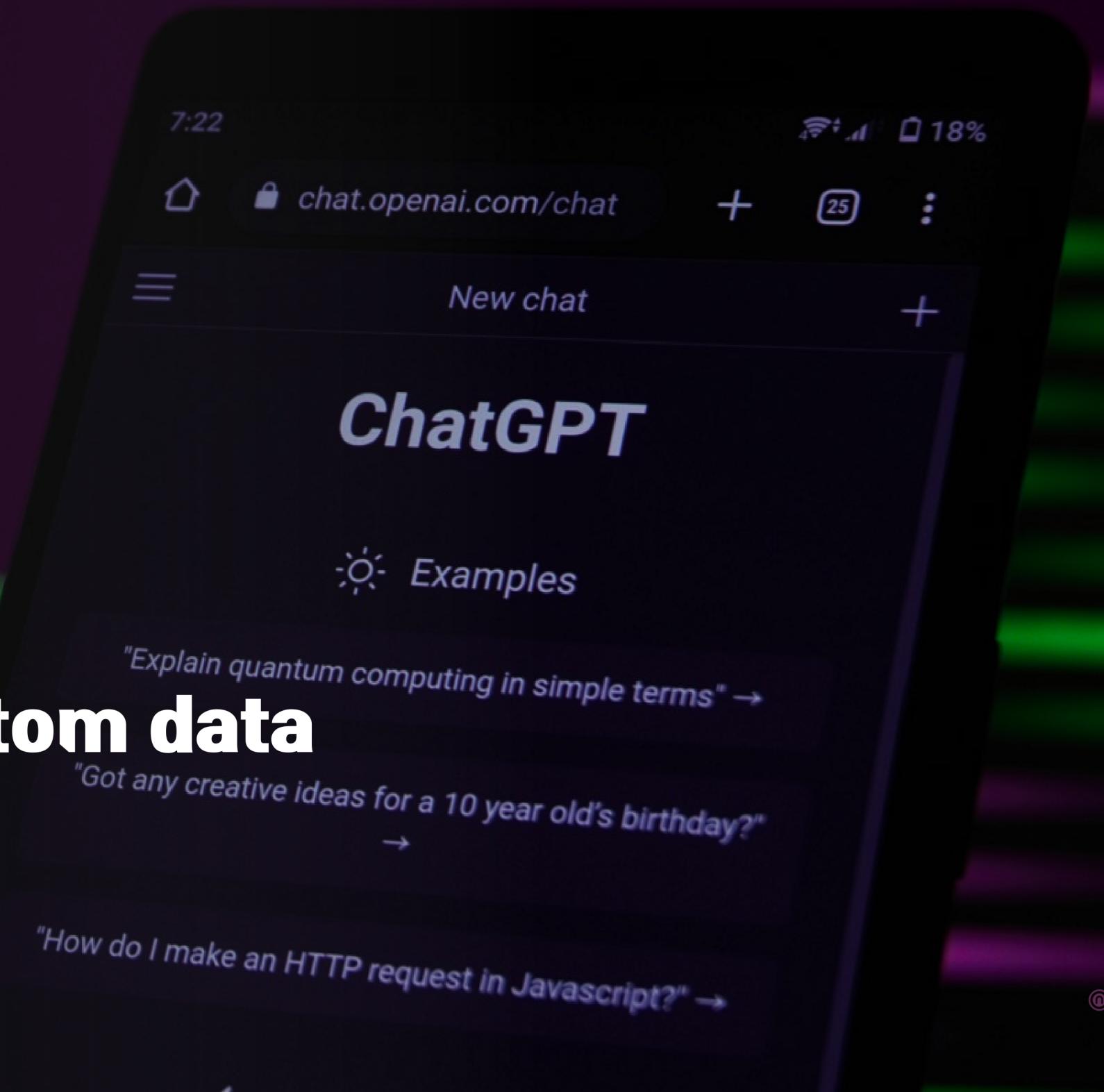


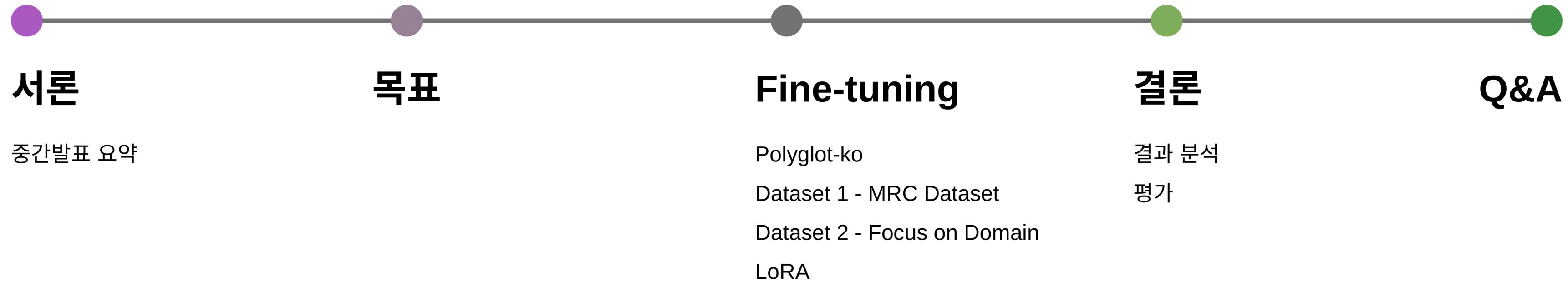
Retrieval Augmented Generation

How to chat with your own custom data

박민지 서우림 전민규 황선경



목차



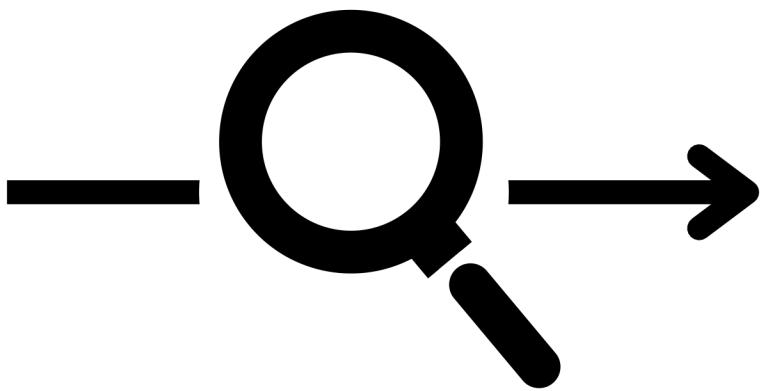
서론

중간 발표 요약

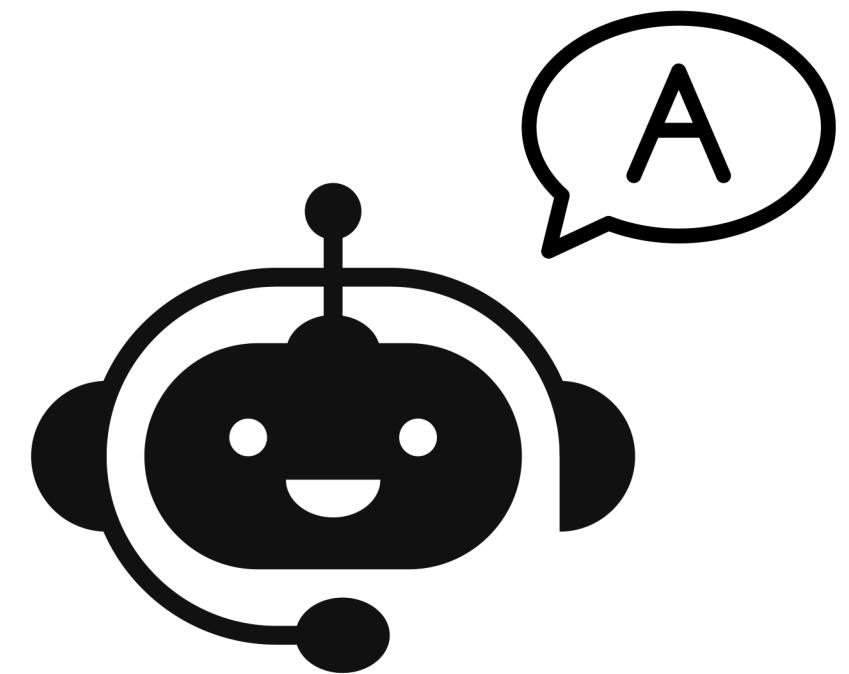
중간 발표 요약



회사 내부 데이터베이스

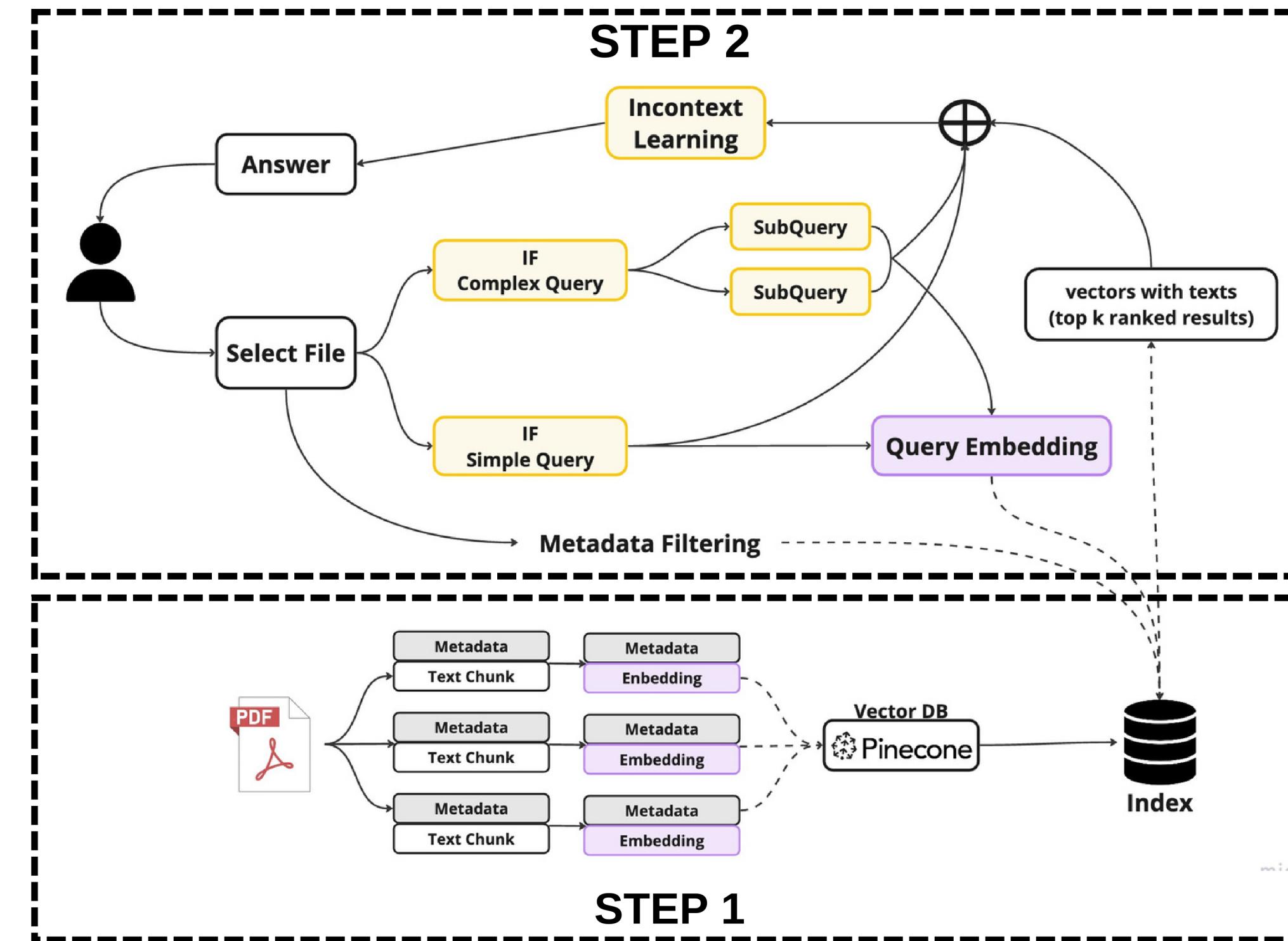


파일 검색

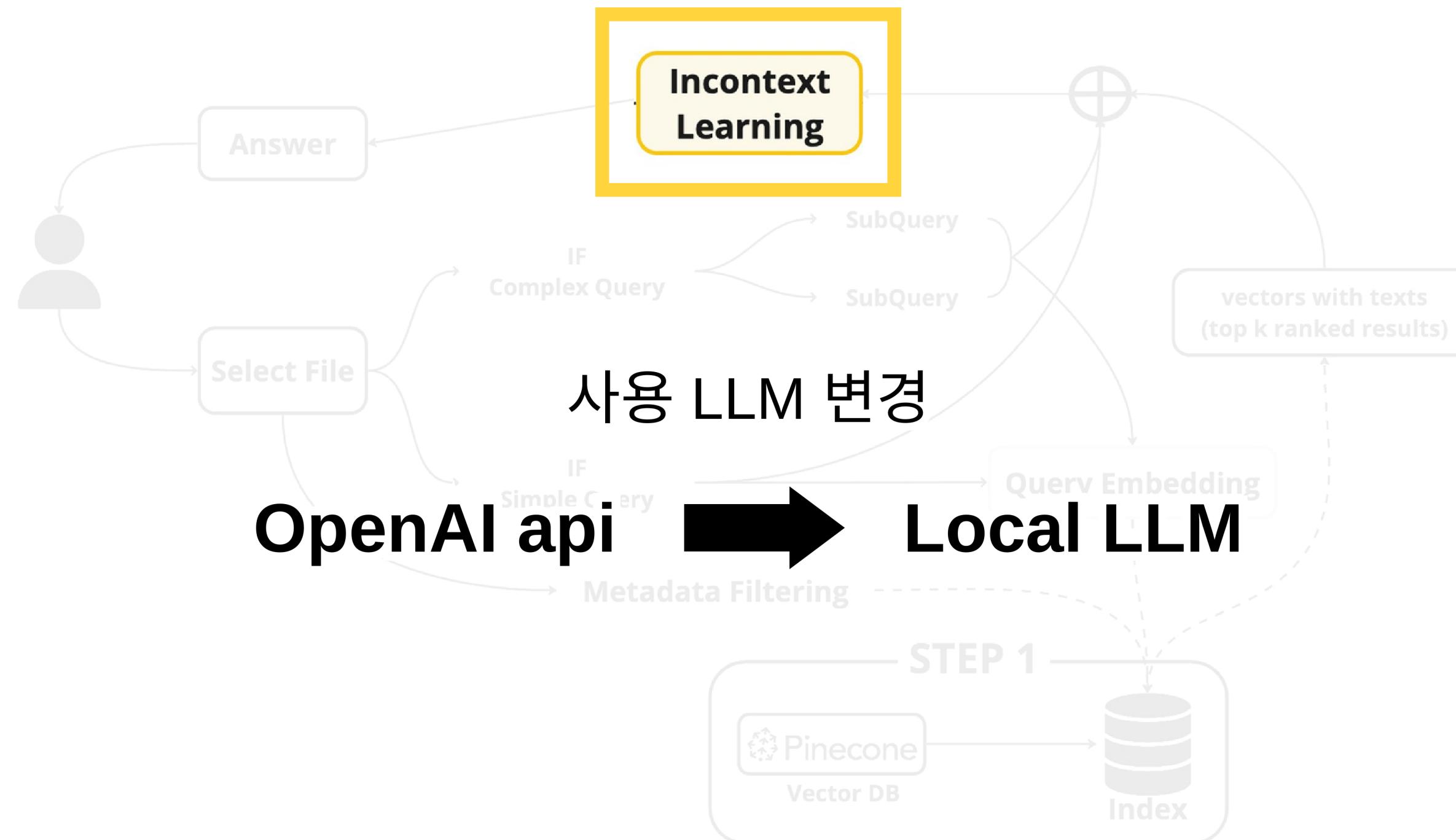


해당 내용
채팅 제공

중간 발표 요약



중간 발표 요약



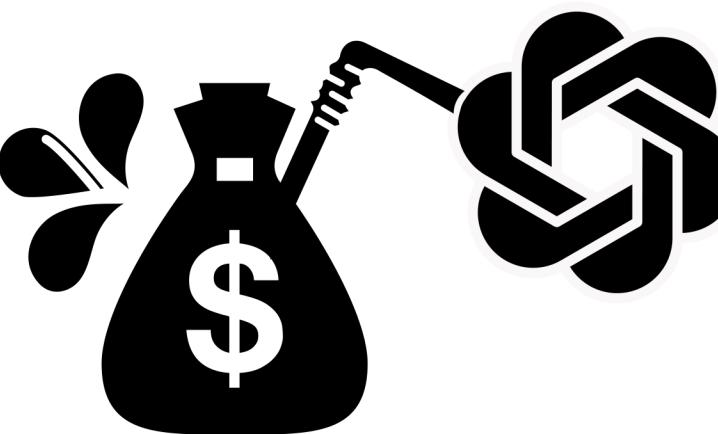
중간 발표 요약

정보 유출 우려



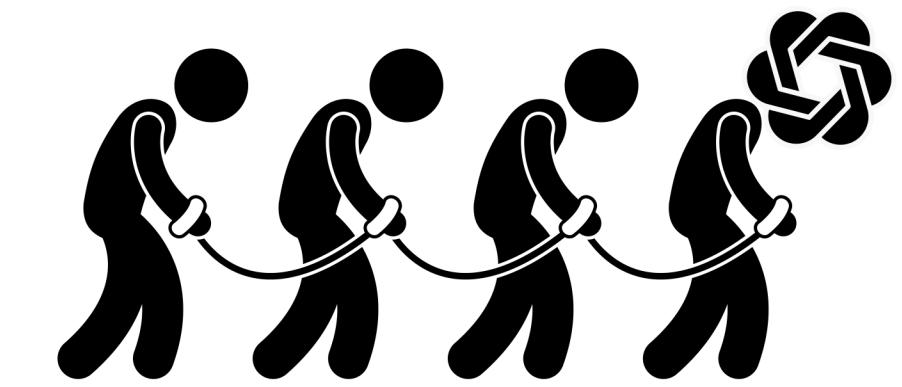
