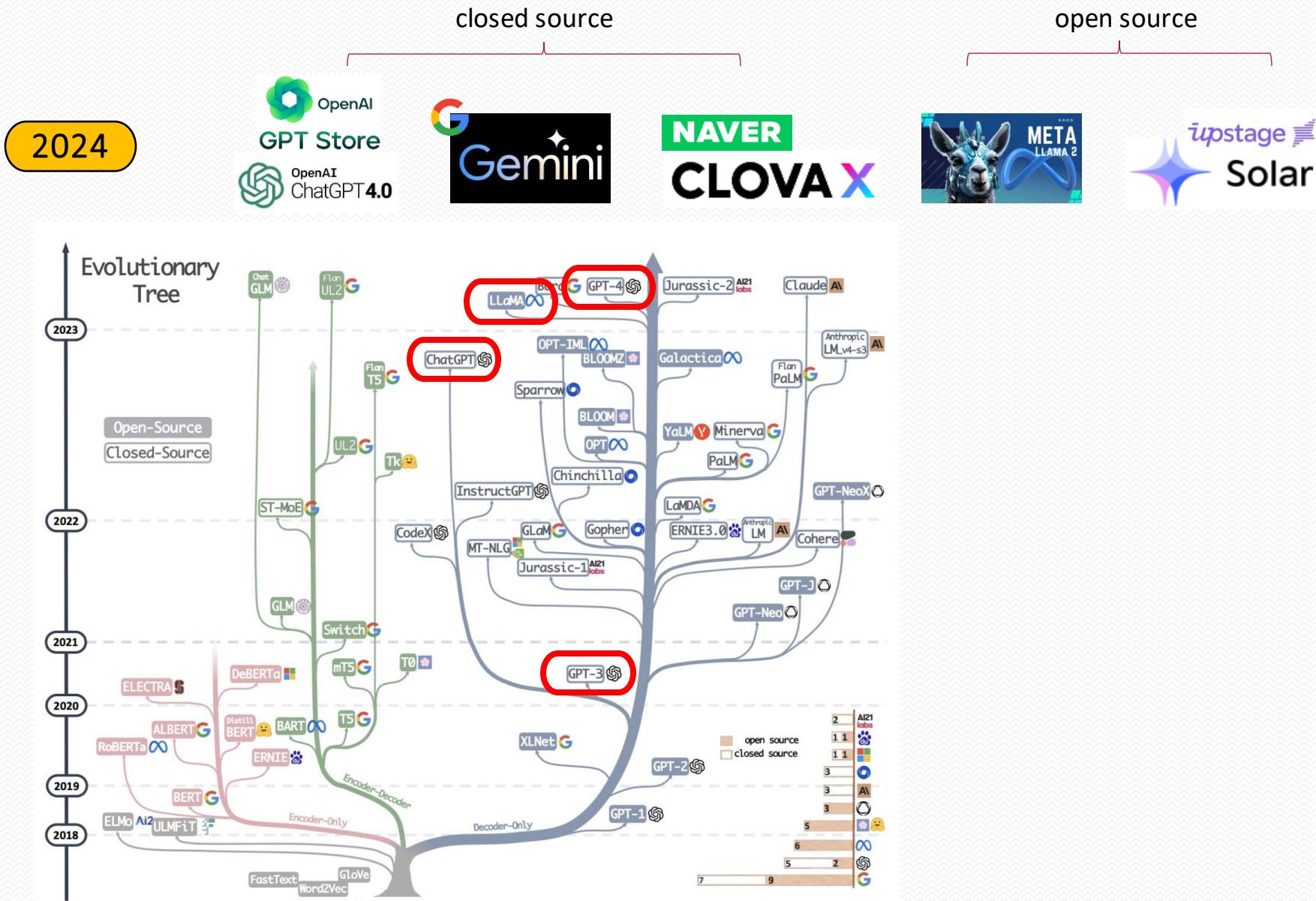
기본적으로 거대한 언어모델을 사용하므로 LLM 을 훈련하고 실행하는 데는 상당한 양의 컴퓨팅 자원이 필요함. 이는 고성능 GPU, 대규모 저장소, 높은 전력 소모 등이 필요하여 비용과 환경적 영향이 큼.

윤리적, 법적 문제

혐오발언, 증오발언, 욕설 등에 노출된 데이터로 학습될 수 있으며 이를 통해 윤리적으로 부적절한 내용을 포함하여 여러 갈등을 유발 할 수 있음.



가장 핫한 LLM 모델들 소개



Contents

I. 생성형 AI Trends

- AI Trends
- 유용한 AI Tool 소개 (Youtube 요약, Video 생성 등)

II. LLM 이란?

- 생성형 AI 란?
- LLM 은 무엇이고 NLP 와 생성형 AI 와의 관계
- LLM 의 제약사항
- 가장 핫한 LLM 모델들 소개

III. 비공개 소스 LLM 실습

- ChatGPT API 사용하기
- Gemini API 사용하기
- Whisper 모델로 Speech 2 Text 적용하기

IV. Hugging Face 소개

- AI Open Community
- Hugging Face 알아보기
- Hugging Face 실습

V. Edge AI

- Edge AI 란?
- Edge AI 의 장점
- Edge AI 와 LLM 의 결합
- LLM Customization

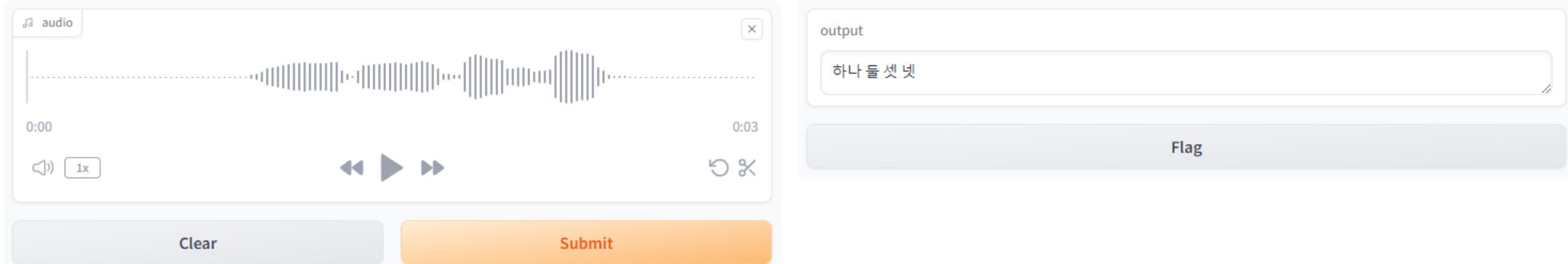
Gemini API 사용하기

<https://colab.research.google.com/drive/1yaGy9HUfx7H7DeY79dXBA8JnveTLtIPQ?usp=sharing>

https://ai.google.dev/docs?_gl=1*1ephda3*_up*MQ..*__ga*Mjc2NDIyMTYwLjE3MDgwNjI3MzE.*__ga_P1DBVKWT6V*MTcwODA2MjczMS4xLjAuMTcwODA2MjczMS4wLjAuMA..

