

Introducing ChatGPT Plus

Retrieval Augmented Generation

How to chat with your own custom data



MINDS
company

서우림, 황선경, 전민규, 박민지

목차



서론

프로젝트 목표

Flow Chart

후반기 계획

Q&A

Llama Index

Step1

Step2

서론

LLM 설명

기존 서비스 소개

COOPERATION

LLM(Large Language Model)



What LLMs are good at

Language understanding

Instruction following

Basic reasoning

Code understanding



What they need help with

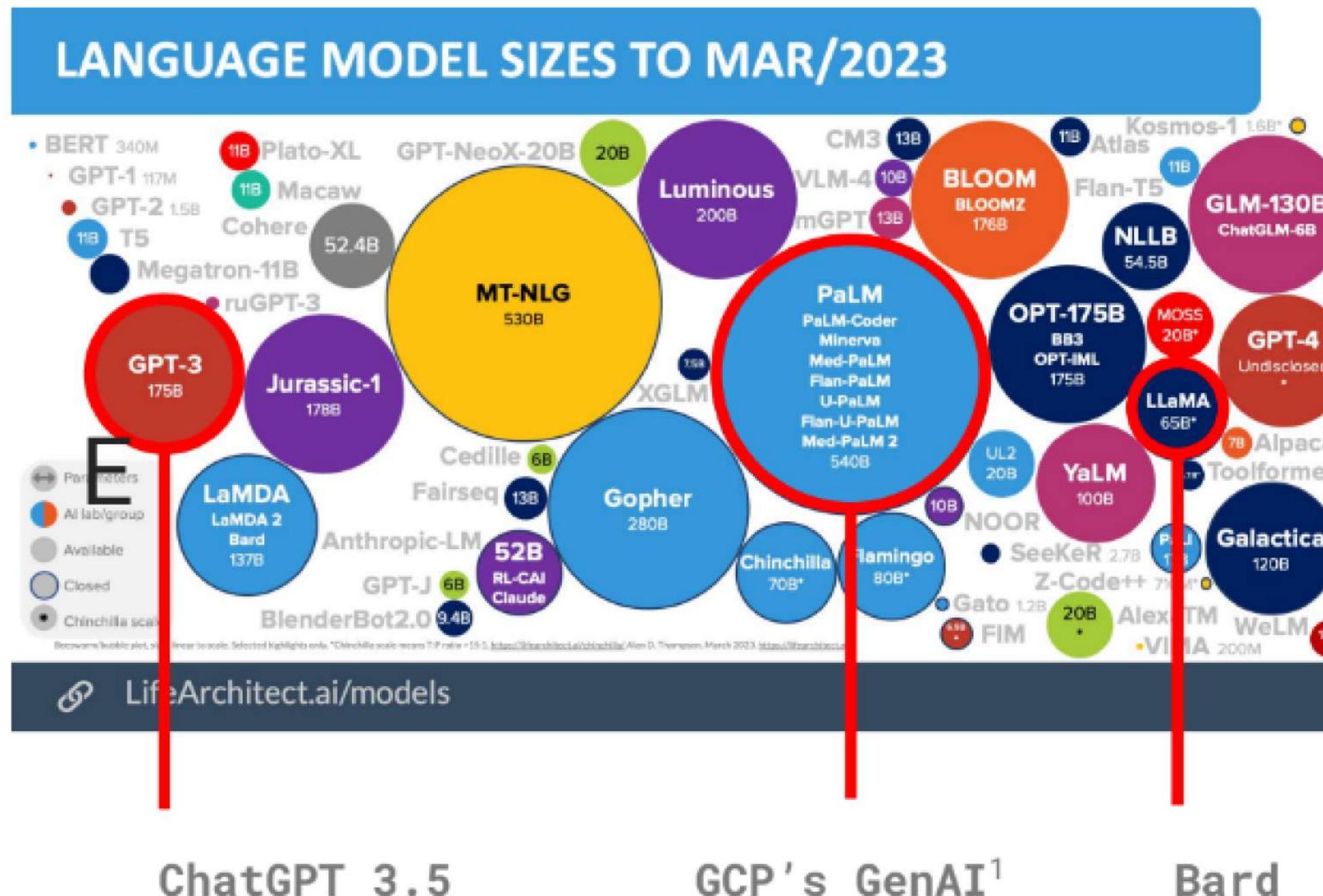
Up-to-date knowledge

Knowledge of untrained data

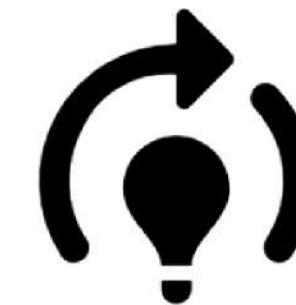
More challenging reasoning

Interacting with the world

How can we help LLM



More training?



No

Too Expensive and Large

What we need is

RAG

Existing service - Chat PDF

Chat with any PDF

Join Discord Post to Twitter Share on Facebook

Drop PDF here

Browse my Computer

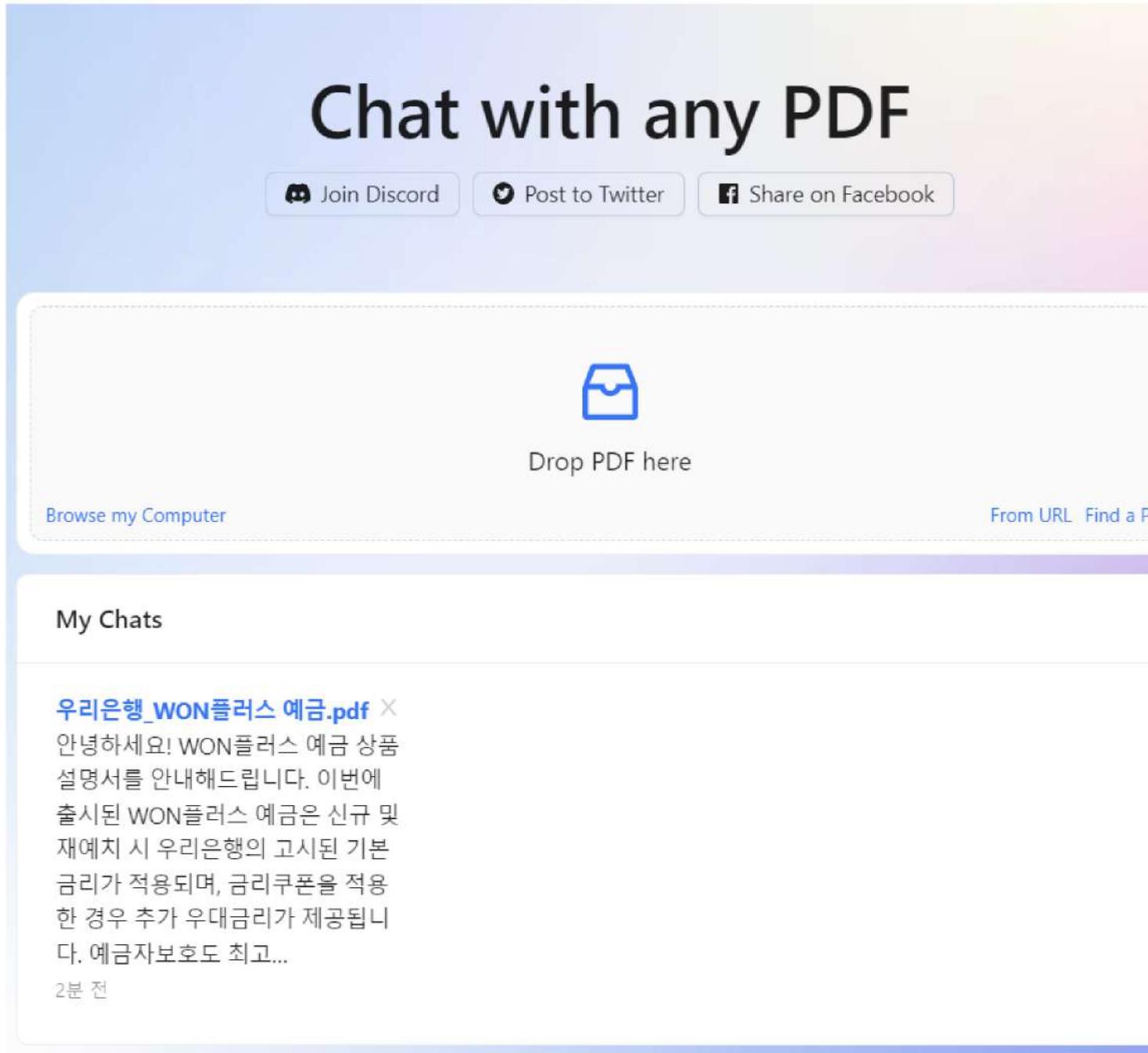
From URL Find a PDF

My Chats

우리은행_WON플러스 예금.pdf X

안녕하세요! WON플러스 예금 상품 설명서를 안내해드립니다. 이번에 출시된 WON플러스 예금은 신규 및 재예치 시 우리은행의 고시된 기본 금리가 적용되며, 금리쿠폰을 적용한 경우 추가 우대금리가 제공됩니다. 예금자보호도 최고...