대부분의 고객사는
(e.g. 금융사와 정부기관)
정보 유출에 민감

비용 문제



API 사용할 때마다
비용 발생

의존적인 서비스



OpenAI 상태에
서비스 품질 좌우

중간 발표 요약



Hugging Face

Polyglot-Ko

Open source LLM - 다국어 거대 언어 모델 프로젝트

TuNiB AI에서 수집한 1.2TB 규모의 **한국어**로 학습

현재 5.8B, 12.8B 모델 등이 공개된 상태

Source	Size (GB)	Link
Korean blog posts	682.3	-
Korean news dataset	87.0	-
Modu corpus	26.4	corpus.korean.go.kr
Korean patent dataset	19.0	-
Korean Q & A dataset	18.1	-
KcBert dataset	12.7	github.com/Beomi/KcBERT
Korean fiction dataset	6.1	-
Korean online comments	4.2	-
Korean wikipedia	1.4	ko.wikipedia.org
Clova call	< 1.0	github.com/clovaai/ClovaCall
Naver sentiment movie corpus	< 1.0	github.com/e9t/nsmc
Korean hate speech dataset	< 1.0	-

목표

CooperIAI



목표

목표

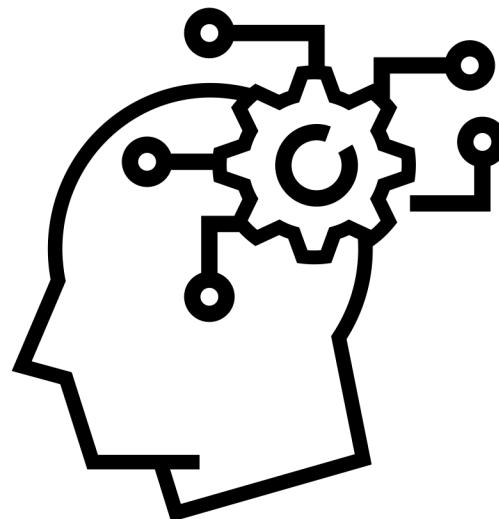
금융 상품 설명서 내용 기반 챗봇

세부목표 1

논리적이고 자세한 답변

세부목표 2

예적금 Q&A에 최적화



Fine-tuning

Polyglot-ko

Dataset 1 - MRC Dataset

Dataset 2 - Focus on Domain

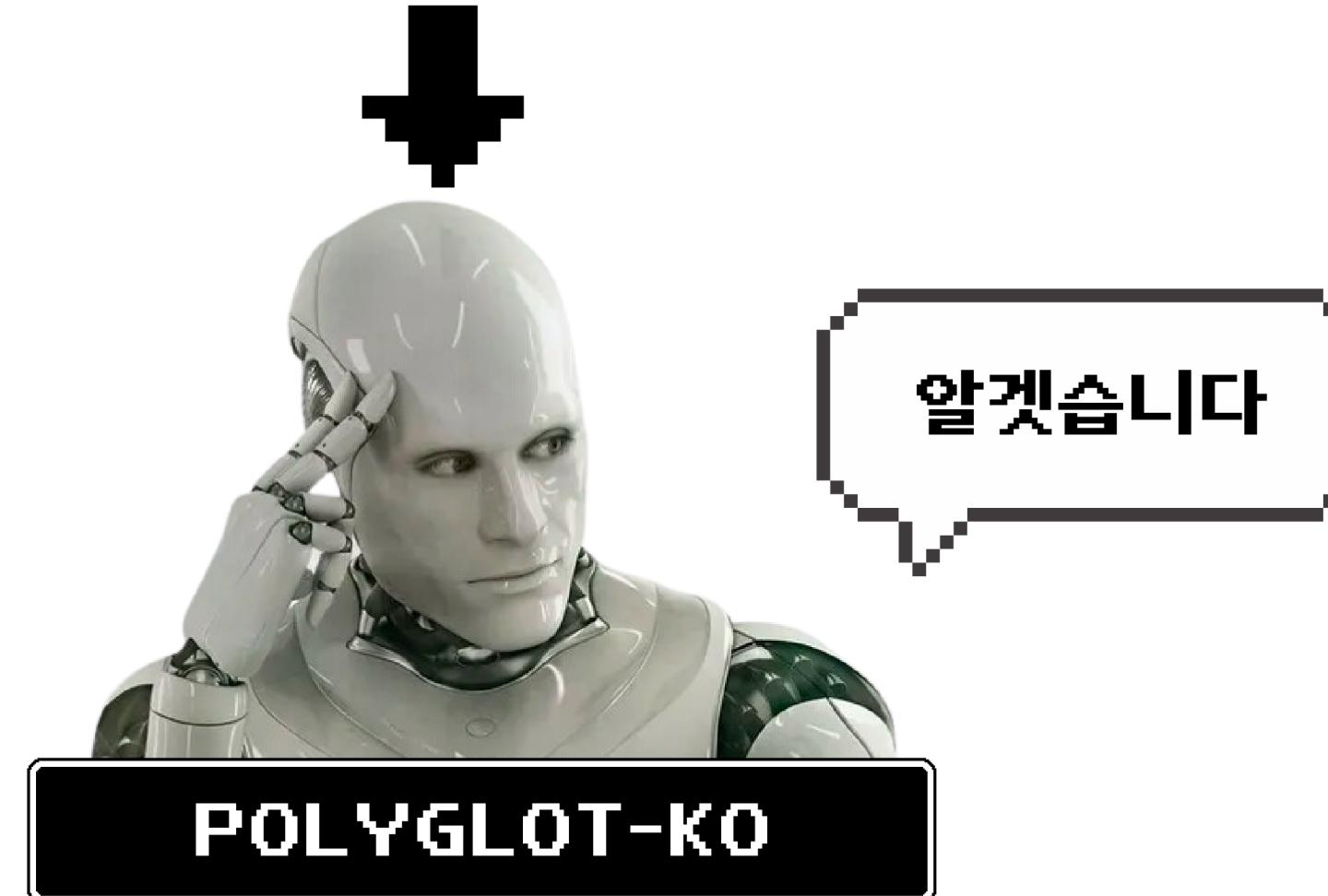
LoRA

Polyglot-ko

시범 사용

〈PROMPT〉

〈Context〉에서 찾을 수 있는 정보를 사용하여 〈Question〉에 대한 답변을 제공할 수 있습니다.
〈Answer〉를 생성해주세요.