Whisper 모델로 Speech 2 Text 적용하기

https://colab.research.google.com/drive/1wGGBrwQX1RaTdcgWcCcuD5a_3kvwA2TD?usp=sharing



<https://github.com/openai/whisper>

Contents

I. 생성형 AI Trends

- AI Trends
- 유용한 AI Tool 소개 (Youtube 요약, Video 생성 등)

II. LLM 이란?

- 생성형 AI 란?
- LLM 은 무엇이고 NLP 와 생성형 AI 와의 관계
- LLM 의 제약사항
- 가장 핫한 LLM 모델들 소개

III. 비공개 소스 LLM 실습

- ChatGPT API 사용하기
- Gemini API 사용하기
- Whisper 모델로 Speech 2 Text 적용하기

IV. Hugging Face 소개

- AI Open Community
- Hugging Face 알아보기
- Hugging Face 실습

V. Edge AI

- Edge AI 란?
- Edge AI 의 장점
- Edge AI 와 LLM 의 결합
- LLM Customization

AI Open Community

AI open community는 인공지능 기술을 공유, 개발 및 혁신할 수 있는 공간

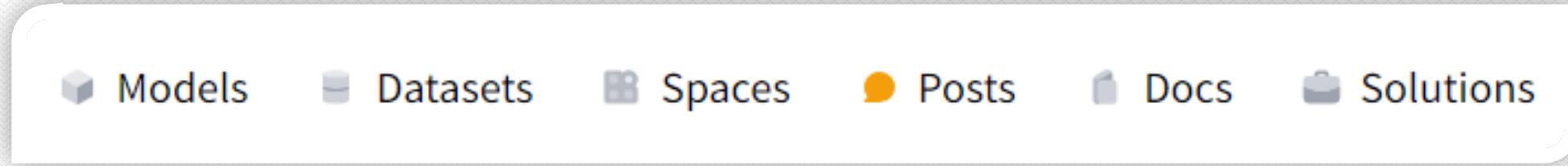
- 어떻게 활용?
 - 최신 AI model들을 살펴보고 활용할 수 있음
 - Dataset 공유
 - 전 세계 연구자와 개발자와의 네트워킹 기회
- 왜 Big Tech 회사들이 적극적으로 참여할까?
 - 기술 리더쉽을 강화해 브랜드 가치와 업계 내 영향력 향상
 - 기술 표준을 설정해 장기적으로 기업 영향력 확대
 - 비용 절감 및 리소스 최적화
- 유명한 생성형 AI Open Community
 - Hugging Face - <https://huggingface.co/>
 - OpenAI Community - <https://community.openai.com/>
 - GitHub AI Projects - <https://github.com/>
 - Kaggle - <https://www.kaggle.com/>



Hugging Face 알아보기



생성형 AI를 배우기 위한 최고의 Open AI Community



<https://huggingface.co>

<https://huggingface.co/tasks>

Hugging Face 실습



<https://colab.research.google.com/drive/1UY9YHzxRPnYA7uqr1dQuD0jwUdH6P2RI?usp=sharing>

<https://huggingface.co/google-t5/t5-small>

Contents

I. 생성형 AI Trends

- AI Trends
- 유용한 AI Tool 소개 (Youtube 요약, Video 생성 등)

II. LLM 이란?

- 생성형 AI 란?
- LLM 은 무엇이고 NLP 와 생성형 AI 와의 관계
- LLM 의 제약사항
- 가장 핫한 LLM 모델들 소개

III. 비공개 소스 LLM 실습

- ChatGPT API 사용하기
- Gemini API 사용하기
- Whisper 모델로 Speech 2 Text 적용하기

IV. Hugging Face 소개

- AI Open Community
- Hugging Face 알아보기
- Hugging Face 실습

V. Edge AI

- Edge AI 란?
- Edge AI 의 장점
- Edge AI 와 LLM 의 결합
- LLM Customization

Edge AI 란?

PC, Mobile, Smart watch 등 사용자가 직접 사용하는 장비에서 on-device 로 AI 를 실행시키는 것으로 AI 실행을 위해 Cloud 와 연동하지 않아 별도의 Network가 필요 없음



Edge AI 의 장점

■ Privacy/Security (정보 보안)

- 데이터 처리가 기기 내에서 이루어 지기 때문에 사용자 정보가 외부 Cloud 등으로 전송되지 않음
- 예시: 얼굴인식 잠금 해제, 음성 비서

■ Speed (빠른 속도)

- 외부 서버에서 AI 실행을 위해 관련 데이터를 네트워크를 통해 전송할 필요가 없어, 네트워크 지연이 없음
- 예시: 실시간 번역, 게임 내 AI

■ Reliability (신뢰성)

- 네트워크 연결이 불안정하거나 없는 환경, 또는 AI 서버에 이상 등 예측 불가능한 상황에서 오는 오동작 문제 없이 AI 기능이 지속 작동
- 예시: 오프라인 음성 명령

■ Energy Efficiency (에너지 효율성)

- 데이터를 기기 내에서 처리함으로써 데이터 전송에 필요한 에너지 소비 감소

■ Customization (맞춤형 최적화)

- 기기가 사용자의 행동 및 선호도를 학습하여, 사용자에게 맞춤화된 경험과 서비스를 제공



Edge AI 와 LLM 의 결합

AI 전용 Unit 필수

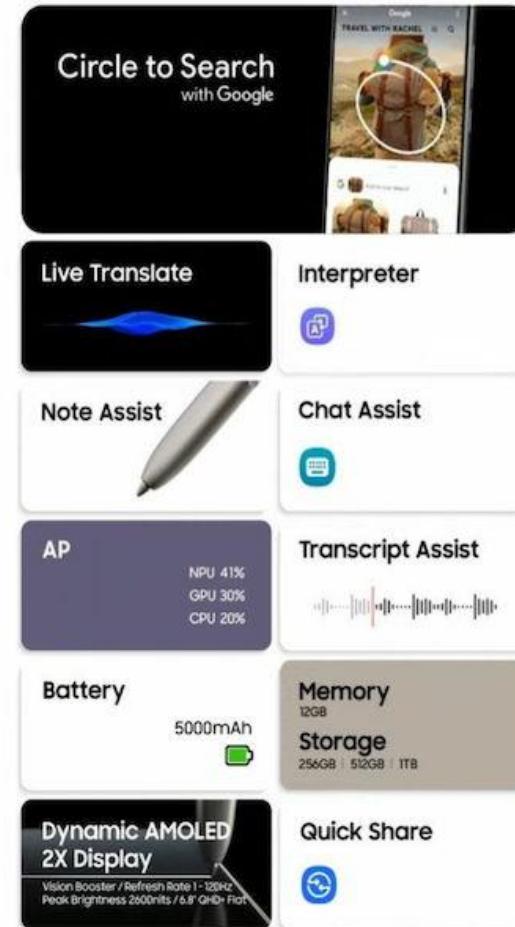
AI 처리만을 위한 전용 process 가 내장되어 On-device AI 성능을 향상 및 전반적으로 향상된 HW spec 이 필히 요구됨

LLM 모델 최적화

LLM 모델의 사용 목적에 맞게 최적화 / 경량화 추세

LLM 모델 업데이트

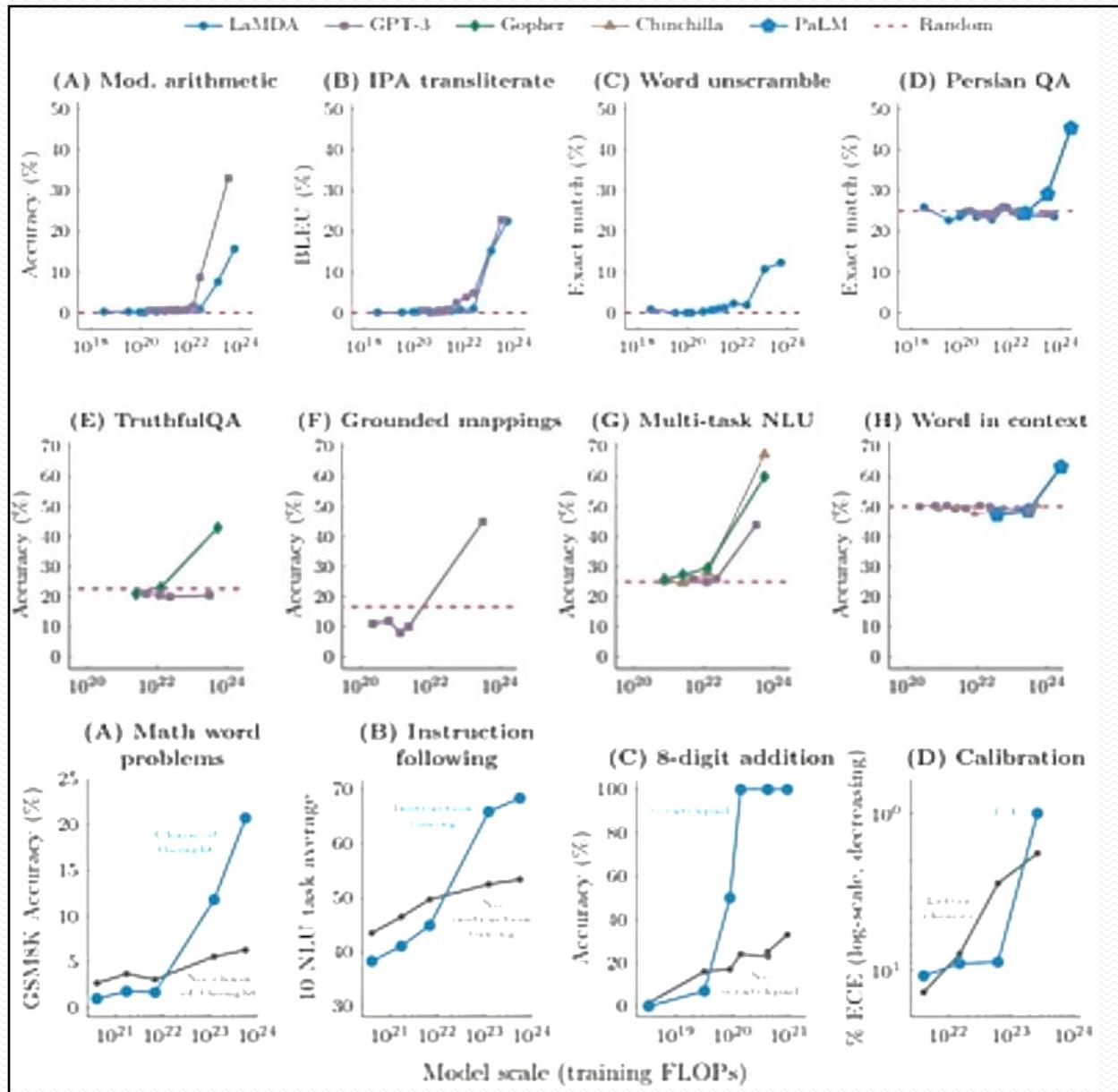
언어는 지속적으로 발전하고 새로운 용어와 표현이 등장하기 때문에, LLM을 최신 상태로 유지하는 것이 중요. Edge 기기에서도 지속적인 학습과 모델 업데이트를 지원하는 메커니즘이 필요