2분 전



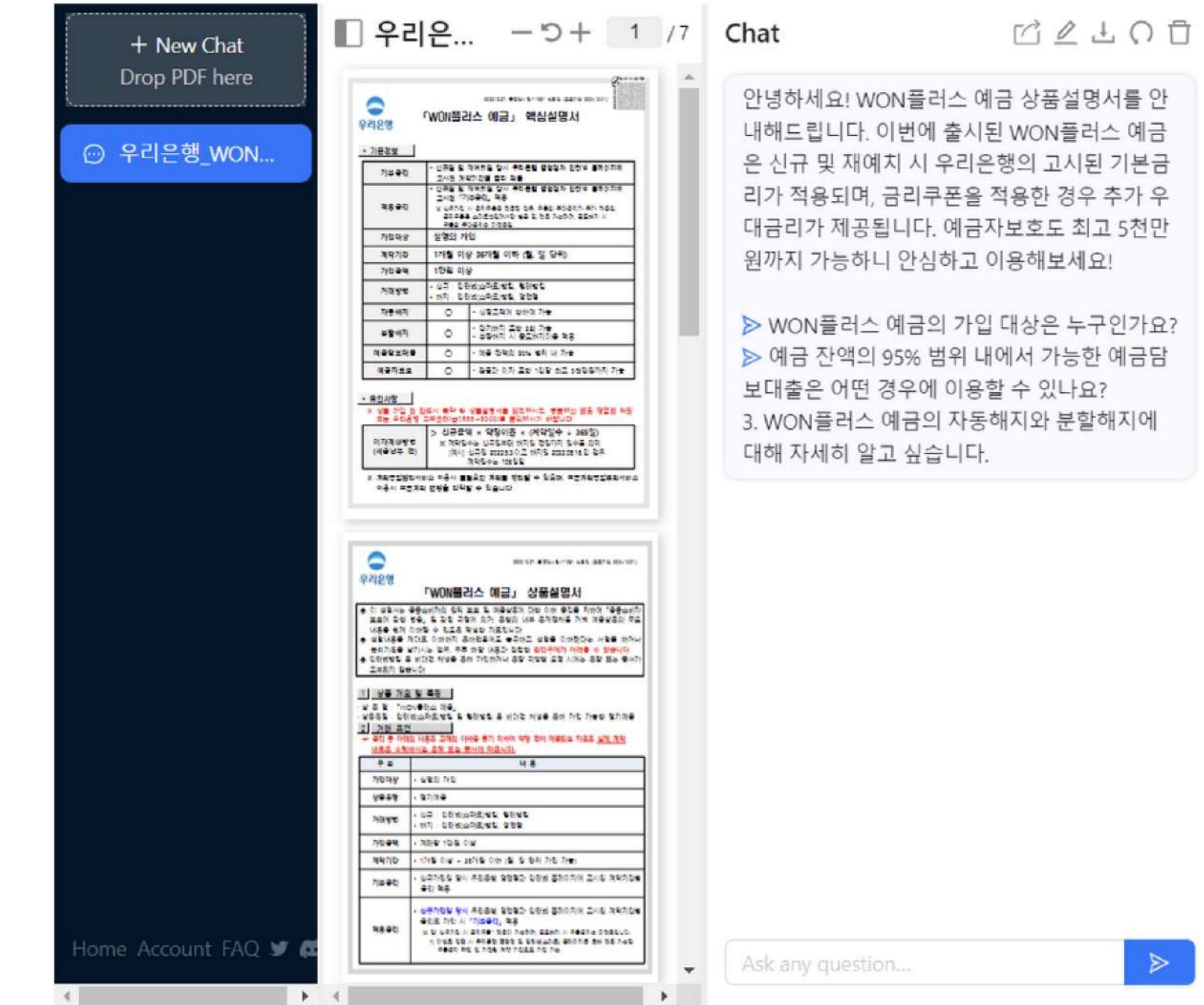
+ New Chat Drop PDF here

우리은행 WON...

안녕하세요! WON플러스 예금 상품설명서를 안내드립니다. 이번에 출시된 WON플러스 예금은 신규 및 재예치 시 우리은행의 고시된 기본금리가 적용되며, 금리쿠폰을 적용한 경우 추가 우대금리가 제공됩니다. 예금자보호도 최고 5천만 원까지 가능하니 안심하고 이용해보세요!

▷ WON플러스 예금의 가입 대상은 누구인가요?
▷ 예금 잔액의 95% 범위 내에서 가능한 예금당 보대출은 어떤 경우에 이용할 수 있나요?
3. WON플러스 예금의 자동해지와 분할해지에 대해 자세히 알고 싶습니다.

Ask any question... ▶



Why Chat PDF?

Hallucination 방지

ChatGPT와 같은 언어 모델은 **실제로는 존재하지 않는 정보**를 생성하거나,

사실이 아닌 내용을 제공할 수 있으며 그 출처도 불분명

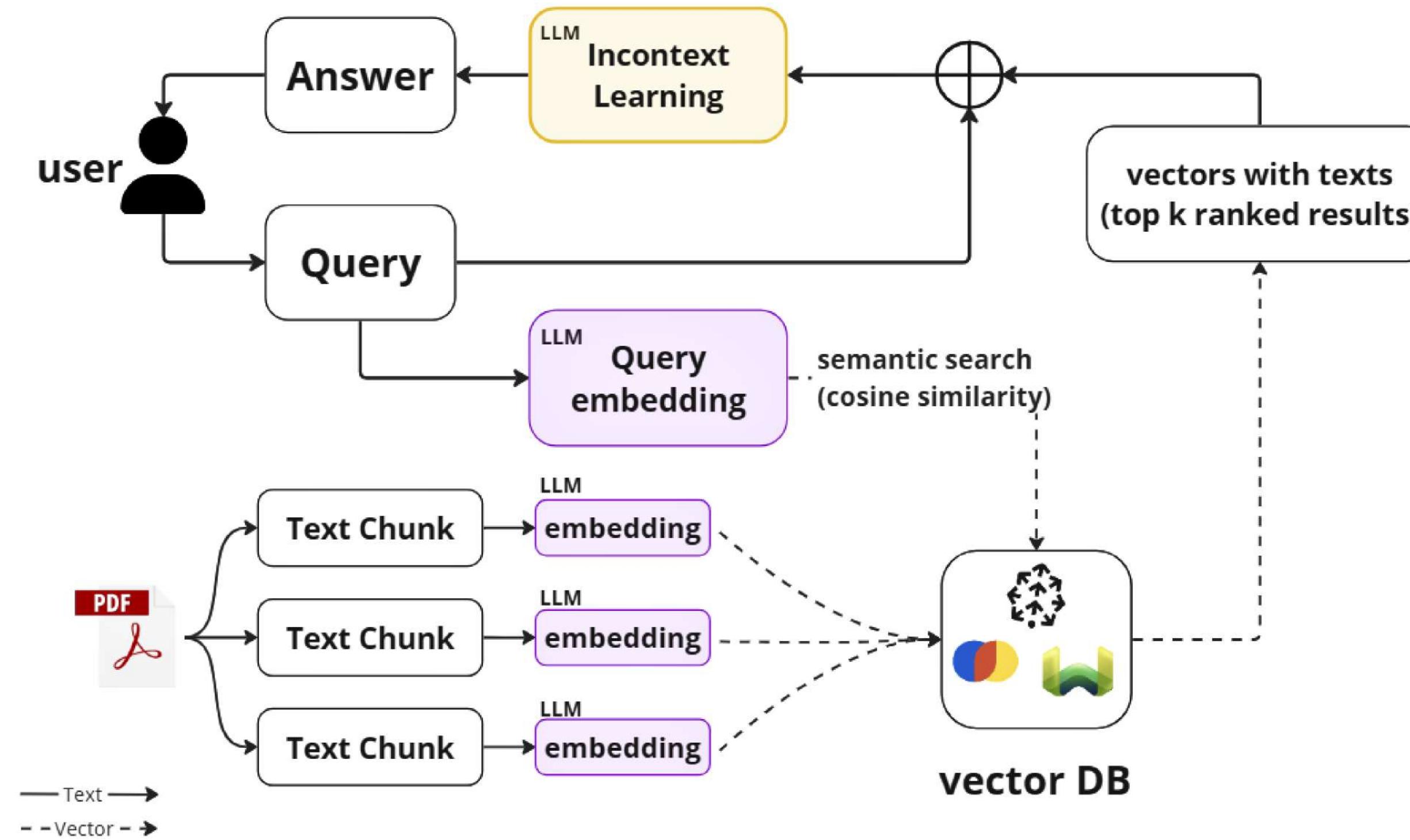
→ 생성된 응답을 **검증하고 신뢰할 수 있는 출처**에서 정보를 확인하는 것이 중요