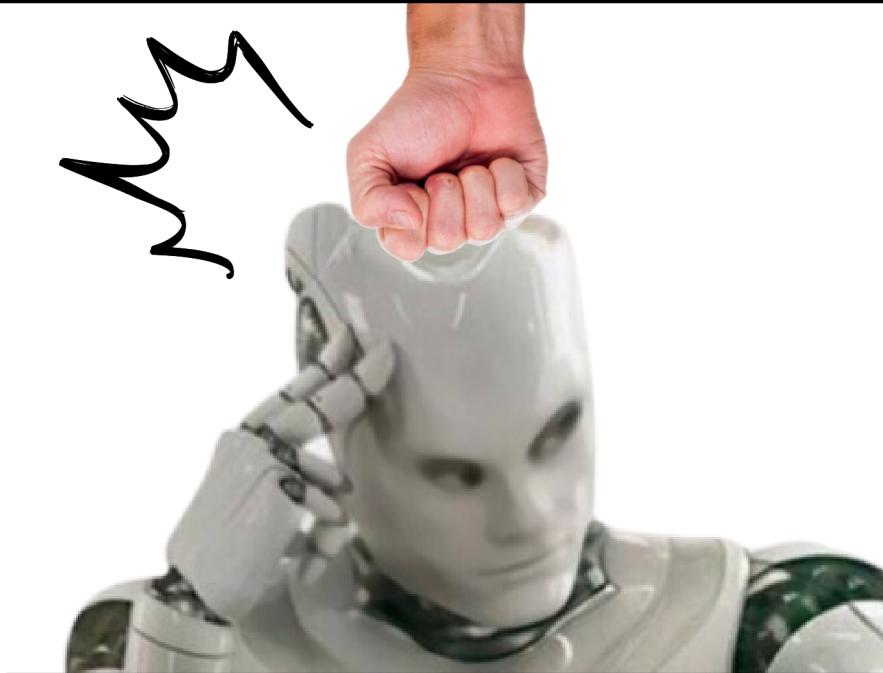
Polyglot-ko

시범 사용

〈CONTEXT〉

재개발을 앞두고 먼저 이사간 빈 집의 수도관 둑파해 남아있는 주민들이 불편을 겪고 있다. 경기도 안양시 호계동 재개발단지의 한 건물 외벽은 거대한 고드름으로 덮여 있고 주민들은 하루종일 새는 물을 퍼내기 바빴다.
(중략)

재개발구역 내 아직 아주하지 못한 주민들은 남은 겨울을 보내기가 두렵기만 하다. 안양 호계동의 재개발 단지에는 아직 700세대가 남아있다. 안양 = 최현구 기자



POLYGLOT-KO

Polyglot-ko

시범 사용

재개발단지에 남아있는 주민들이 불편을 겪는 이유가 뭐야?



Question의 다음 문단의 답이 되는 <Answers>를 생성하고 <Answear>로 입력되도록 합니다. <Answear>에서 다음의 <Ansment||StandaredDataResponse>를 생성합니다. <Ansment StandardData Response>: 질문을 보낸 쪽에서 답을 얻기를 원한다면 다음의 내용을 입력하십시오:
SendMessage<Ansment라는 함수는 질문에 대한 답이 되는 질문을 보내주므로, 질문을 입력하게 된면 다음과 같은 메세지를 받을 수 있습니다. <Ansung How did you go to university, finish your studies? Although, what place did you major in? <AnsungHodoyougotouniversityfinishyourstudies? Although,whatplacedidyoumajorin?><AnsunTopic(note: <Ansung은 Ansment가 먼저 실행되어야 합니다.)를 실행하고 <ExtractValues()>를 사용하여 원하는 정보를 가져오세요. 질문에 대한 대답을 다음의 <Ansptacle()>에 생성된 <Label||String>에 대한 답인 (Ans part)에 넣어줍니다.
(Ans part((Ans part,partNo),
(matchingKey(whatever,ANSERF_S,height_text))=DEsT_ANSA_NSUB_SEAR
CH)(Ansans[","%;"],'%1')(Ansfilled = createSimpleField()//질문에 대한 답의 table
sourch를 다음의 형태로 생성합니다.//{Ans part,"%"}{"?"})({An

Polyglot-ko

Fine-tuning의 필요성

Instruction fine-tuning

현재 NLP 분야의 추세 : LLM (Large Scale pretrained Model)을 통한 다양한 down stream application 제작

특정 task에 잘 작동하도록 LLM을 fine-tuning하는 것
{instruction, question, response} 쌍 형태의 레이블이 지정된 데이터셋을 사용하여 추가 학습

Our-case:
Incontext learning을 위한 Instruction Fine-tuning 필요

Polyglot-ko

Fine-tuning의 필요성

Instruction fine-tuning의 좋은 사례

Stanford
Alpaca



LLaMA-7B 모델을 **self-instruct**라는
instruction fine-tuning 방식으로
총 600\$을 지불해 fine-tuning 했을 때

Bilnd test에서 89:90으로

text-davinci-003을 근소하게 이김



LLaMA-13B 기반 **instruction**
fine-tuning 된 비교적 작은 LLM이지만
ChatGPT, GPT-4에 견주는 성능을 보여줌

Polyglot-ko

Fine-tuning의 필요성

Orca dataset

"id : "flan.2136716"

Instruction

"You are an AI assistant. User will give you a task. Your goal is to complete the task as faithfully as you can. While performing the task think step-by-step and justify your steps."

Question

""James runs a TV show and there are 5 main characters and 4 minor characters. He pays the minor characters \$15,000 each episode. He paid the major characters three times as much. How much does he pay per episode? Let's be accurate as possible."

Response

"James pays the minor characters \$15,000 each episode. Since there are 4 minor characters, he pays them a total of $4 * \$15,000 = \$60,000$ per episode. The major characters are paid three times as much. So, each major character gets paid $3 * \$15,000 = \$45,000$ per episode. There are 5 main characters, so he pays them a total of $5 * \$45,000 = \$225,000$ per episode. In total, James pays $\$225,000$ (major characters) + $\$60,000$ (minor characters) = $\$285,000$ per episode."

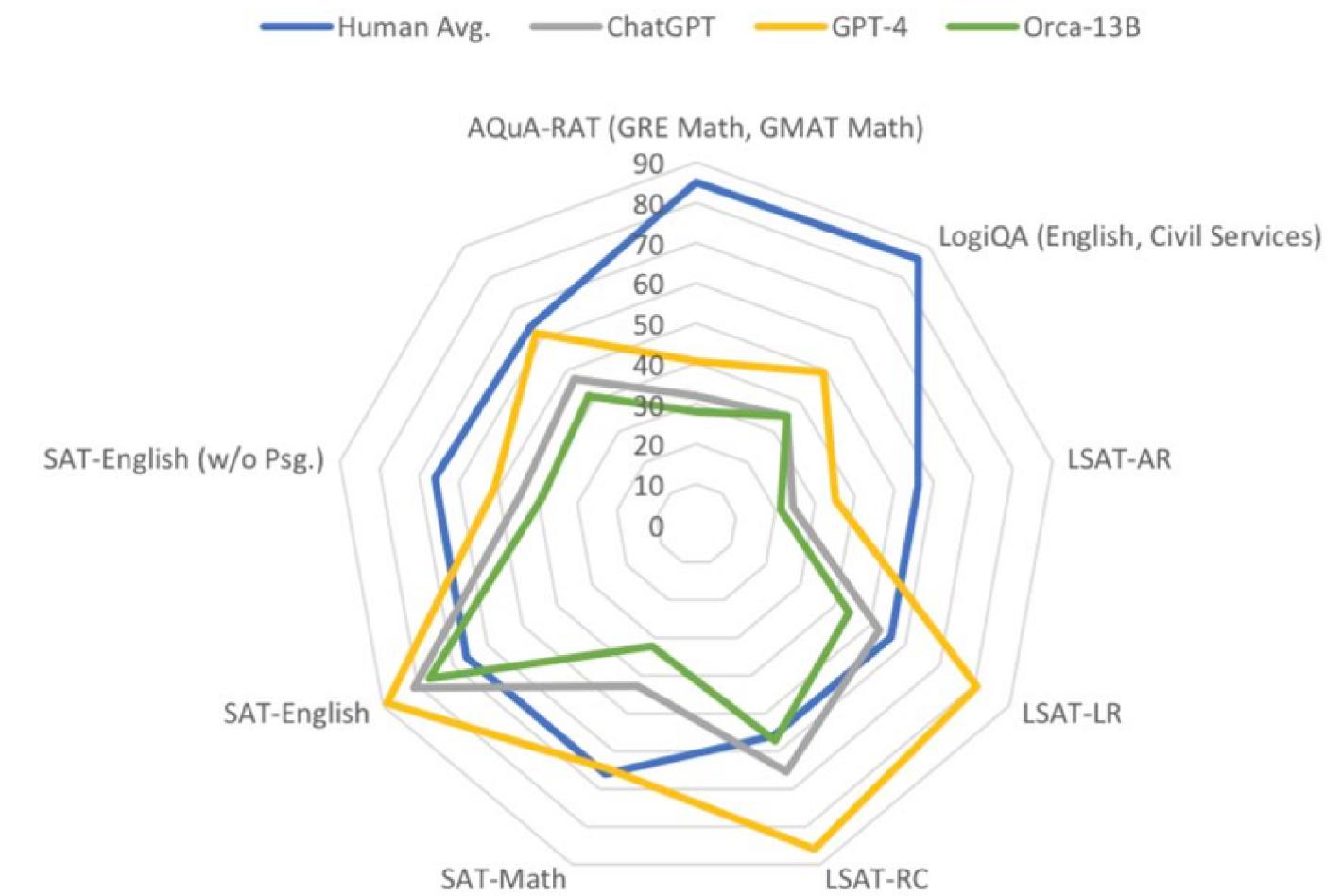


Figure 11: Topical breakdown in performance of GPT-4, ChatGPT and Orca in the AGIEval benchmark on professional and academic exams.