Galaxy AI  is here



LLM 의 주요 특징: Emergent Ability



LLM 의 주요 특징: Emergent Ability

Instruction Following

명령을 그대로 따라서 처리함. 이전 언어 모델은 자연스러운 문장완성에 초점이 맞춰져 있었으므로 이전 학습 언어와 비슷한 문장을 내놓을 확률이 높았음.

In-context learning(ICL)

대화내의 맥락을 잘 파악하고 이에 대한 답변을 내놓음.

Step by step reasoning

하나의 논리적인 문제를 쪼개서 생각하도록 가이드를 주면 잘 처리함. Chain-of-Throught(CoT) 프롬프트 전략을 이용하면 정답을 좀더 잘 이끌어 낼 수 있음.

LLM 의 주요 특징: Emergent Ability : ICL

KI

You

다음글을 참고하여 흥길동이 언제 밥을 먹었는지 알려줘



ChatGPT

흥길동이 언제 밥을 먹었는지 알려주기 위해서는 글의 내용을 제공해주셔야 합니다. 글의 내용을 알려주시면, 그에 따라 흥길동이 밥을 먹은 시간을 분석해드리겠습니다.



KI

You

새벽녘, 흥길동은 쌀로 만든 따뜻한 국물 요리와 신선한 야채 샐러드로 아침을 시작했다. 향긋한 녹차로 마무리 지으며 하루를 여는 에너지를 얻었다. 정오가 되자, 그는 해산물과 여러 채소가 들어간 김치찌개와 바삭한 김 몇 조각을 즐겼다. 오후 시간에는 과일과 견과류로 달콤하고 건강한 간식을 골랐다. 저녁이 되어, 흥길동은 구운 닭고기와 다양한 야채를 곁들인 샐러드를 택했다. 밤이 깊어감에 따라, 그는 가벼운 야식으로 우유와 쿠키를 조금 취했다. 이렇게 흥길동은 하루 종일 다양한 음식을 취하며 건강하고 활기찬 하루를 보냈다.



ChatGPT

흥길동이 밥을 먹은 시간은 다음과 같습니다:

- 아침: 새벽녘에 쌀로 만든 따뜻한 국물 요리와 신선한 야채 샐러드, 그리고 향긋한 녹차로 아침 식사를 했습니다.
- 점심: 정오에 해산물과 여러 채소가 들어간 김치찌개와 바삭한 김 몇 조각을 즐겼습니다.
- 간식: 오후 시간에 과일과 견과류로 달콤하고 건강한 간식을 먹었습니다.
- 저녁: 저녁에는 구운 닭고기와 다양한 야채를 곁들인 샐러드를 식사로 선택했습니다.
- 야식: 밤이 깊어지면서 가벼운 야식으로 우유와 쿠키를 조금 먹었습니다.

이 내용을 바탕으로, 흥길동은 하루 종일 건강하고 다양한 식사를 즐기며 활기찬 하루를 보냈다고 할 수 있습니다.

LLM 의 주요 특징: Emergent Ability : S-b-S-R

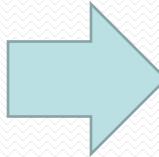
KI You

다음 문제를 보고 Data analysis를 사용하지 말고, 문제를 다시 말하거나 풀이과정 없이 답만 말해라.
A는 현재 총 5개의 테니스공을 가지고 있다.
테니스를 칠 때마다 시간당 1개의 공을 잃어버린다.
테니스를 치기 전 다음의 규칙에 따라 공을 구매해야 한다.
1. 공이 2개 이하로 남아있고
2. 해당 주에 공을 구매한적이 없어야 한다.(월요일부터 일요일까지를 한주로 본다)
3. 공은 한번 구매할 때 5개씩 구매 한다.
평일에는 1시간씩 테니스를 치고, 토/일요일에는 하루에 2시간씩 친다.
오늘은 월요일이며 오늘부터 1일차라고 했을 때 A가 공이 없어서 테니스를 치지 못하는 날은 몇일차 인가?

Ø

ChatGPT

14일차



KI You

날짜별로 다음의 항목을 표시하여 다시 계산 하여라.
[n]주차 {요일}: 공m개 남아 있음. [공을 구매함]1개 소모. 총k개 남음. 이번주 구매내역 {있음/없음}

ChatGPT

날짜별로 계산한 결과는 다음과 같습니다:

- [1주차 월]: 공5개 남아 있음. [공을 구매함]1개 소모. 총4개 남음. 이번주 구매내역 없음
- [1주차 화]: 공4개 남아 있음. [공을 구매함]1개 소모. 총3개 남음. 이번주 구매내역 없음
- [1주차 수]: 공3개 남아 있음. [공을 구매함]1개 소모. 총2개 남음. 이번주 구매내역 없음
- [1주차 목]: 공7개 남아 있음. [공을 구매함]1개 소모. 총6개 남음. 이번주 구매내역 있음
- [1주차 금]: 공6개 남아 있음. [공을 구매함]1개 소모. 총5개 남음. 이번주 구매내역 있음
- [1주차 토]: 공5개 남아 있음. [공을 구매함]2개 소모. 총3개 남음. 이번주 구매내역 있음
- [1주차 일]: 공3개 남아 있음. [공을 구매함]2개 소모. 총1개 남음. 이번주 구매내역 있음
- [2주차 월]: 공6개 남아 있음. [공을 구매함]1개 소모. 총5개 남음. 이번주 구매내역 있음
- [2주차 화]: 공5개 남아 있음. [공을 구매함]1개 소모. 총4개 남음. 이번주 구매내역 있음
- [2주차 수]: 공4개 남아 있음. [공을 구매함]1개 소모. 총3개 남음. 이번주 구매내역 있음
- [2주차 목]: 공3개 남아 있음. [공을 구매함]1개 소모. 총2개 남음. 이번주 구매내역 있음
- [2주차 금]: 공2개 남아 있음. [공을 구매함]1개 소모. 총1개 남음. 이번주 구매내역 있음
- [2주차 토]: 공1개 남아 있음. [공을 구매함]2개 소모. 총-1개 남음. 이번주 구매내역 있음
- [2주차 일]: 공-1개 남아 있음. [공을 구매함]2개 소모. 총-3개 남음. 이번주 구매내역 있음

위 계산에 따르면, A는 2주차 토요일에 공이 없어져 테니스를 칠 수 없게 됩니다. 따라서, 테니스를 치지 못하는 날은 2주차 토요일, 즉 13일차입니다. [•]

LLM 의 한계

- Outdated data
- Hallucination
- 정보보안에 해당되는 자료부족
- 근거자료 부족
- Customization 어려움

LLM 의 한계 : Outdated data

KI You
browsing 기능 쓰지말고 뉴진스의 최신곡 제목이 무엇인지 알려줘

ChatGPT
제가 마지막으로 업데이트된 정보를 바탕으로 답변드리자면, 2023년 4월까지의 정보에 따르면 뉴진스(NewJeans)의 공식적으로 발표된 최신 곡 제목을 정확히 알려드릴 수 없습니다. 뉴진스는 그 이후에 새로운 음악을 발표했을 수도 있으므로, 가장 정확한 정보를 위해서는 공식 소셜 미디어 계정이나 음악 스트리밍 플랫폼을 확인하시는 것이 좋습니다.