To chat with own dataset

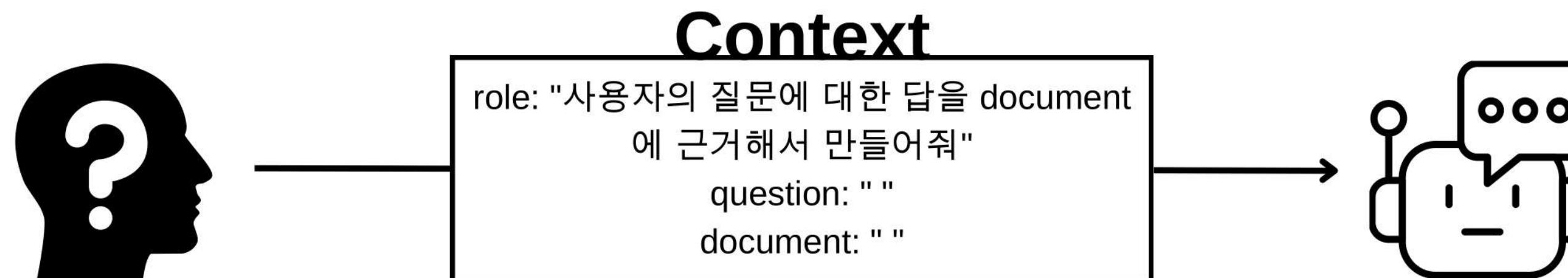
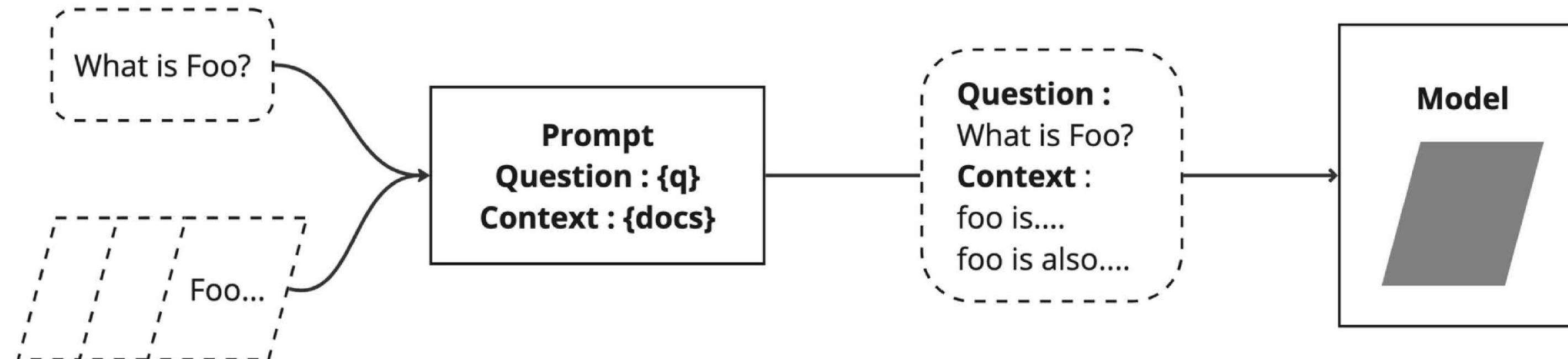
자신의 데이터를 기반으로 하는 챗봇은 **사용자에게 맞춤화된 경험**과 **개인 정보 보호**를 제공하며,