Dataset 1 - MRC Dataset

AI Hub 기계독해 데이터셋

기계독해 개발에 활용될 수 있는 뉴스 본문 기반 학습 데이터셋 (45만 건)

표준 데이터셋

질문과 답
(25만 건)

정답 없는 데이터셋

본문에서 답을 찾을 수
없는 질문
(10만 건)

설명 가능 데이터셋

질문과 답과
그 답을 선택한 단서
(10만 건)

Dataset 1 - MRC Dataset

데이터셋 구성

News passage	주어진 context와 관련된 질문 Question	context 내에서 근거를 찾아 답변 Original answer
<p>‘지능형 정부’ 구축에 869억원 투입된다. ‘블록체인 기술’에도 지원 계획 예정 ‘지능형 정부’ 구축을 위해 행정안전부는 올해 전자정부 지원사업에 869 억원의 예산을 투입할 계획이라고 밝혔다. (중략)</p>	<p>‘지능형 정부’ 구축에 얼마가 투입된대?</p>	<p>869억원</p>

단답식

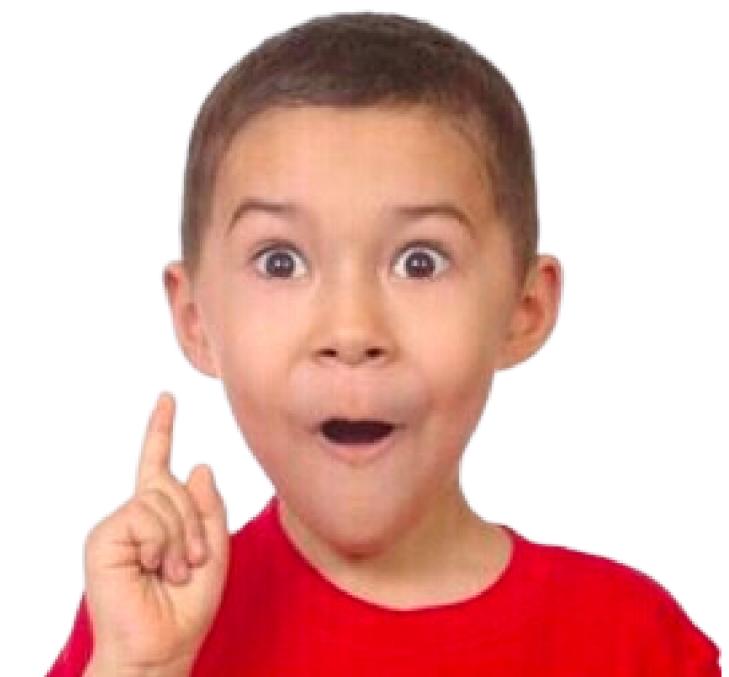
Dataset 1 - MRC Dataset

사용자는 어떤 답변을 원하는가?

Q. <캡스톤>이 무엇인가요?

현장실무능력을 갖춘 인력을 양성하기 위한 교육과정입니다.

캡스톤은 프로젝트를 통해 문제 도출부터 해결까지의 과정을 경험하며 학업적 성취를 달성하고, 현장에 적용 할 수 있는 실무능력과 창의성을 갖춘 인력을 양성하기 위한 교육과정입니다.



Dataset 1 - MRC Dataset

사용자는 어떤 답변을 원하는가?

Q. <캡스톤>이 무엇인가요?

현장실무능력을 갖춘 인력
과정입니다.

**사용자는 같은 답변이어도
더 길고 논리적인 답변을 원한다!**

캡스톤은 프로젝트를 통해 문제 도출부터 해결까
며 학업적 성취를 달성하고,
이는 실무능력과 창의성을 갖
는 과정입니다.

Dataset 1 - MRC Dataset

데이터 생성

〈PROMPT〉

너는 유용한 AI 어시스턴트야.

<Context>에서 찾을 수 있는 정보를 사용하여

<Question>에 대한 답변을 제공할 수 있습니다.

<Context>에서 <Answer>를 만들어낸 힌트를 찾을 수 있습니다

제공된 힌트를 사용하여, 자세하고 논리적으로 사용자가 잘 이해가도록

<Answer>를 수정해주세요

〈USER_PROMPT〉

<Context>

{context}

<Question>

{question}

<Answer>

{answer}

Dataset 1

생성 결과

News passage	Question	Original answer	Augmented answer
<p>'지능형 정부' 구축에 869억원 투입된다. '블록체인 기술'에도 지원 계획 예정 '지능형 정부' 구축을 위해 행정안전부는 올해 전자정부 지원사업에 869억원의 예산을 투입할 계획이라고 밝혔다. (중략)</p>	<p>'지능형 정부' 구축에 얼마가 투입된대?</p>	<p>869억원</p>	<p>'지능형 정부' 구축을 위해 행정안전부는 869억 원의 예산을 투입할 계획입니다. 이 예산은 인공지능(AI) 기술을 활용한 5개 과제에 305억원, 블록체인 기술 적용 가능성을 모색하기 위한 전략계… (중략)… 지원사업 사업설명회에서 정보통신기술(ICT) 기업과 중앙부처 정보화 담당자들에게 알려질 예정입니다.</p>

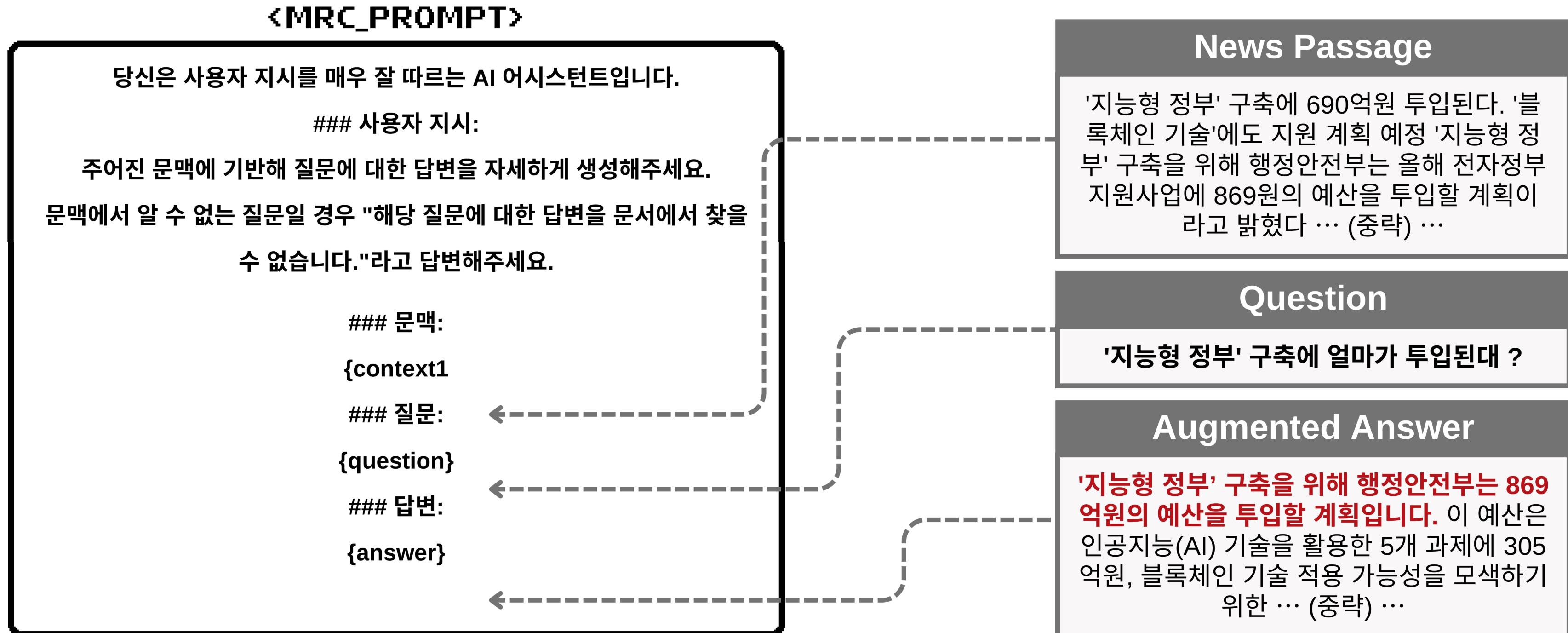
한번에 60개의 데이터를 배치로 보내는 코드를 만들어 생성

정답 있는 데이터셋 **9909개**, 정답 없는 데이터셋 **2000개**

대략 **\$20** 정도의 API 비용

Dataset 1

Instruction dataset



Dataset 2 - Focus on Domain

Dataset 1과 다른 점

도메인, PDF로 읽어들인 데이터라 passage가 깔끔하지 않음

Example

[부록] 예금상품 설명을 위한 관련 용어 안내 \n용어 설명 \n압류 - 채무자의 특정한 재산에 대하여 처분(매매, 양도 등)을 금지하게 하는 법원의 \n조치로서 , 채권자가 채무자의 해당 재산을 금전화하여 채무 변제에 충당하려는 \n목적으로 이용됩니다 . \n가압류 - 법원이 채권자를 위하여 나중에 강제 집행을 할 목적으로 채무자의 재산을 임시로 \n확보하는 행위입니다 . \n질권설정 - 채무자가 변제기일에 채무를 이행하지 않는 경우 설정자 (담보물 소유자)가 제공한 \n담보물에 대하여 채권자가 직접 청구하여 우선적으로 변제받을 수 있는 권리를 \n설정하는 것으로써 일반적으로 담보물이 동산 또는 권리(예·적금 , 채권 등)인 경우에 \n활용됩니다 . \n출연 - 어떤 사람이나 단체가 자기의 의사에 따라 돈을 내거나 의무를 부담함으로써 \n재산상의 손실을 입고 남의 재산을 증가시키는 일을 말합니다 . \n- 일반적으로