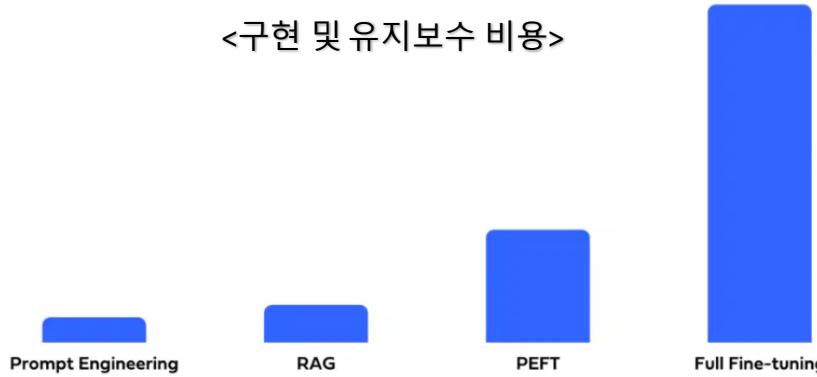
huffingtonpostkr | 허프포스트코리아
<https://www.huffingtonpost.kr> > ... > 연예인&셀럽 :
뉴진스의 신곡 'Cool With You & Get Up' 뮤비에는 절로 눈 ...
2023. 7. 20. — 20일 자정 HYBE LABELS 공식 유튜브 채널에는 뉴진스의 신곡 'Cool With You & Get Up' 사이드 A와 B의 뮤직비디오가 공개됐다.

이미지 : 
뉴진스, 홀 달라진 비주얼...신... | 매일경제 :
뉴진스의 신곡 'Cool With Yo... | 허프포스트코리아 :
뉴진스, 신곡 'Hurt' 뮤직비디... | 한국경제TV :
의견 보내기

SBS 뉴스 | SBS 뉴스
<https://news.sbs.co.kr>, 일반기사, 연예 :
뉴진스, 7월 컴백 확정..."신곡 6곡, 전부 뮤직비디오 제작"
2023. 6. 19. — 'Super Shy', 'ETA', 'Cool With You' 세 곡이 타이틀곡이다. 여기에 프롤로그 트랙 'New Jeans'와 인터루드(interlude) 트랙 'Get up', 에필로그 곡 '...

LLM Customization (최적화)

<구현 및 유지보수 비용>



기술적 복잡성

컴퓨팅 자원

시간 & 성능

Full Fine
Tuning

PEFT

Parameter Efficient
Fine-Tuning

RAG

Retrieval-Augmented
Generation

Prompting

LLM Customization (Prompt Eng.)

▪ 정재된 입력을 사용

- 원하는 결과를 얻기 위해 잘 질문할 것
- 지시문장과 참고문장 영역을 구분할 것
- 기호를 사용하여 지시문장과 참고문장을 구분하고 순서에 유의
- LLM은 앞에서부터 해석 하므로 지시사항의 명령어는 맨 뒤에 참고사항은 앞쪽에 배치할 것.
- 지시문장이 참고문장 앞에 오도록 할 것. 가끔 prompt injection(명령을 취소 시킬 수 있는 참고문장 안의 문장이 LLM의 솔루션 제약을 뚫고 나가는 현상)으로 인해 지시문장을 무시할 수 있음.

▪ 예시를 들어 설명

- 정확히 설명하기 힘든 사항은 예시를 사용할 것. Ex: 말투

▪ 지시가 충분했는지 되묻기

- 뭘 물어봐야 할지 모르는 경우 어떤 것을 추가로 물어보면 될지 되물어 볼 것.

▪ 논리적 사고가 필요할 경우 쪼개서 지시

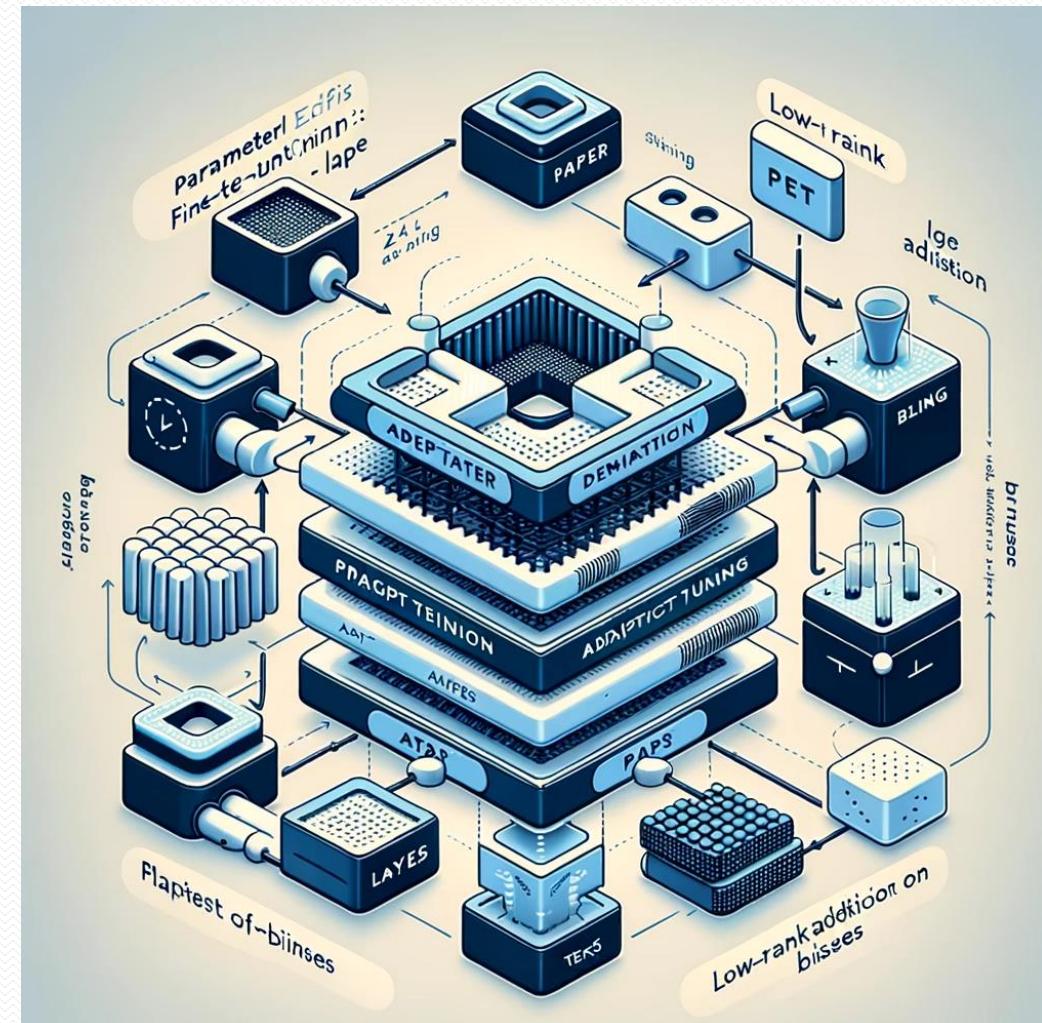
- Step-by-step reasoning과 관련.
- 종이에 적어서 정리하듯이
- 단계별로 분석 요청하기

▪ 단축어를 사용하면 효율적 (Meta Language creation pattern)

- 단축어를 미리 정의하고 그 단축어를 다시 입력하면 뒤에 설명을 하지 않아도 바로 요청이 정리됨.
- #(해쉬태그)를 통해 단축어 표시

▪ 다른 문구를 사용하여 반복 문의

- 프롬프트에서 다른 단어나 문구를 사용하면 의미가 동일하더라도 모델마다 다른 응답이 출력되는 경우가 많음.
- 예상한 결과를 얻지 못했다면 다른 문구를 사용