더욱 정확하고 유용한 서비스가 가능

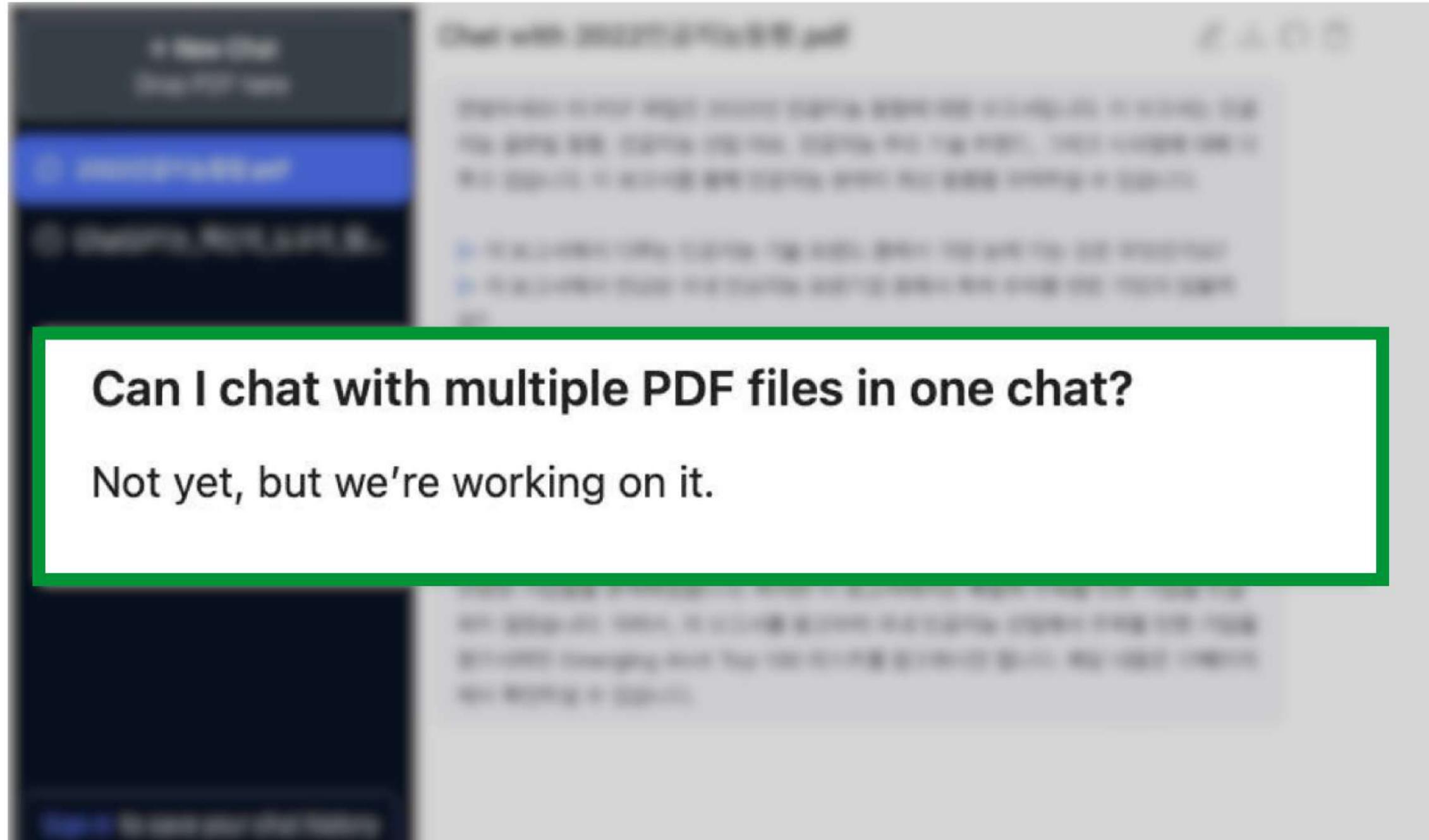
Chat PDF - How it works



Chat PDF - Incontext Learning

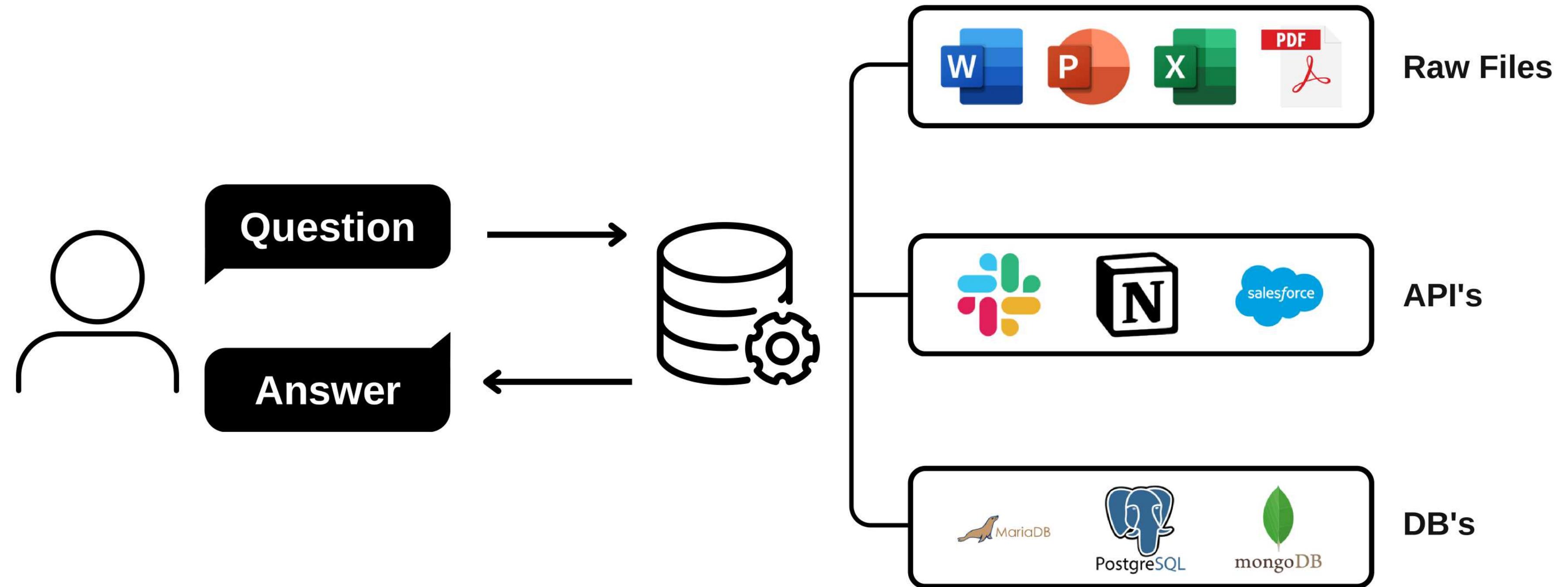


Limitation of Chat PDF



현 시점에서
Multi PDF chat 불가

The Ultimate Goal



목표

데모 소개

COOPERATION

PDF select

Choose an option

Multiple PDF Chat Demo 😎

여기에 질문을 입력하세요: ➤



Frontend



Backend



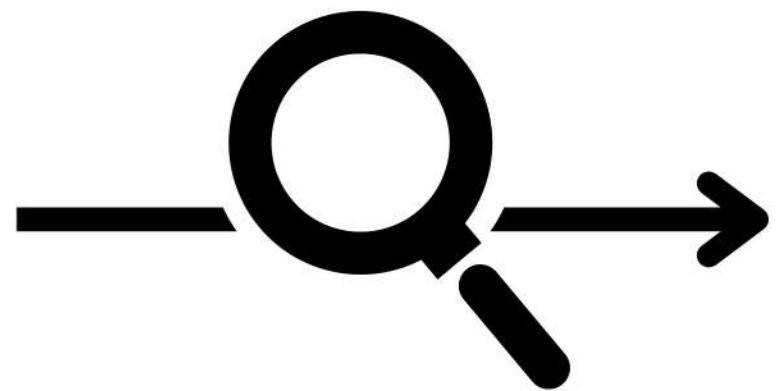
Database

목표 설정

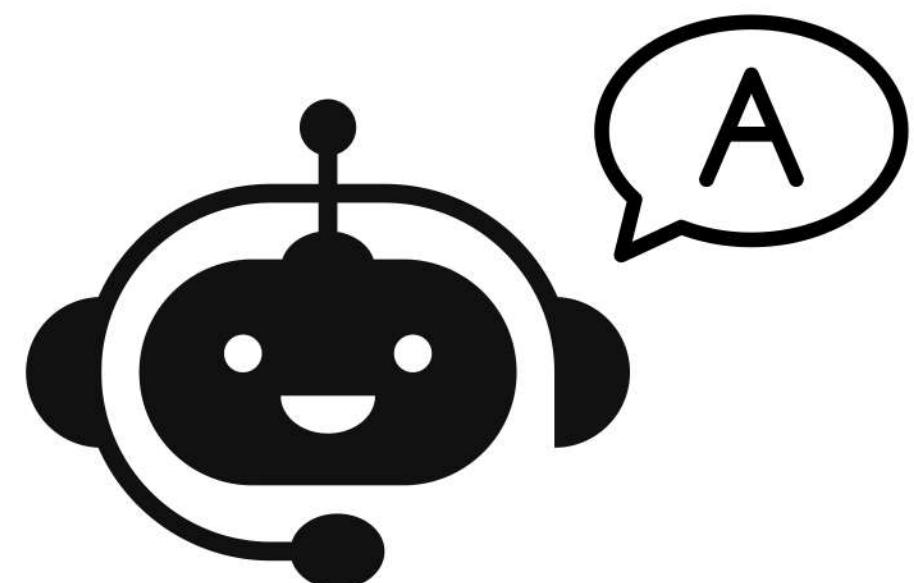
구체적인 기능



회사 내부 데이터베이스

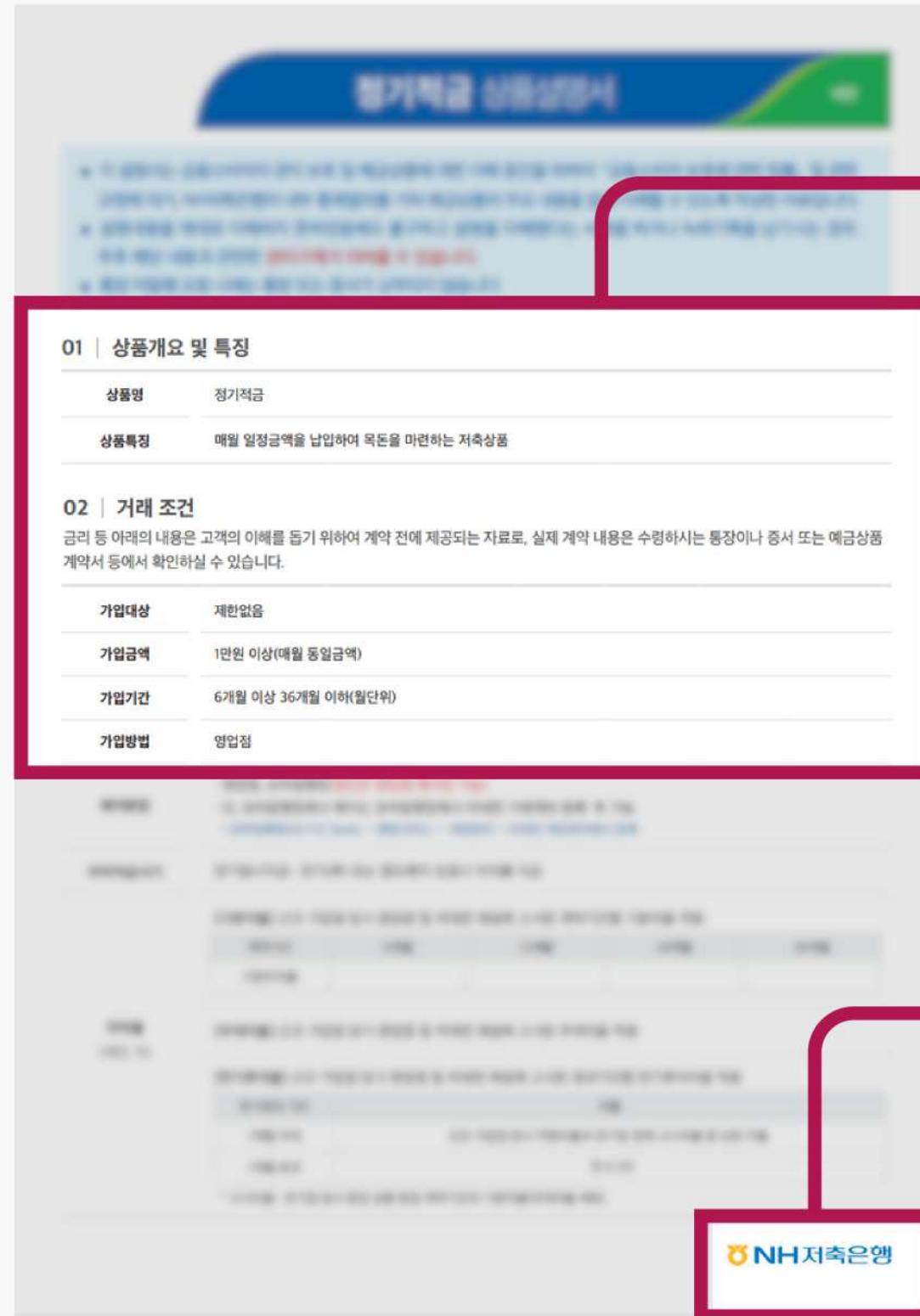


파일 검색



해당 내용
채팅 제공

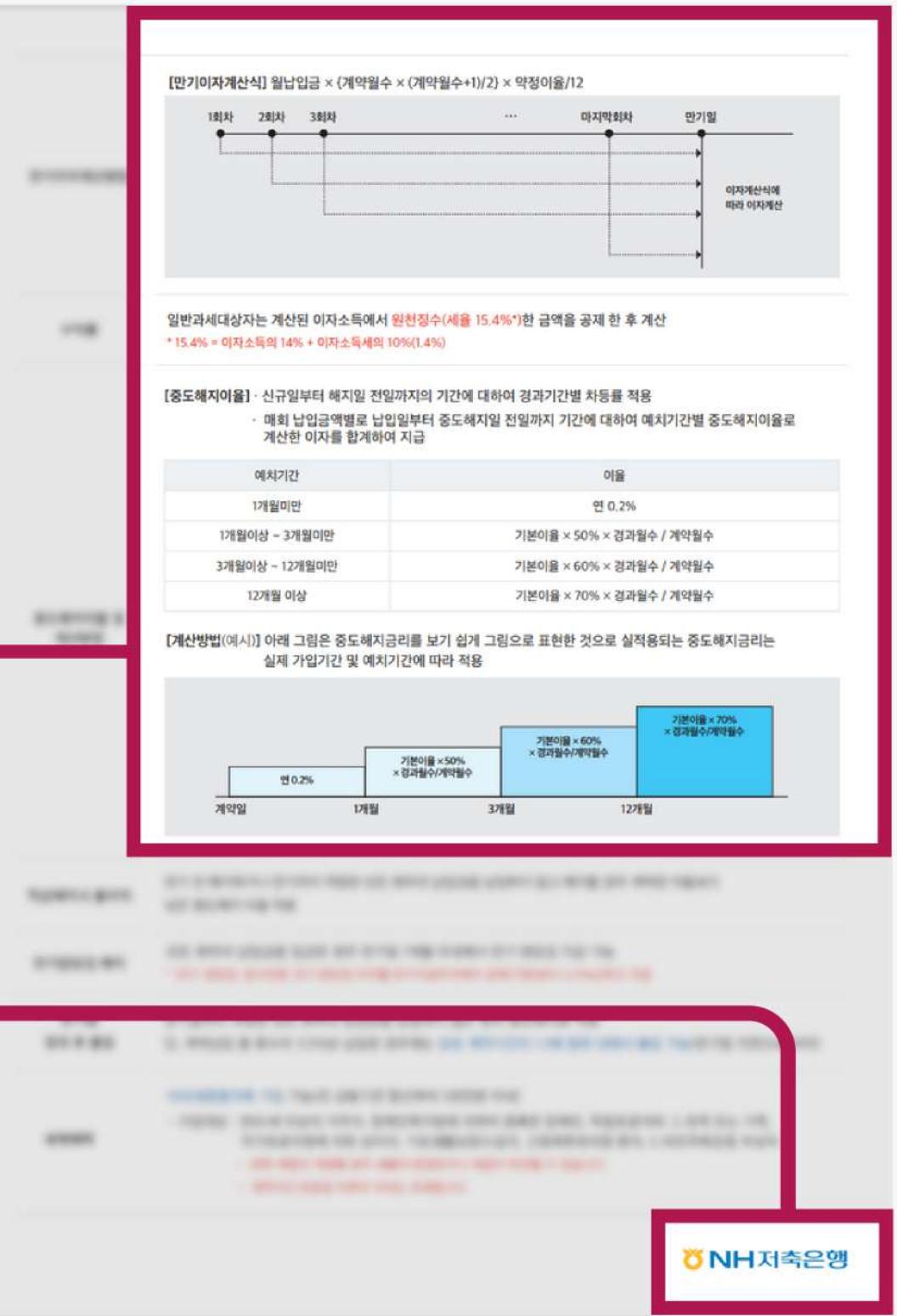
예적금 PDF 파일의 특징



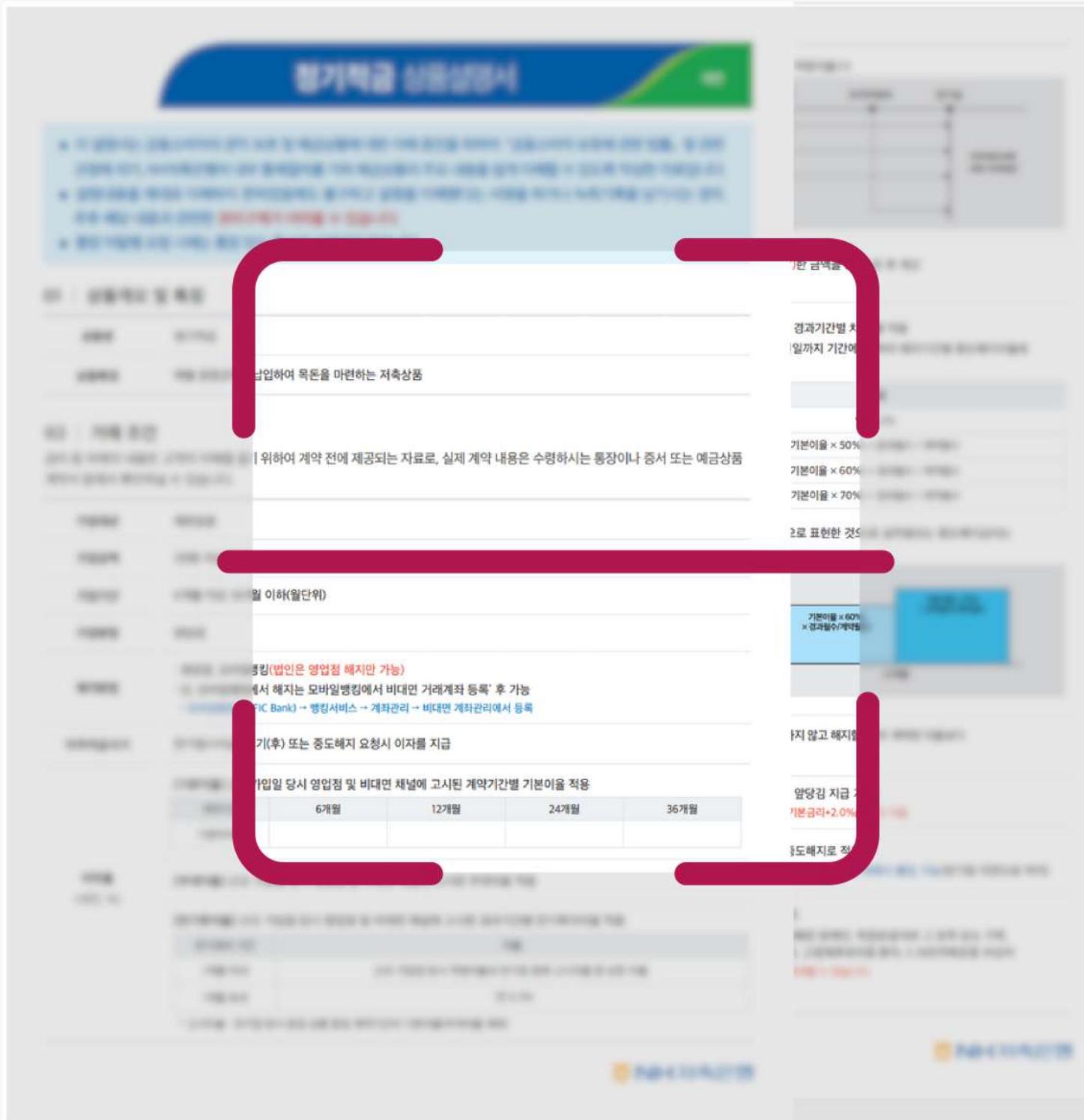
표로 구성된 레이아웃

표 내부의
다른 표나 그래프

반복되는 문구



예적금 PDF 파일의 특징



스캔본 PDF는 텍스트 인식 X

□ OCR 필요

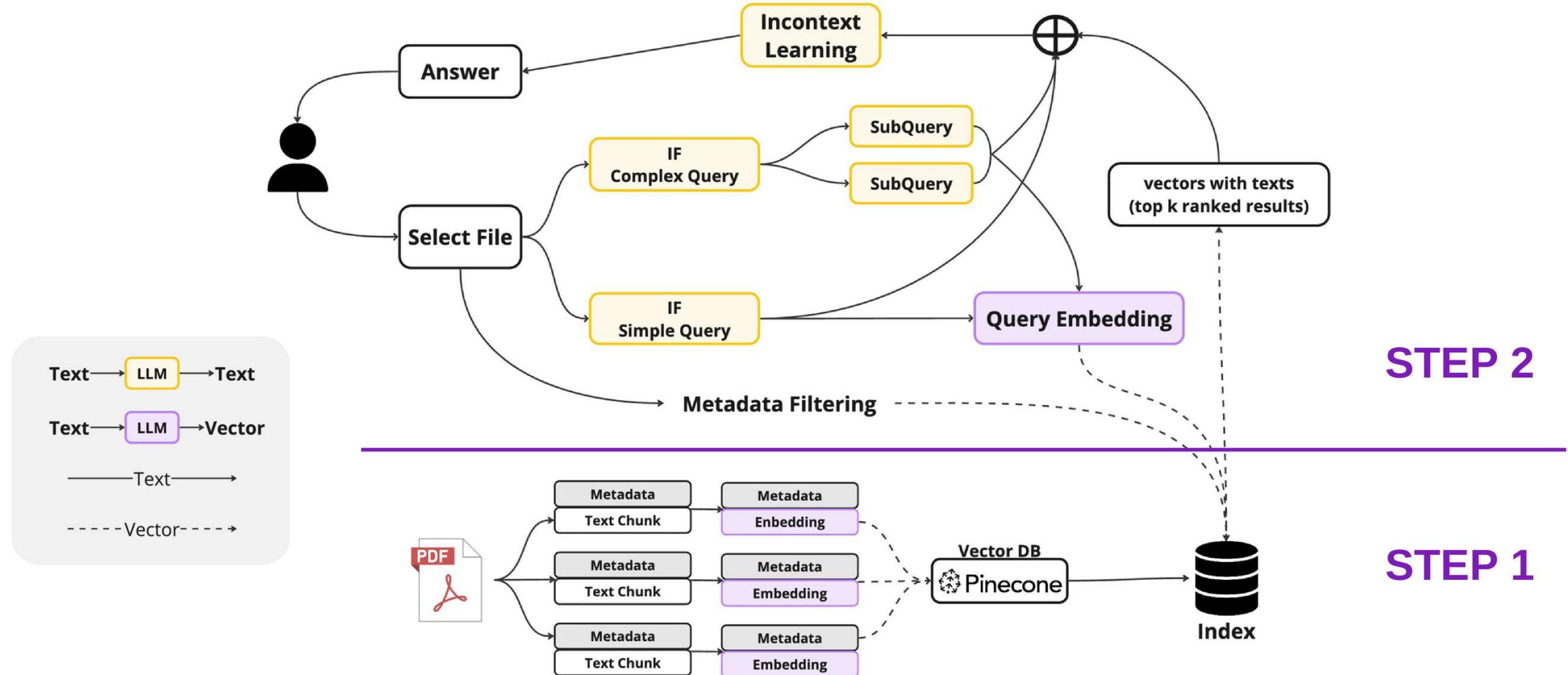
Flow Chart

Llama Index

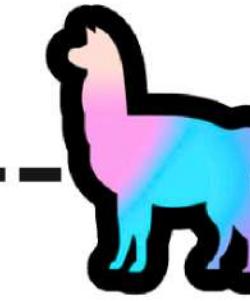
Step 1 Read PDF file

Step 2 Querying & Answering