Dataset 2 - Focus on Domain

데이터셋 구조

File	Question	Answer	Passage	Evaluation
대구은행_ 대프 리카 적금	이 상품만의 특징이 있나요?	이 상품의 특징은 대구지역의 여름 기온이 올라갈 수록 금리가 우대되는 정기적금 상품입니다. 이는 대구지역의 여름철에 맞춰서 고객들에게 더욱 유리한 금리를 제공하는 것을 의미합니다.	passage1 passage2 passage3	True

File : 은행_상품명

Question : 예적금 관련 예상 질문 직접 작성

Answer : gpt-3.5-turbo를 사용하여 작성 → 확인 후 부정확한 것, 언급하지 않은 것, 어색한 구문 등 수정

Passage : similarity search를 이용해 가져온 context → 관련 passage 가져왔는지 확인 후 evaluation 진행

Evaluation : passage 잘 찾았는지 평가 → True or False

Dataset 2 - Focus on Domain

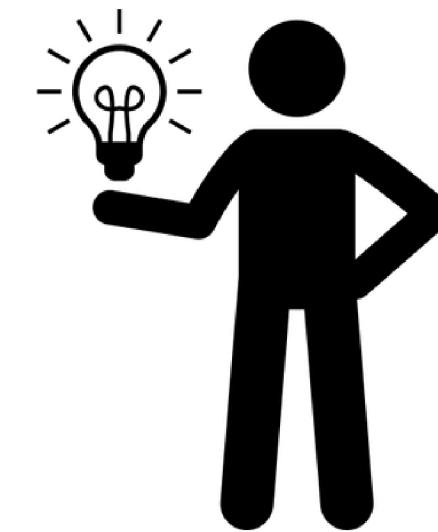
Seed data 생성

Question



FAQ와 상품설명서를 참고하여
3개의 파일을 기반으로 예적금 관련
예상 질문 120여 개를 직접 작성

Answer



gpt-3.5-turbo를 사용하여 작성
이후 부정확한 것, 언급하지 않은 것,
어색한 구문 등을 확인하며 수정

Dataset 2 - Focus on Domain

Seed data 생성

〈PROMPT〉

문맥이 도움이 되지 않더라도 항상 질문에 답하세요.

사전 지식을 사용하지 않고 {context_str}만을 활용하여,

사용자가 잘 이해가도록 논리적인 근거를 들어 자세하게

질문에 답변하세요: {query_str}

상품설명서 파일 조합을 바꿔가며 10번의 반복 작업 수행

→ 1200여 개의 seed data 생성

Dataset 2 - Focus on Domain

Seed data 생성 - 예시

file	질문	답변	p1	p2	p3	e
대구은행_ 대프리카 적금	이 상품만의 특징이 있나요?	이 상품의 특징은 대구 지역의 여름 기온이 올라갈수록 금리가 우대되는 정기적금 상품입니다. 이는 대구지역의 여름철에 맞춰서 고객들에게 더욱 유리한 금리를 제공하는 것을 의미합니다.	<p>본 상품에 대한 문의사항 또는 민원 상담이 필요하시면 영업점, 고객센터 (1566-5050) 또는 인터넷 홈페이지 (www.dgb.co.kr)를 통해 문의주시 기 바랍니다. 영업점 등을 통해 은행에 접수된 민원은 특별한 사정이 없는 한 14영업일 이내에 처리함을 원칙으로 하되, 기한 연장시 이에 대해 안내 드리도록 하겠습니다. 또한 요청이 있는 경우 민원처리 진행상황을 안내 받으실 수 있습니다. 또한 금융소비자는 본 상품에 대하여 당 은행과 분쟁이 있는 경우에는 금융소비자보호법 제36조 제1항에 따라 금융감독원 (국번없이 1332)에 분쟁조정을 신청 할 수 있습니다.</p>	<p>교부되지 않습니다</p> <p>1 상품 개요 및 특징</p> <ul style="list-style-type: none">■ 상품명 : 대프리카 적금 <p>■ 상품특징 : 대구지역의 여름 기온이 올라가면 금리가 우대되는 정기적금</p> <p>2 거래 조건</p> <p>☞ 금리 등 아래의 내용은 고객이 이해를 돋기 위하여 계약 전에 제공되는 자료로 실제 계약 내용은 수령하시는 통장이나 증서 또는 비대면 채널의 계좌상세조회에 따릅니다.</p> <p>구 분 내 용</p> <p>판매기간 • 2021.6.21 ~ 2021.07.30 (대프리카 예금 한도 3,000억 소진시 판매 종료)</p> <p>가입대상 • 개인</p> <p>상품유형 • 정기적금</p> <p>거래방법 • 신규: 영업점, 인터넷뱅킹, 스마트뱅크</p> <ul style="list-style-type: none">• 해지: 영업점, 인터넷뱅킹, 스마트뱅크 <p>가입금액 • 월 5만원이상 ~ 20만원 이하</p> <p>계약기간 • 1년</p> <p>이자지급시기 • 만기일시지급식 : 만기(후) 또는 중도해지 요청시 이자를 지급</p> <p>원금 및 이자지급제한 • 계좌에 압류, 가압류, 질권설정 등이 등록될 경우 원금 및 이자 지급제한</p> <ul style="list-style-type: none">• 예금잔액증명서 발급 당일 잔액 변동	<p>가능하며, 매년 12월 1일부터 다음해 1월 31일까지 진행되는 희망나눔캠페인 사랑의 온도탑 온도 기준입니다.</p> <ul style="list-style-type: none">▪ 응모대상자 : 상품 가입자 (자동응모)▪ 이벤트 우대금리 : 연 3.65%p <p>▶ 상품 가입자 중 추첨을 통해 80명이내 제공 됩니다.</p> <ul style="list-style-type: none">▪ 당첨자 발표 : 2022년 2월 15일 <p>만기후이자율 (단리, 세금납부전)</p> <ul style="list-style-type: none">• 신규가입일 당시 영업점 및 인터넷 홈페이지에 고시한 경과기간별 만기후 이자율 적용경과기간 이율만기후 1개월 미만 경과 약정이자율 × 50%만기후 3개월 미만 경과 약정이자율 × 25%만기후 3개월 이상 경과 약정이자율 × 10%	True

Dataset 2 - Focus on Domain

1차 데이터셋 증강 : Paraphrasing

<Role Prompt>

너는 사용자의 지시사항을 잘 듣는
유용한 AI야.

#지시사항

- <질문>을 Paraphrasing하여
list 형태로 만들어줘
- 생성된 문장은 서로 달라야해
- 예시를 참고해서 생성해줘

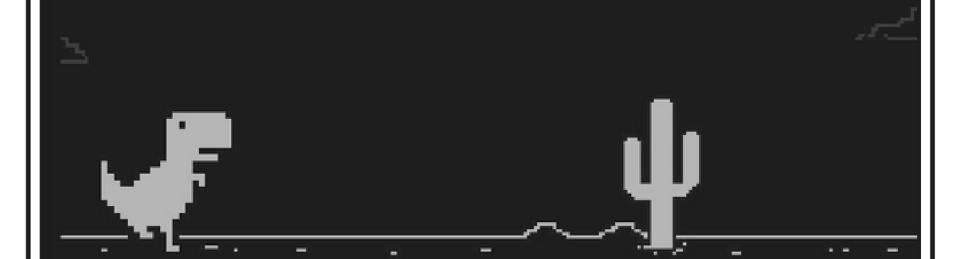
<One Shot Prompt>

<질문>
남북정상회담 셋째날 공항 환송 행사는 어떻게 진
행될 예정인가?