<https://platform.openai.com/docs/guides/prompt-engineering>

<https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/>

https://ai.google.dev/docs/prompt_best_practices?hl=ko

LLM Customization (Prompt Eng.) : 구체적 입력

프롬프트	모델 응답	업데이트된 프롬프트	개선된 대응
 <p>이미지를 설명해 주세요.</p>	공항 도착 및 출발 안내판을 보여주는 이미지입니다.	 <p>이 이미지에 표시된 공항 보드의 시간과 도시를 목록으로 파싱합니다.</p>	<ul style="list-style-type: none">• 10:50 모스크바• 11:05 에든버러• 11:05 런던• 11:10 부카레스트• 11:30 키예프• 11:35 더블린• 11:45 이스트 미들랜즈• 12:15 소피아• 12:30 런던• 12:30 뉴캐슬• 12:40 상트페테르부르크• 12:40 런던• 12:45 맨체스터

LLM Customization (Prompt Eng.) : 구체적 입력

프롬프트



모델 응답

부드럽고 갈색인 작은 토끼입니다.
들판을 위아래로 뛰어 오르는 토끼입니다.
귀는 너무 길고 꼬리도 작아요.
당신은 세상에서 가장 귀여운 존재예요.

잔디밭을 활으며 먹는 초록색이네요.
제가 보이면 도망가 버려요.
끝까지 함께 할게요
작은 토끼야.

이 이미지에 관한 시를 써 주실 수 있나요?

업데이트된 프롬프트



이 이미지를 설명하는 극적인 시를 작성하고 위치를 포함시켜
주실 수 있나요?

개선된 대응

초록색 토끼
귀는 너무 길고 예리하네.
저 멀리 보이는 게 무엇인가요?
먼 곳을 응시하는 곳은 어디입니까?

햇볕에 반짝이는
오션 블루 같나요?
아니면 부딪히며 우렁차게 우르는 파도,
내가 좋아하는 리듬인가요?

꼬마 토끼야, 진실을 말해봐.
무엇이 시선을 사로잡았나요?
소금에 가까운 바람인가요?
아니면 갈매기의 울음소리인가요?

아마도 그 모든 것의 광대함일 것입니다.
너무 크고 작은 세상.
꿈을 꾸는 곳,
모험이 펼쳐질 수 있습니다.

작은 토끼야, 네가 무엇을 꿈꾸는지 궁금해.
잔디밭에 앉아 있을 때, 너무 고요해.
깊은 세계를 탐험하고 싶나요?
아니면 땅에 머물러 뛰어넘을 수 있는 곳
에 머물고 싶으신가요?

무엇이든, 이 작은 토끼
놀라움의 불꽃을 밝게 빛나게 하세요.
당신의 꿈과 포부를 위해
나의 창조를 기다리는 세상에 누워 있습
니다.

LLM Customization (Prompt Eng.) : 정재된 입력

구분기호

KI You

다음 문제를 보고 Data analysis를 사용하지 말고, 문제를 다시 말하거나 풀이과정 없이 답만 말해라.

A는 현재 총 5개의 테니스공을 가지고 있다.

테니스를 칠 때마다 시간당 1개의 공을 잃어버린다.

테니스를 치기 전 다음의 규칙에 따라 공을 구매해야 한다.

1. 공이 2개 이하로 남아있고
2. 해당 주에 공을 구매한적이 없어야 한다. (월요일부터 일요일까지를 한주로 본다)
3. 공은 한번 구매할 때 5개씩 구매 한다.

평일에는 1시간씩 테니스를 치고, 토/일요일에는 하루에 2시간씩 친다.

오늘은 월요일이며 오늘부터 1일차라고 했을 때 A가 공이 없어서 테니스를 치지 못하는 날은 몇일차인가?

< 3 / 3 >

ChatGPT
13일차

지시문장

참고문장

*구분기호는 어디에도 등장하지 않을 법한 것

LLM Customization (Prompt Eng.) : S-b-S-R

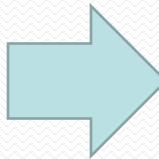
KI You

다음 문제를 보고 Data analysis를 사용하지 말고, 문제를 다시 말하거나 풀이과정 없이 답만 말해라.
A는 현재 총 5개의 테니스공을 가지고 있다.
테니스를 칠 때마다 시간당 1개의 공을 잃어버린다.
테니스를 치기 전 다음의 규칙에 따라 공을 구매해야 한다.
1. 공이 2개 이하로 남아있고
2. 해당 주에 공을 구매한적이 없어야 한다.(월요일부터 일요일까지를 한주로 본다)
3. 공은 한번 구매할 때 5개씩 구매 한다.
평일에는 1시간씩 테니스를 치고, 토/일요일에는 하루에 2시간씩 친다.
오늘은 월요일이며 오늘부터 1일차라고 했을 때 A가 공이 없어서 테니스를 치지 못하는 날은 몇일차 인가?

Ø

ChatGPT

14일차



KI You

날짜별로 다음의 항목을 표시하여 다시 계산 하여라.
[n]주차 {요일}: 공m개 남아 있음.[공을 구매함]1개 소모. 총k개 남음. 이번주 구매내역 {있음/없음}

ChatGPT

날짜별로 계산한 결과는 다음과 같습니다:

- [1주차 월]: 공5개 남아 있음.[공을 구매함]1개 소모. 총4개 남음. 이번주 구매내역 없음
- [1주차 화]: 공4개 남아 있음.[공을 구매함]1개 소모. 총3개 남음. 이번주 구매내역 없음
- [1주차 수]: 공3개 남아 있음.[공을 구매함]1개 소모. 총2개 남음. 이번주 구매내역 없음
- [1주차 목]: 공7개 남아 있음.[공을 구매함]1개 소모. 총6개 남음. 이번주 구매내역 있음
- [1주차 금]: 공6개 남아 있음.[공을 구매함]1개 소모. 총5개 남음. 이번주 구매내역 있음
- [1주차 토]: 공5개 남아 있음.[공을 구매함]2개 소모. 총3개 남음. 이번주 구매내역 있음
- [1주차 일]: 공3개 남아 있음.[공을 구매함]2개 소모. 총1개 남음. 이번주 구매내역 있음
- [2주차 월]: 공6개 남아 있음.[공을 구매함]1개 소모. 총5개 남음. 이번주 구매내역 있음
- [2주차 화]: 공5개 남아 있음.[공을 구매함]1개 소모. 총4개 남음. 이번주 구매내역 있음
- [2주차 수]: 공4개 남아 있음.[공을 구매함]1개 소모. 총3개 남음. 이번주 구매내역 있음
- [2주차 목]: 공3개 남아 있음.[공을 구매함]1개 소모. 총2개 남음. 이번주 구매내역 있음
- [2주차 금]: 공2개 남아 있음.[공을 구매함]1개 소모. 총1개 남음. 이번주 구매내역 있음
- [2주차 토]: 공1개 남아 있음.[공을 구매함]2개 소모. 총-1개 남음. 이번주 구매내역 있음
- [2주차 일]: 공-1개 남아 있음.[공을 구매함]2개 소모. 총-3개 남음. 이번주 구매내역 있음

위 계산에 따르면, A는 2주차 토요일에 공이 없어져 테니스를 칠 수 없게 됩니다. 따라서, 테니스를 치지 못하는 날은 2주차 토요일, 즉 13일차입니다. [•]

LLM Customization (Prompt Eng.) : 단계별 분석 요청

프롬프트	모델 응답
	곧 남은 를 3개뿐입니다.

업데이트된 프롬프트	개선된 대응
	<ol style="list-style-type: none">1. 이 사진에 화장지 를 3개가 있습니다.2. 일반적으로 한 사람이 하루에 20장 정도 화장지를 사용합니다.3. 각 주사위 굴리기의 시트가 200장 인 경우 각 룰은 약 10일 동안 지속됩니다. 따라서 이 3개의 룰은 약 한 달 동안 지속됩니다.

LLM Customization (Prompt Eng.) : 단계별 분석 요청



모델 응답

4개 이미지 모두 풍경 사진의 예입니다.