Flow Chart



Llama Index

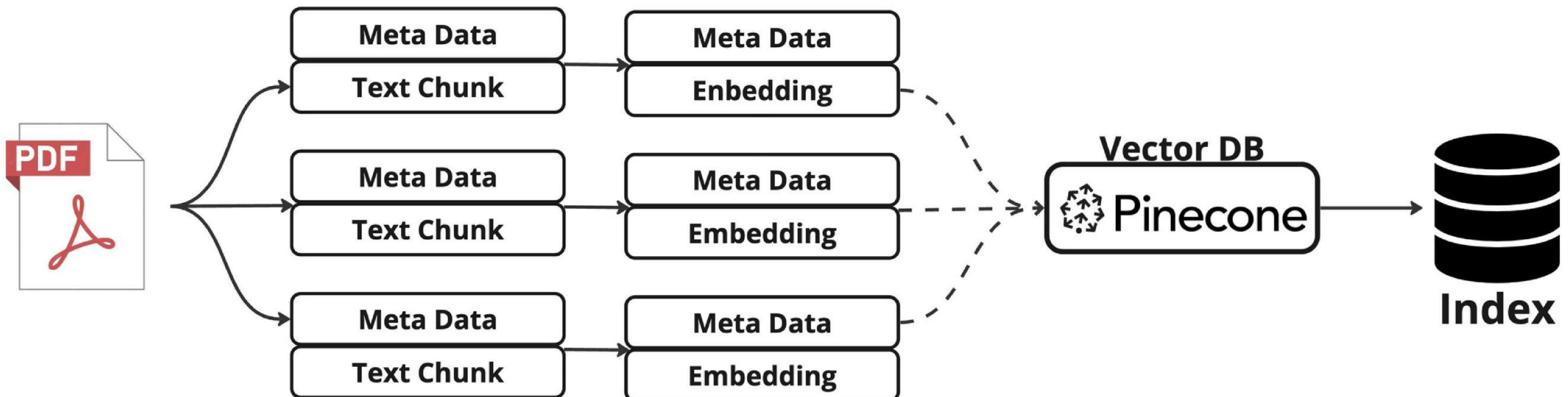


LLM application 개발을 용이하게 하는 종합적인 데이터 프레임워크

사용자의 개인 데이터와 LLM 간의 연계를 담당

text chunk들을 node로 정의하고 hash 값으로 앞뒤 chunk들을 찾기 가능

STEP 1



STEP 1 - PDF Reader



<https://llama-hub-ui.vercel.app/>

file/pdf
ravi03071991 · 4 months ago · ⭐ 1

docugami
tjaffri · 4 months ago · ⭐ 0

file/flat_pdf
emmanuel-oliveira · 4 months ago · ⭐ 1

file/unstructured
thejessezhang · 4 months ago · ⭐ 1

file/cjk_pdf
JiroShimaya · 4 months ago · ⭐ 1

file/pdf_miner
thunderbug1 · 4 months ago · ⭐ 1

Challenge 1 - Tables in PDF

표를 처리하기 위한 시도들

최고이자율	연 400%
대출금액	221101 현재 대출 우대이율은 연 15%로, 일반 우대이자율 요건은 상품설명서 참조)
계약기간	6개월
가입대상	실명인증 및 개인사업자
가입 및 활용방법	1천원 이상 30만원까지 가능
재매지	X
일부매지	O 2회까지 가능
예금인출통장	O 날당액의 100%까지 가능
예금자보호	O 원금과 이자 포함 5천만원까지 가능
세계여행	O 비과세 종합자산으로 가능 가능



 Camelot —

인식률이 좋지 않음
표의 위치 정보를 알 수 없음

 HTML5 Html —

PDF 중 html로 바꾸면 인식 안되는 경우 O