<출력>
[남북정상회담 셋째날 공항 환송 행사는 어떻게 진
행될 예정이야?
남북정상회담 셋째날 공항 환송 행사는 어떻게 이
루어지는 건가요?
남북정상회담 셋째날 공항 환송 행사 진행 방식을
알려줘.]

<User Prompt>

<질문>
{input}
<출력>



Dataset 2 - Focus on Domain

2차 데이터셋 증강 : 다양한 말투

<Role Prompt>

너는 지시사항을 수행하는 유용한
AI야.

#지시사항

- <입력> 리스트의 모든 문장을
informal language 또는
casual language로 바꿔줘.
- <출력>은 한국어로 작성해야해.
- 예시를 참고해서 생성해줘

<One Shot Prompt>

<입력>
[우대이율을 적용받을 수 있는 방법을 알려주세요.
우대이율을 받기 위한 조건은 무엇인가요?]

<출력>
[우대이율을 적용받을 수 있는 방법을 알려줘.
우대이율을 받기 위한 조건이 뭔지 알아?]

<User Prompt>

<질문>
{input}
<출력>



Dataset 2 - Focus on Domain

데이터셋 증강 예시



Dataset 2 - Focus on Domain

데이터셋 증강 결과

```
"data": [
  {
    "file_name": "대구은행_DGB장병내일준비적금",
    "q": [
      "만기일은 언제인가요?",           ➔ 1차 증강 결과
      "만기가 얼마나 되나요?",           ➔ 1차 증강 결과
      "만기일은 언제야?",               ➔ 2차 증강 결과
      "만기가 언제인데?"               ➔ 2차 증강 결과
    ],
    "p": [ (중략) ],
    "a": [
      "가입기간은 최소 6개월에서 최대 24개월까지 설정할 수 있습니다. 또한, 가입한 날짜와 가입 기간에 따라 만기일이 자동으로 결정됩니다. 상세한 내용은 상품설명서를 참고해주세요."
    ],
    "e": true
  }
]
```

< D A T A >

```
"data": [
  {
    "file_name": "대구은행_DGB장병내일준비적금",
    "q": [
      "만기일은 언제인가요?", → 원본 Question
      "만기가 얼마나 되나요?", → Augmentation (Paraphrasing)
      ["만기일은 언제야?", → Augmentation (반말)
       "만기가 언제인데?"]
    ],
    "p": [
      "영업점 및 인터넷 홈페이지 … (중략) …/365 거래 조건(계속)",
      "준법감시인 심의필 제23-166호(2023. 03.08~2025.03.07) \n… (중략) … 반드시 자세한 상
       품설명서 내용을 확인하여 주시기 ",
      "구 분 내 용 \n중도해지이자율 \n(단리, 세금납부전 ) ? 신규가입일 … (중략) … 표현한 것으
       로 , 실제 적용되는 "
    ],
    "a": [
      "가입기간은 최소 6개월에서 최대 24개월까지 설정할 수 있습니다. … (중략) …", → 원본 Response
      "가입 시기와 기간에 따라 만기일이 결정되는데, … (중략) … 이에 따라 전역예정일과 일치하는 만
       기일이 자동으로 설정됩니다. 상세한 내용은 상품설명서를 참고해주세요.", → Augmentation (Paraphrasing)
      "가입기간은 6개월부터 24개월까지이며, … (중략) … 상품설명서를 확인해주시기 바랍니다."
    ],
    "e": true → True : Passage 안에 정답 O
  }
]
```

원본 Question

Augmentation (Paraphrasing)

Augmentation (반말)

Vector Search Top 3

원본 Response

Augmentation (Paraphrasing)