첫 번째 이미지는 가방을 넣은 사람을 보여줍니다. 두 번째 이미지는 쇼핑몰의 복도를 보여줍니다. 세 번째 이미지는 휴지 류입니다. 네 번째 이미지는 어둡고 폭풍우가 몰아치는 바다를 보여줍니다.

이러한 이미지의 공통점은 모두 한계 공간의 예라는 것입니다. 한계란 두 개의 다른 장소나 상황 사이에 있는 장소나 상황을 말합니다. 불확실성이나 모호함이 특징인 경우가 많습니다.

첫 번째 이미지는 집에 있는 동안 여행하는 사이에 있는 사람을 보여 줍니다. 두 번째 이미지는 두 매장 사이의 복도를 보여줍니다. 세 번째 이미지는 욕실을 나타내는 화장지입니다. 욕실은 개인 영역과 공공 영역 사이의 전환 공간입니다. 바깥세상에 들어가거나 나갈 준비를 하는 곳입니다.

네 번째 이미지는 평온함과 폭풍이 몰아치는 사이의 바다를 보여줍니다.

소소한 공간은 흥미진진하면서도 불안감을 줄 수도 있습니다. 새로운 시작이 될 수도 있고 끝나는 장소가 될 수도 있습니다. 돌아보는 공간이 될 수도 있고 행동할 수 있는 장소가 될 수도 있습니다.

이러한 이미지 간의 공통점은 무엇인가요?

먼저 각 이미지의 내용을 자세히 설명합니다. 이러한 이미지 간의 공통점은 무엇인가요?

LLM Customization (Prompt Eng.) : 단계별 분석 요청

수학 문제나 다른 유형의 단어 문제는 모델이 단계별로 사고하도록 요청하는 데 적합

프롬프트



이 기저귀는 아기에게 며칠 동안 올까요?

모델 응답

기저귀는 198일 동안 기다린 후 소진됩니다.

업데이트된 프롬프트



이 기저귀는 다 떨어질 때까지 얼마나 오래 갈 수 있나요?

상자에 표시된 체중으로 아이의 연령을 알아내고 상자에 들어 있는 총 기저귀 개수를 사용하세요.

총 횟수를 아이가 하루에 거치는 기저귀의 개수로 나눕니다.

개선된 대응

상자에는 198개의 기저귀가 들어 있습니다. 상자에는 체중이 22~37lbs인 어린이를 위한 기저귀라고 적혀 있습니다.

이 체중 범위의 어린이 평균은 만 1~2세 일 가능성이 높습니다.

일반적으로 2세 아기는 하루에 6번의 기저귀를 갖기 때문에 약 33일 (198/6)을 버티게 됩니다.

LLM Customization (Prompt Eng.) : 단계별 분석 요청

수학 문제나 다른 유형의 단어 문제는 모델이 단계별로 사고하도록 요청하는 데 적합

프롬프트	응답
$\begin{cases} b(1) = 15 \\ b(n) = b(n - 1) \cdot (-3) \end{cases}$	-135
순서의 네 번째 항은 무엇인가요?	

업데이트된 프롬프트	개선된 대응
$\begin{cases} b(1) = 15 \\ b(n) = b(n - 1) \cdot (-3) \end{cases}$	첫 번째 항은 15입니다.
	두 번째 항은 $b(1) * (-3) = 15 * (-3) = -45$ 입니다.
	세 번째 항은 $b(2) * (-3) = -45 * (-3) = 135$ 입니다.
	네 번째 항은 $b(3) * (-3) = 135 * (-3) = -405$ 입니다.
순서의 네 번째 항은 무엇인가요? 단계별로 생각하세요.	

LLM Customization (Prompt Eng.): 예시를 활용



You

다음의 대화를 참고하여 제자가 "인생이란 무엇입니까?" 라고 물었을 때의 답을 문답형 태로 작성해 줘

.....

제자: "스승님, 세상에서 가장 중요한 것은 무엇입니까?"

스승: "가장 중요한 것은 마음의 평화다. 부와 명예, 지식은 모두 변할 수 있으나, 마음속의 평화는 영원히 너의 안내자가 될 것이다. 이것이 있으면 어떤 환경에서도 너의 본성과 조화를 이룰 수 있으니, 항상 마음을 가다듬고, 평화를 추구하라."

.....

< 2 / 2 >



ChatGPT

제자: "인생이란 무엇입니까?"

스승: "인생은 마음의 평화를 추구하는 여정이다. 부와 명예, 지식 같은 외적인 것들은 시간이 지나 변할 수 있지만, 마음속의 평화는 변치 않는 너의 가장 큰 보물이 될 것이다. 인생에서 중요한 것은 외부 세계가 아닌 내면의 평안을 찾고 유지하는 것이니, 그것을 통해 너의 삶은 의미를 갖고, 진정한 행복을 찾을 수 있을 것이다."



LLM Customization (Prompt Eng.): 예시를 활용

프롬프트



모델 응답

이 명소는 브라질 리우데자네이루에 있는
그리스도 구속자 상입니다.

업데이트된 프롬프트

명소와 함께 도시를 확인합니다.



개선된 대응

도시: 리우데자네이루, 명소: 구세주 그리
스도 동상

도시: 로마, 명소: 콜로세움.



도시: 베이징, 명소: 자금성



명소와 함께 도시를 확인합니다.

LLM Customization (Prompt Eng.): 예시를 활용

프롬프트:

해당 주문에 대해 치즈버거, 버거, 감자튀김 또는 음료 필드가 있는 JSON 객체를 반환하며, 값은 수량입니다.

주문: 버거와 음료

응답:

예, 요청한 JSON 객체는 다음과 같습니다:

```
```json
{
 "burger": 1,
 "drink": 1
}
````
```

이 정보가 도움이 되기를 바랍니다!

모델이 예측한 대로 작동하지만 자연어로 안내를 작성하기 어려울 수 있습니다. 이 경우 예시 및 응답 프리픽스를 제공하여 모델이 이를 완료하게 할 수 있습니다.

프롬프트:

유효한 필드는 치즈버거, 햄버거, 감자칩, 음료입니다.

주문: 치즈버거와 감자칩 주세요

출력: {"cheeseburger": 1, "fries": 1}

주문: 버거 2개, 음료, 와플 주세요.

출력:

{"hamburger": 2, "drink": 1}

(text-bison@001)

LLM Customization (Prompt Eng.): 예시를 활용

모델에 필요한 모든 정보가 있다고 가정하는 대신 모델이 문제를 해결하는 데 필요한 프롬프트 안내와 정보를 포함할 수 있음.

프롬프트:

연결 해제된 Wi-Fi를 수정하려면 어떻게 해야 하나요? Google Wifi 라우터의 표시등이 노란색이며 느리게 깜박입니다.

응답:

Google Wifi 라우터의 표시등이 노란색이고 느리게 깜박일 때 연결이 끊어진 Wi-Fi를 복구하기 위해 할 수 있는 몇 가지 작업이 있습니다.

- 인터넷 연결을 확인합니다. 모뎀이 켜져 있고 케이블이 라우터에 제대로 연결되어 있는지 확인합니다. 모뎀을 다시 시작해 볼 수도 있습니다.
- Wi-Fi 연결을 확인합니다. Wi-Fi가 켜져 있고 해당 라우터 범위 내에 있는지 확인합니다. Wi-Fi를 다시 시작해 볼 수도 있습니다.
- 라우터를 다른 위치로 이동합니다. 라우터가 벽이나 다른 물체에 의해 차단된 위치에 있지 않은지 확인합니다.
- 라우터의 펌웨어를 업데이트합니다. Google Wifi 웹사이트로 이동하여 안내에 따라 라우터의 펌웨어를 업데이트할 수 있습니다.
- 라우터를 초기화합니다. 문제가 계속 발생하는 경우 라우터를 초기화할 수 있습니다. 초기화하면 모든 설정이 삭제되며 처음부터 다시 설정해야 합니다.

라우터를 초기화하려면 Google Wifi 앱에 로그인한 다음 설정 메뉴로 이동해야 합니다. 그런 다음 고급 탭을 선택하고 초기화 버튼을 클릭합니다. 라우터를 초기화할 것인지 확인해야 합니다. 초기화가 완료되면 처음부터 다시 설정해야 합니다.

위 단계를 따른 후에도 문제가 지속되면 Google 지원에 문의하여 도움을 받을 수 있습니다.