Challenge 1 - Tables in PDF

전처리가 필수적인가?

원본 PDF 표

기본금리	• 신규일 및 재예치일 당시 우리은행 영업점과 인터넷 홈페이지에 고시된 계약기간별 금리 적용
적용금리	• 신규일 및 재예치일 당시 우리은행 영업점과 인터넷 홈페이지에 고시된 「기본금리」 적용 ※ 신규가입 시 금리쿠폰을 적용한 경우, 쿠폰의 우대금리가 추가 제공됨. 금리쿠폰은 스마트뱅킹에서만 발급 및 적용 가능하며, 중도해지 시 쿠폰의 우대금리는 미적용됨.
가입대상	실명의 개인
계약기간	1개월 이상 36개월 이하 (월, 일 단위)
가입금액	1만원 이상
거래방법	• 신규 : 인터넷(스마트)뱅킹, 텔레뱅킹 • 해지 : 인터넷(스마트)뱅킹, 영업점
자동해지	<input type="radio"/> - 신청고객에 한하여 가능

Text Extract →

줄별로 표 내용 인식

기본금리• 신규일 및 재예치일 당시 우리은행 영업점과 인터넷 홈페이지에 고시된 계약기간별 금리 적용 적용금리• 신규일 및 재예치일 당시 우리은행 영업점과 인터넷 홈페이지에 고시된 「기본금리」 적용 ※ 신규가입 시 금리 쿠폰을 적용한 경우, 쿠폰의 우대금리가 추가 제공됨 . 금리쿠폰은 스마트뱅킹에서만 발급 및 적용 가능하며 , 중도 해지 시 쿠폰의 우대금리는 미적용됨 . 가입대상 실명의 개인 계약기간 1개월 이상 36개월 이하 (월, 일 단위)가입 금액 1만원 이상 거래방법• 신규 : 인터넷 (스마트)뱅킹, 텔레뱅킹 • 해지 : 인터넷 (스마트)뱅킹, 영업점자동해지 ○ -신청고객에 한하여 가능

생각보다 별도의 전처리 없이도
언어 모델이 정보를 잘 인식

Challenge 1 - Tables in PDF

질문 : "중도해지이율을 보유기간 별로 알려줘"

답변: 보유기간에 따른 중도해지이율은 다음과 같습니다.