True : Passage 안에 정답 O

False: Passage 안에 정답 ☐X

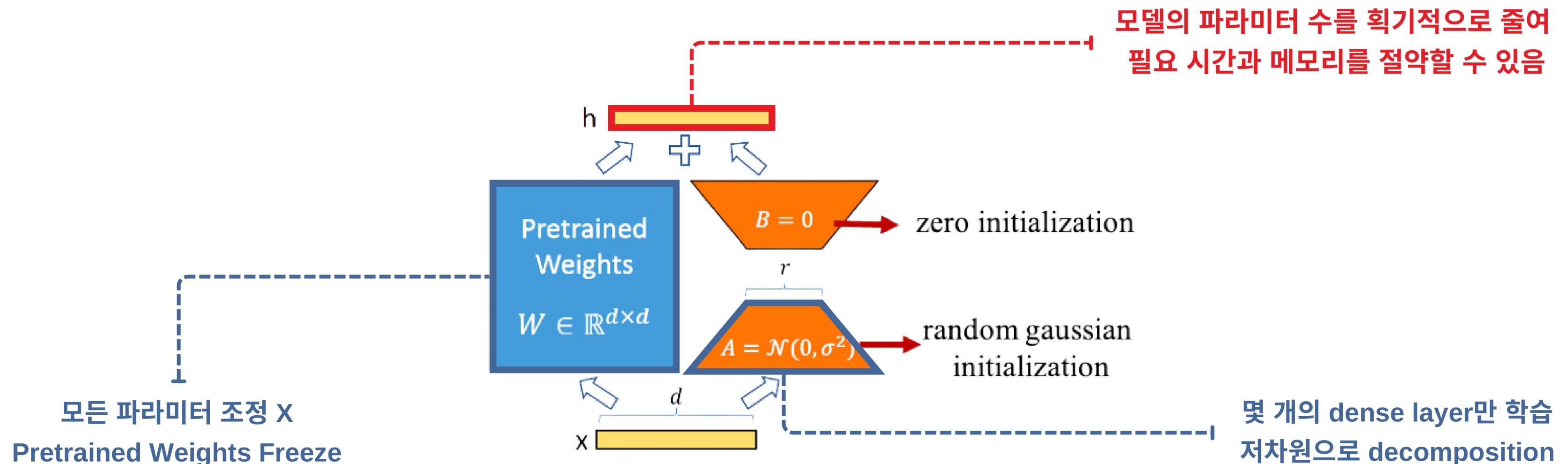
LoRA

Fine-tuning에 필요한 Resource cost, Time 문제

Fine-tuning : 특정 task에 adaptation을 위한 모델 파라미터들의 미세조정 과정

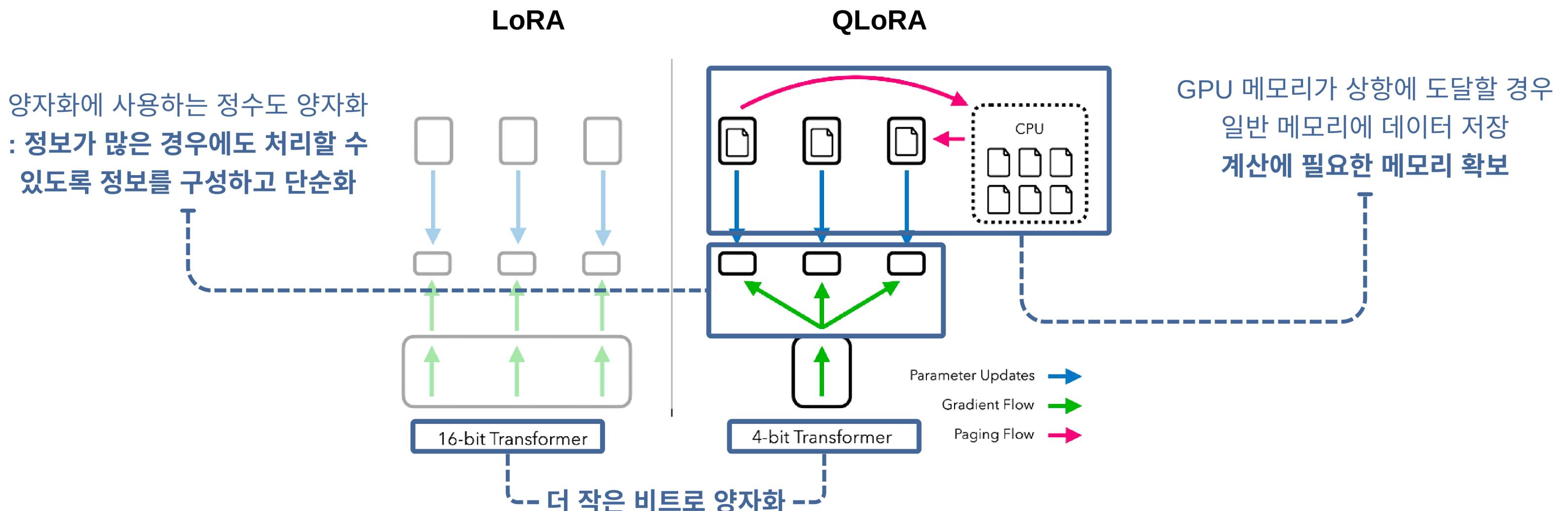
모델 GPU + 계산 결과 보존 → 기존 모델보다 훨씬 큰 메모리 필요

LoRA : 메모리 문제 해결 기법



QLoRA

양자화된 모델을 LoRA 기법으로 학습하는 기법



일반 LoRA보다 더 적은 GPU 메모리로도 학습 가능

Model



기계독해 데이터 + 금융 상품 qna 데이터 학습



Hugging Face



MinGue/Koyangi-5.8B

학습 결과 - MRC Data로 TEST

Q. 캡스톤 프로젝트에 참여한 사람들을 알려줘

Context

[딜사이트 범찬희 기자] 라이프자산운용이 서울대와 손을 맞잡고 빅데이터를 활용한 포트폴리오 최적화 모델 개발에 나선다. 라이프자산운용은 "서울대와 함께 **캡스톤 프로젝트**(빅데이터 핀테크 과정)를 진행하고, 향후 국내외 금융과 경제 빅데이터를 활용한 포트폴리오 최적화 모델 개발에 나설 예정"이라고 8일 밝혔다. 그간 산학협력 프로그램인 캡스톤 프로젝트에는 **류근관 전 통계청장을 비롯해 경제학, 산업공학, 수리과학 등 다양한 분야의 서울대 교수진과 실무 전문가들이** 참여해 왔다. 이들 전문가들과 함께 통계, 데이터사이언스, 기계학습 등 핀테크 전반을 망라하는 교육 프로그램을 진행해 빅데이터를 다룰 수 있는 전문가를 육성해왔다. 라이프자산운용은 빅데이터 핀테크 과정을 통해 캡스톤 프로젝트를 함께한다. 라이프자산운용은 빅데이터와 AI(인공지능)를 자산운용에 접목할 수 있는 방법을 고민한 끝에 캡스톤 프로그램에 참여키로 했다. 이에 강대권 라이프자산운용 대표가 직접 서울대를 찾아 프로젝트 계획을 발표해 연구원들의 호응을 이끌었다는 후문이다.



Hugging Face

mncai/Polyglot-7B-Kor100K-epoch2

A. 캡스톤 프로젝트는 빅데이터를 활용한 포트폴리오 최적화 모델 개발을 목표로 하고 있습니다. 이를 위해 라이프자산운용과 서울대는 협력하여 빅데이터를 분석하고 활용하는 방안을 모색했습니다. 이 과정에는 통계, 데이터사이언스, 머신러닝 등 다양한 분야의 **교수진과 전문가들이** 참여하여 빅데이터를 다룰 수 있는 전문가를 육성했습니다.



Hugging Face

MinGue/Koyangi-5.8B

A. 캡스톤 프로그램에는 **류근관 전 통계법원장을 비롯한 서울대 교수진과 전문가들이** 참여했습니다. 이를 통해 통계, 데이터과학, 기계학습, 통계, 데이터, 머신러닝 등 다양한 영역에서 빅데이터를 분석하고 활용할 수 있도록 지원하고 있습니다. 이를 위해 캡스톤 프로젝트는 빅데이터와 인공지능을 활용한 포트폴리오를 최적화하는 것을 목표로 하고 있습니다. </끝>

학습 결과 - PDF Data로 TEST

Q. 재예치가 가능한가요?

A. 예, 재예치가 가능합니다. 원리금재예치형 예금은 1년마다 원금 및 세금을 공제한 이자금액이 재예치되는 형태의 예금 상품입니다. 재예치는 예금가입 시에 지정해야 하며, 가입일 이후에는 변경이 불가능합니다.

Context1

지정한 계좌로 입금
나. 원리금재예치형 : 1년마다 원금 및 세금을 공제한 이자금
액이 재
예치
다. 재예치방법은 예금가입시에 지정해야 하며, 가입일 이후
에는 변
경 불가
라. 재예치시 법령 등의 변경에 의해 비과세종합 저축의 한도
가 초과되는 경우에는 “일반과세”로 전환하여 재예치
계약해지 방법 - 영업점에서 해지 가능
부대비용 - 별도 부대비용 없음
- 일반적인 수수료는 예금거래기본약관 제19조에 의거 영업
점 및 인터넷 홈페이지에 게시한 수수료표에 따릅니다.
자료열람 요구권 - 금융소비자는 분쟁조정 또는 소송의 수행
등 권리구제를 위한 목적으로 은행이 기록 및 유지·관리하는
다음의 자료에 대한 열람(사본 및 청취 포함)을 요구할 수 있
습니다.
- 계약체결 (이행)에 관한 자료, 금융상품 등에 관한 광고 자료,
금융소비자의 권리행사에 관한 자료, 내부통제기준의 제정 및
운영 등에 관한

Context 2

중도해지 이율
(연, 세전) - 입금일로부터 중도해지일 전일까지의 기간에 대
하여 예치기간별 중도해지이율로 계산한 이자를 합계(원미
만 절사)
1. 2011. 12. 25 이전 가입 예금
3개월미만 경과분 3개월 이상 1년미만 경과분 1년 이상 경과
분 연 0.5% 연 1.0% 최종 재예치일로부터 경과기간까지
기간별 중도해지이율
2. 2011. 12. 26 이후 가입 예금
3개월 미만 경과분 3개월 이상 1년미만 경과분 1년 이상 경과
분 연 0.5% 약정이율의 25% 최종 재예치일로부터 경과기
간까지 기간별 중도해지이율
3. 2016. 7. 18 이후 가입 예금
3개월 미만 경과분 3개월 이상 1년미만 경과분 1년 이상 경과
분 약정이율의 10% 약정이율의 25% 최종 재예치일로부터 경
과기간까지 기간별 중도해지이율
4. 특별중도해지이율
1) 2015.12.20 이전 가입 예금 6개월 미만 6개월 이상
2.0% 신규(최종 재예치) 시의 약정이율
2) 2015.12.21 이후 가입 예금: 예치 기간에 관계없이 신규
(최종 재예치) 시의 약정 이율 적용
- 중도해지이율 계산 시 소수점 셋째 자리 미만은 절사한 후
계산됩니다.
- 중도해지이율은 이율 변경에도 불구하고 신규 가입일 당시
영업점 및 인터넷 홈페이지에 고시한 예치기간별 중도해지
이율을 적용 받습니다.
만기후 이율 (연, 세전) 1. 2015.4.19 이전 가입 예금
만기 후 1개월 이내 만기 후 1개월 초과 1년 이내 만기 후 1년
초과 1.5% 1.0% 0.1%

Context 3

학습 결과 - PDF Data로 TEST

Q. 재예치가 가능한가요?



Hugging Face

mncai/Polyglot-7B-Kor100K-epoch2

재예보는 불가능합니다. 지정된 계좌에 입금된 원금과 세금만을 공제한 후, 재예치하는 것은 불가능합니다(중도해지시에는 이율이 적용되며, 만기해지 시에는 이와 유사한 이율이 제공됩니다). 중도 해지시에 제공되는 이율은 해당 계좌에 적용되는 것이 아니라, 재 예치를 위한 자금이 이율에 따라 계산된 후, 해당 계좌의 해지를 처리하는 데 사용됩니다. 재예치가 가능한지 여부는, 해당 계좌가 실제로 재예치를 허용하는 최소 이율보다 낮은 이율로 만기되었는지 여부에 따라 달라집니다.<|endoftext|>



Hugging Face

MinGue/Koyangi-5.8B

재예보는 불가능합니다. 재예치를 가능하려면 예금자가 예금을 중도해지할 수 있는 조건이 충족되어야 합니다. 이는 예금자가 정기예금을 중도해지한 경우에만 가능합니다. 이는 정기 예금의 중도해지이유가 예금자가 가입한 정기예금의 만기일로부터 경과한 기간에만 가능하기 때문입니다. 따라서 예금자가 중도해지를 원할 경우에만 재예치를 허용됩니다.

결론

학습 결과 분석
평가

COOPERATION

평가 한계점

PDF reading

PDF를 읽어오는 과정에서 많은 문제가 발생

- 예금/적금 파일의 특성상 **대부분 표**로 이루어져 있음 → 아직 general한 해결 방법이 없음
- 위의 이유로 text를 **전처리하기 어려움**

Vector search

- Vector search로 question에 **유사한 passage를 가져오지 못함**
성능이 좋은 **OpenAI의 embedding모델**을 사용하였음에도 불구하고 대략 **3분의 1 정도는** question과 관련있는 passage를 가져오지 못함
- **Hybrid search의 가능성**
dense vector의 similarity search 결과와 sparse vector의 search 결과를 혼합하여 사용(하려 했지만 **기각당함**)