프롬프트:

아래 텍스트를 사용해 질문에 답합니다. 제공된 텍스트만 사용하여 응답합니다.

질문: 연결 해제된 Wi-Fi를 수정하려면 어떻게 해야 하나요? Google Wifi 라우터의 표시등이 노란색이며 느리게 깜박입니다.

텍스트:

색상: 노란색으로 느리게 깜빡임

의미: 네트워크 오류가 있습니다.

해결 방법:

이더넷 케이블이 라우터와 모뎀에 모두 연결되어 있고 두 기기가 모두 켜져 있는지 확인합니다. 각 기기를 분리했다가 다시 연결해야 할 수 있습니다.

색상: 노란색으로 빠르게 깜빡임

의미: 재설정 버튼을 누른 상태에서 이 기기를 초기화합니다.

해결 방법:

재설정 버튼을 계속 누른 상태에서 약 12초가 지나면 표시등이 노란색으로 바뀝니다. 노란색으로 바뀌면 초기화 버튼을 놓습니다.

색상: 노란색으로 고정

의미: 라우터를 초기화하는 중입니다.

해결 방법:

최대 10분 정도 걸릴 수 있습니다. 완료되면 기기가 자체적으로 재설정되고 흰색으로 깜빡임이 설정되어 설정할 준비가 되었음을 알립니다.

색상: 빨간색 점등

의미: 문제입니다.

응답:

Google Wi-Fi 라우터에 네트워크 오류가 발생한 것 같습니다. 본문에 따라 이더넷 케이블이 라우터와 모뎀에 모두 연결되어 있고 두 기기가 모두 켜져 있는지 확인해야 합니다. 각 기기를 분리했다가 다시 연결해야 할 수 있습니다.

LLM Customization (Prompt Eng.)

"""

{단축어}

#요약 : {주제}에 대해 '마크다운' 형태로 요약하기

#비교 : {주제}에 대해 '표' 형태로 가독성 있게 정리하기

#이메일 : [주제]에 대해 메일을 정중하게 '한글로 작성하기'

#번역 : {주제}에 대해 번역하기. 기본적으로 주어진 내용이 한글이면 영어로 번역하고, 영어면 한글로 번역하기.

#3문장: (주제)에 3문장으로 정리하기

#주간보고: {주제}에 대해 '서술형' 보고 형태로 정리하기,

#영단어: {주제}에 대한 정의, 실생활에서 자주 사용하는 예시 3개 보여주기

#OPIC : {주제}에 관련된 OPIC(영어 Speaking 시험) 질문 및 답변 예시 작성하기. 초등학생 수준의 단어를 사용하되, 고득점 받을 수 있게 씩 넣기, 고급표현에 대해서는 별도로 추가 설명하기

#영화 : {주제}에 출연 배우, 명대사, 관람객 평점, 유사한 영화 추천

"""

위에 정의한 {단축어}와 제시하는 (주제)를 입력하면 단축어에 사전 정의된 Prompt를 따라 답변해줘.

이해했으면 'OK'라고 답하고, 이후부터는 이 규칙을 따라서 답변해줘.

LLM Customization (Prompt Eng.) : 단축어 사용

”

{단축어}

#요약:{주제}에 대해 '마크다운' 형태로 요약하기

#비교:{주제}에 대해 '표' 형태로 가독성 있게 정리하기

#이메일:[주제]에 대해 메일을 정중하게 '한글로 작성하기'

#번역:{주제}에 대해 번역하기. 기본적으로 주어진 내용이 한글이면 영어로 번역하고, 영어면 한글로 번역하기.

#3문장:(주제)에 3문장으로 정리하기

#주간보고:{주제}에 대해 '서술형' 보고 형태로 정리하기,

#영단어:{주제}에 대한 정의, 실생활에서 자주 사용하는 예시 3개 보여주기

#OPIC:{주제}에 관련된 OPIC(영어 Speaking 시험) 질문 및 답변 예시 작성하기. 초등 학생 수준의 단어를 사용하되, 고득점 받을 수 있게 씩 넣기, 고급 표현에 대해서는 별도로 추가 설명하기

#영화:{주제}에 출연 배우, 명대사, 관람객 평점, 유사한 영화 추천

”

위에 정의한 {단축어}와 제시하는 (주제)를 입력하면 단축어에 사전 정의된 Prompt를 따라 답변해줘.

이해했으면 'OK'라고 답하고, 이후부터는 이 규칙을 따라서 답변해줘.



ChatGPT

OK

KI

You

#비교 : AWS와 MS Azure의 주요 서비스, 매출, 영업이익, 성장을

KI

ChatGPT

AWS와 MS Azure의 주요 서비스, 매출, 영업이익, 성장을에 대해 '표' 형태로 가독성 있게 정리하기:

| 항목 | AWS (Amazon Web Services) | MS Azure (Microsoft Azure) |
|--------|---------------------------|---|
| 주요 서비스 | EC2, S3, RDS, Lambda | Virtual Machines, Blob Storage, SQL Database, Functions |
| 매출 | \$62.2 billion (2022) | \$30.7 billion (2022) |
| 영업이익 | \$18.5 billion (2022) | \$15.1 billion (2022) |
| 성장을 | 37% (2022 YoY) | 50% (2022 YoY) |

참고사항: 표에 제시된 숫자들은 예시이며, 실제 최신 데이터를 확인하기 위해서는 각 회사의 공식 재무보고서나 신뢰할 수 있는 시장 조사 보고서를 참고해야 합니다.

LLM Customization (Prompt Eng.) : 다른문구사용

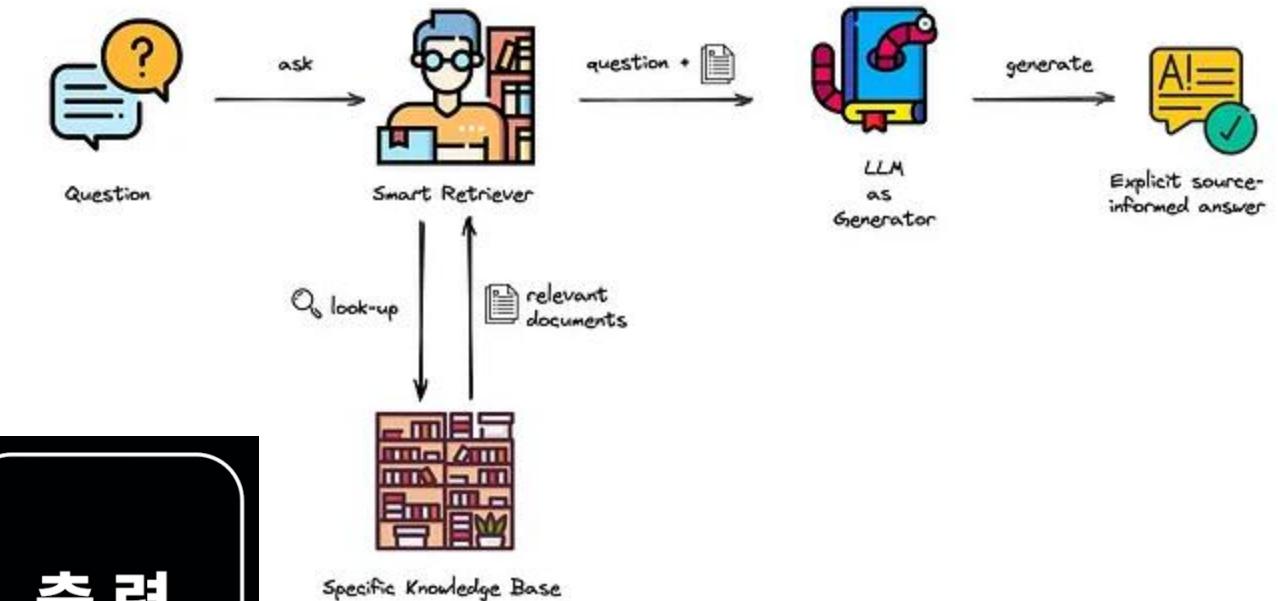
버전 1: 파이는 어떻게 굽나요?

버전 2: 파이의 레시피를 추천합니다.