- 3개월 미만: 연 0.1%
- 3개월 이상 ~ 6개월 미만: 기본이율 $\times 50\% \times$ 보유일수 /계약일수 (최저이율 연 0.15%)
- 6개월 이상 ~ 9개월 미만: 기본이율 $\times 70\% \times$ 보유일수 /계약일수
- 9개월 이상 ~ 11개월 미만: 기본이율 $\times 80\% \times$ 보유일수 /계약일수
- 11개월 이상: 기본이율 $\times 90\% \times$ 보유일수 /계약일수

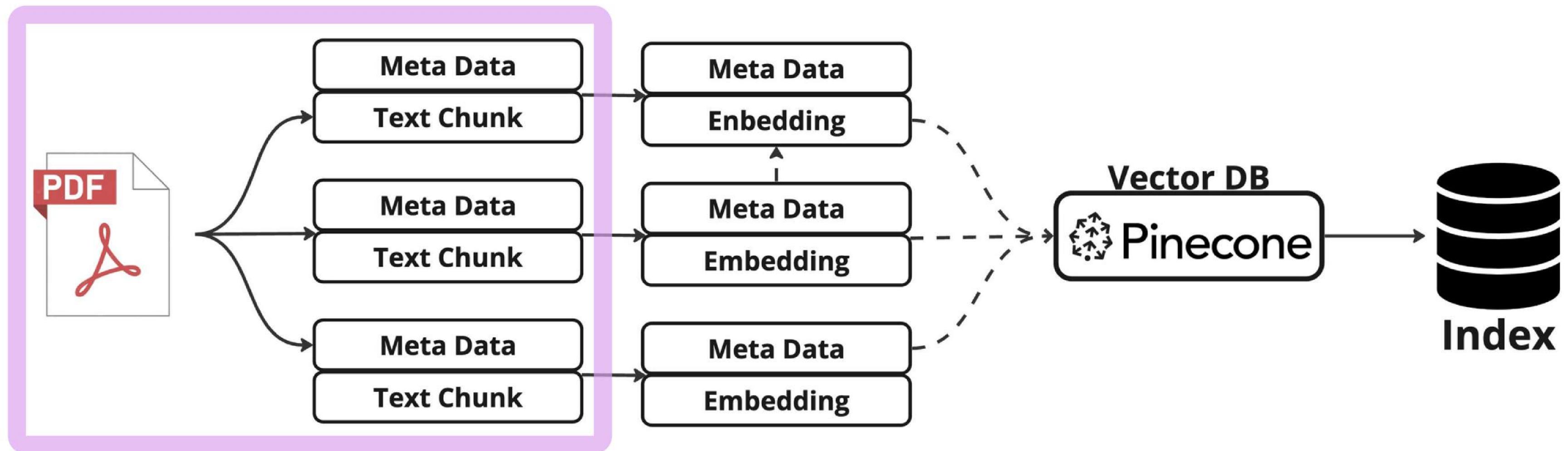
기본이율은 신규(또는 재예치)시점 해당 상품의 기본금리입니다.

Source From: won플러스 예금.pdf, 3.0

중도 해지이율	보유기간			비 고
	3개월 미만	이 율	기타	
	3개월 이상 ~ 6개월 미만	기본이율 $\times 50\% \times$ 보유일수 /계약일수		
	6개월 이상 ~ 9개월 미만	기본이율 $\times 70\% \times$ 보유일수 /계약일수		최저이율 연 0.15%
	9개월 이상 ~ 11개월 미만	기본이율 $\times 80\% \times$ 보유일수 /계약일수		
	11개월 이상	기본이율 $\times 90\% \times$ 보유일수 /계약일수		

※ 기본이율 : 신규(또는 재예치)시점 해당 상품의 기본금리

표 안의 표도 꽤 잘 답변하는 것을 확인할 수 있음

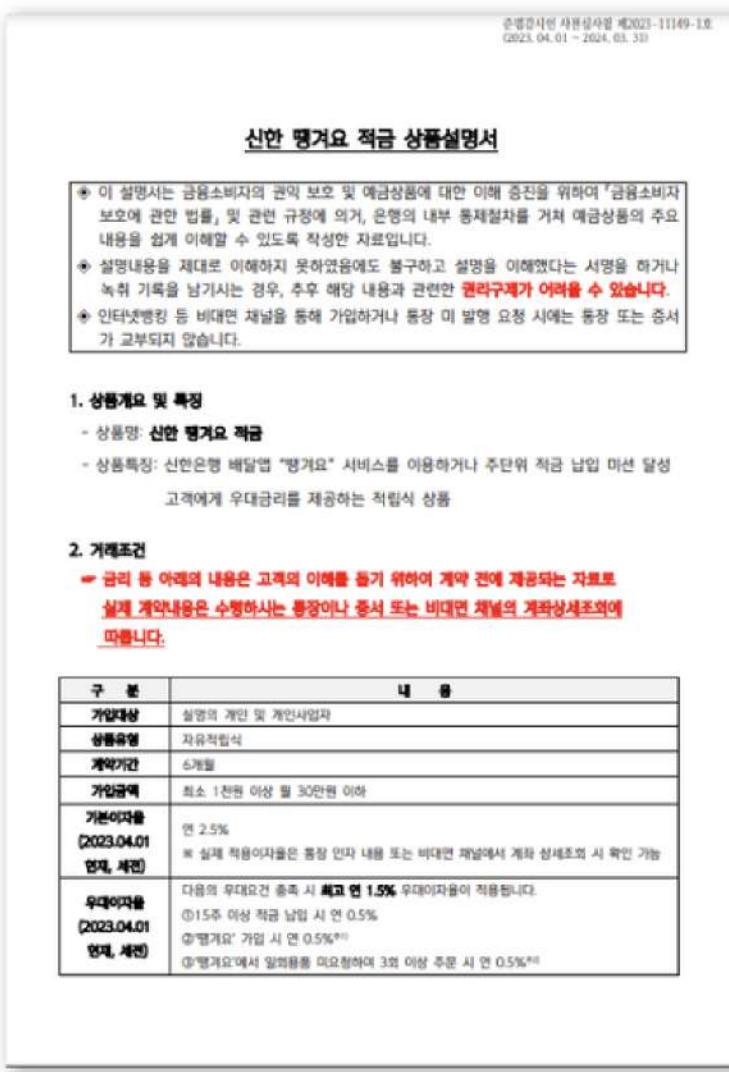


STEP 1 - TextSplitter

Splitting text into chunks

DEFAULT_CHUNK_SIZE = 1024

DEFAULT_CHUNK_OVERLAP = 50



신한 땡겨요 적금 상품설명서 이 설명서는 금융소비자의 권리 보호 및 예금상품에 대한 이해 증진을 위하여 「금융소비자 보호에 관한 법률」 및 관련 규정에 의거, 은행의 내부 통제절차를 거쳐 예금상품의 주요 내용을 쉽게 이해할 수 있도록 작성한 자료입니다. 설명내용을 제대로 이해하지 못하였음에도 불구하고 설명을 이해했다는 서명을 하거나 녹취 기록을 남기시는 경우, 추후 해당 내용과 관련한 권리구제가 어려울 수 있습니다. 인터넷뱅킹 등 비대면 채널을 통해 가입하거나 통장 미 발행 요청 시에는 통장 또는 증서 가 교부되지 않습니다.