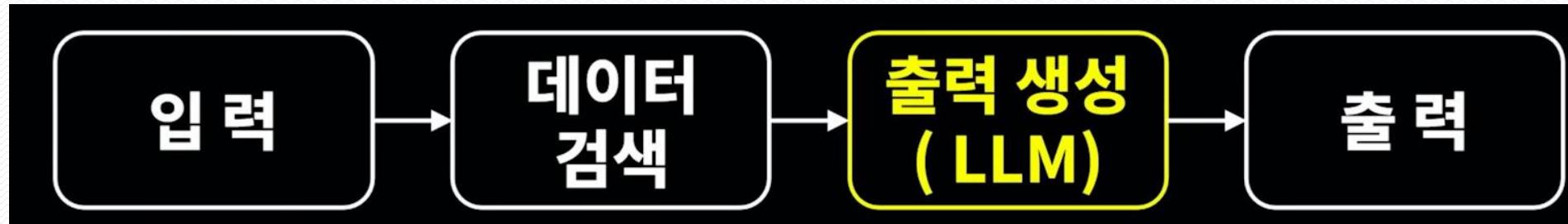
버전 3:
어떤 파이 레시피가 좋은가요?

LLM Customization (RAG)

- LLM 등력에만 의존하지 않고 모델 외부 자료 함께 사용
- 모델 학습없이도 최신 데이터 활용 가능
- 학습없이도 회사 정보 활용 가능



LLM Customization (RAG)



You

다음 정보를 참고하여 뉴진스의 최신곡을 알려줘

=====

제목, 앨범명 발매일

우리의 밤은 당신의 낮보다 아름답다, 마이데몬 (Original Soundtrack), 2023.11.24
우리의 밤은 당신의 낮보다 아름답다 (Inst.), 마이데몬 (Original Soundtrack), 2023.11.24
GODS, Gods Worlds Anthem 2023 (From 'League of Legends') [Epic Version] -
Single, 2023.10.05

아름다운구속, 너의 시간 속으로 OST, 2023.09.01

New Jeans, NewJeans 2nd EP 'Get Up', 2023.07.21

=====



ChatGPT

뉴진스의 최신곡은 "우리의 밤은 당신의 낮보다 아름답다"이며, 앨범명은 "마이데몬
(Original Soundtrack)"으로, 발매 일은 2023년 11월 24일입니다.

LLM Customization (PEFT)

■ Adapter Modules

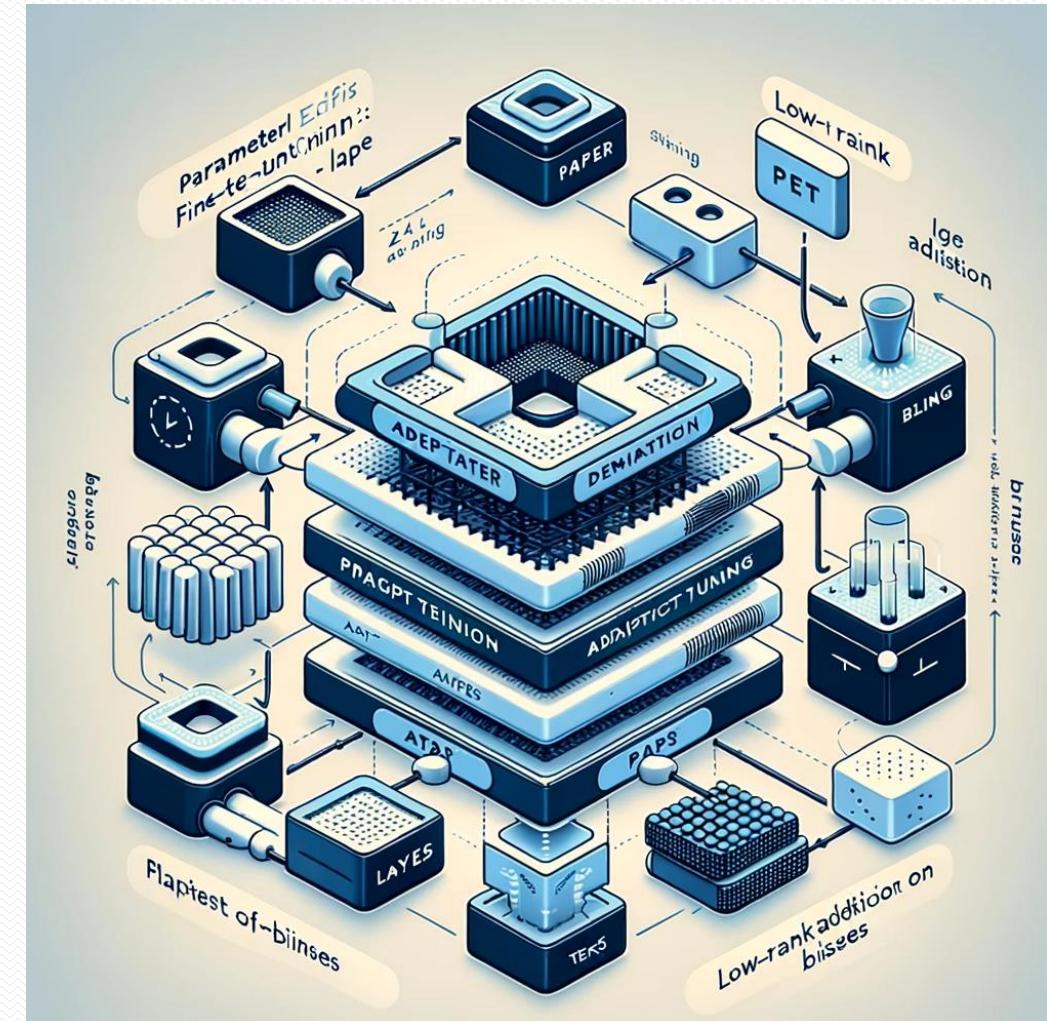
- Adapter 모듈은 사전 학습된 모델의 각 레이어 사이에 작은 신경망을 삽입하는 방식
- 원본 모델의 파라미터는 고정하고, 삽입된 Adapter만을 학습하여 특정 작업에 맞게 모델을 미세 조정
- 파라미터 수를 크게 증가시키지 않으면서도 좋은 성능을 달성

■ Low-Rank Adaptation

- 모델의 가중치 행렬에 저랭크(low-rank) 행렬을 적용하여 파라미터 수를 줄이는 방식
- 기존 모델의 복잡성을 유지하면서도 효율적으로 파라미터 수를 줄일 수 있는 방법

■ BitFit

- 매우 작은 부분집합의 파라미터만을 미세 조정하는 방법
- 예시: 바이어스 파라미터만을 학습
- 계산 비용을 크게 줄이면서도 유의미한 성능 개선을 달성



LLM 모델의 주요 tuning parameter 이해

<https://platform.openai.com/playground>

Temperature

- temperature 매개변수는 모델이 생성할 텍스트의 무작위성을 조정
- 낮은 temperature 값은 보다 확실한(덜 다양한) 텍스트를 생성하는 반면, 높은 temperature 값은 더 다양하고 예측 불가능한 텍스트를 생성
- 예시: 모델에게 "고양이는"이라는 문구로 문장을 시작하라고 요청
 - temperature=0.1 (낮은 값): 모델은 아마도 "고양이는 귀엽다"와 같이 일반적이고 예측 가능한 문장을 생성
 - temperature=2.0 (높은 값): 모델은 "고양이는 파란색 눈물을 흘리며 재즈 음악을 연주한다"와 같은 매우 독창적이고 예측하기 어려운 문장을 생성

Top-p

- 모델이 다음 단어를 선택할 때 고려할 확률 분포의 부분 집합을 결정
- 예시: "고양이는" 이후의 단어를 선택하는 상황. 모델의 예측에 따르면, "잠을", "귀엽다", "뛰어다닌다", "낮잠을", "우아하다"가 다음 단어로 나올 가능성이 가장 높음
 - top-p=0.8: 이 경우, 모델은 "잠을", "귀엽다", "뛰어다닌다"와 같이 누적 확률이 80%에 도달할 때까지의 단어들만 고려할 것입니다. 이 범위 내에서 무작위로 다음 단어를 선택
 - top-p=0.95: 더 많은 단어, 예를 들어 "잠을", "귀엽다", "뛰어다닌다", "낮잠을"까지 포함될 수 있으며, 이는 생성된 텍스트에 더 많은 다양성을 제공



THANK YOU