1. 상품개요 및 특징

- 상품명: 신한 땡겨요 적금
- 상품특징: 신한은행 배달앱 "땡겨요" 서비스를 이용하거나 주단위 적금 납입 미션 달성을 통해 우대금리를 제공하는 적립식 상품

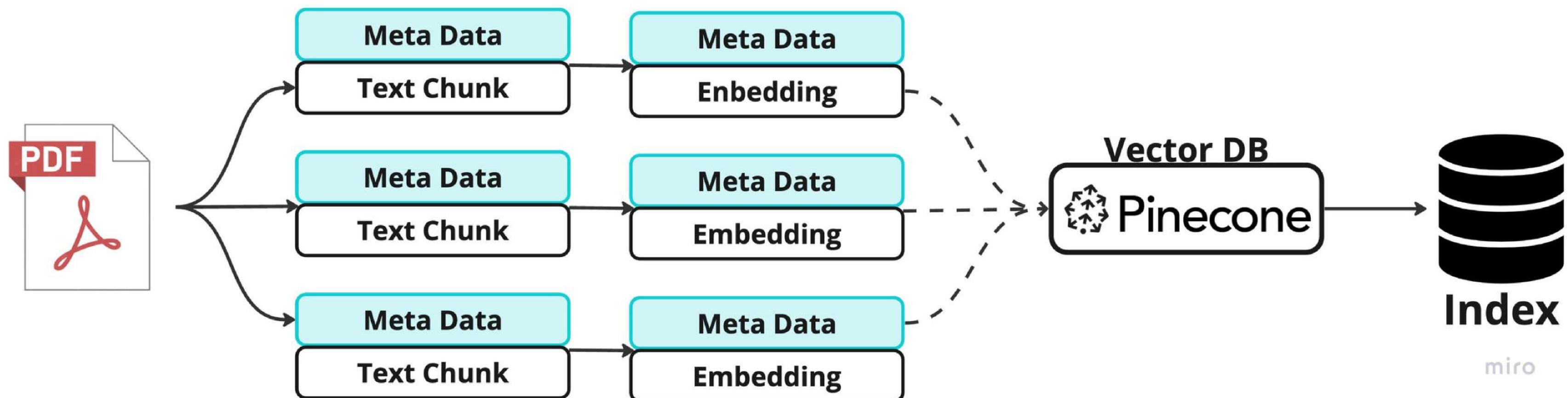
2. 거래조건

금리 등 아래의 내용은 고객의 이해를 돋기 위하여 계약 전에 제공되는 자료로 실제 계약내용은 수령하시는 통장이나 증서 또는 비대면 채널의 계좌상세조회에 따릅니다.

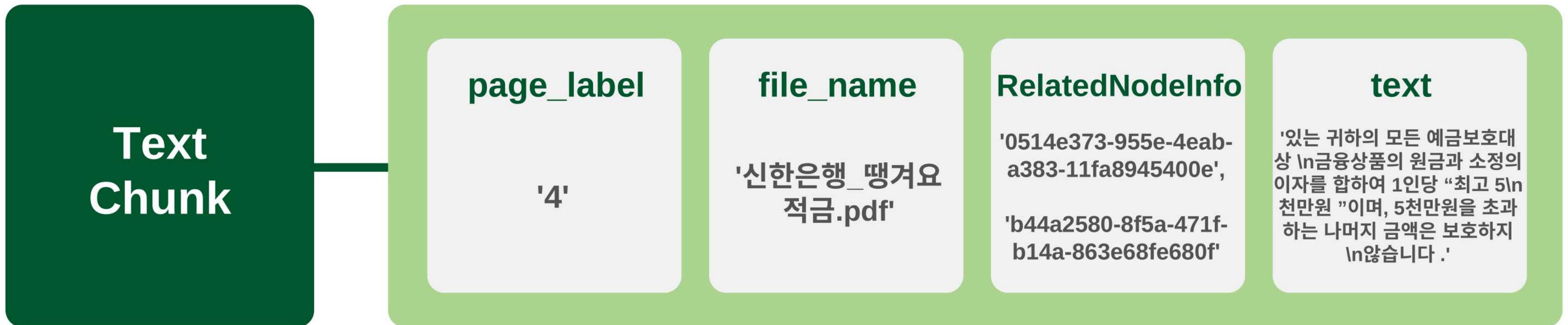
① 15 주 이상 적금 납입 시 연 0.5%



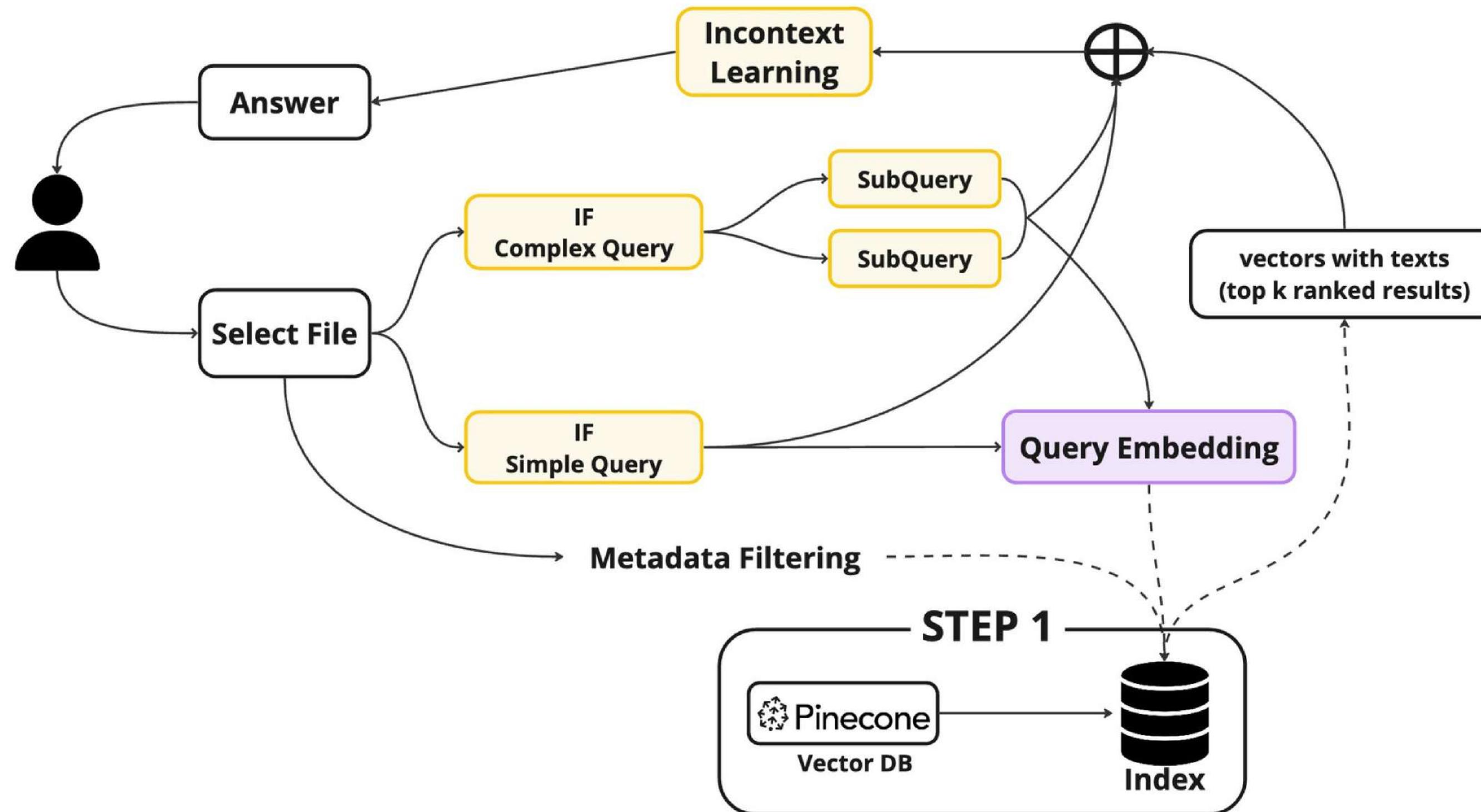
① 15 주 이상 적금 납입 시 연 0.5% ② '땡겨요' 가입 시 연 0.5% 주 1) ③ '땡겨요'에서 일회용품 미요청하여 3 회 이상 주문 시 연 0.5% 주 2) 준법감시인 사전심사필 제2023-11149-1호(2023. 04. 01 ~ 2024. 03. 31)', metadata={'page': 1, 'chunk': 1, 'source': '1-1', 'pdf_name': '땡겨요 적금.pdf', 'text': '① 15 주 이상 적금 납입 시 연 0.5% ② '땡겨요' 가입 시 연 0.5% 주 1) ③ '땡겨요'에서 일회용품 미요청하여 3 회 이상 주문 시 연 0.5% 주 2) 준법감시인 사전심사필 제2023-11149-1호(2023. 04. 01 ~ 2024. 03. 31)'}



STEP 1 - Metadata

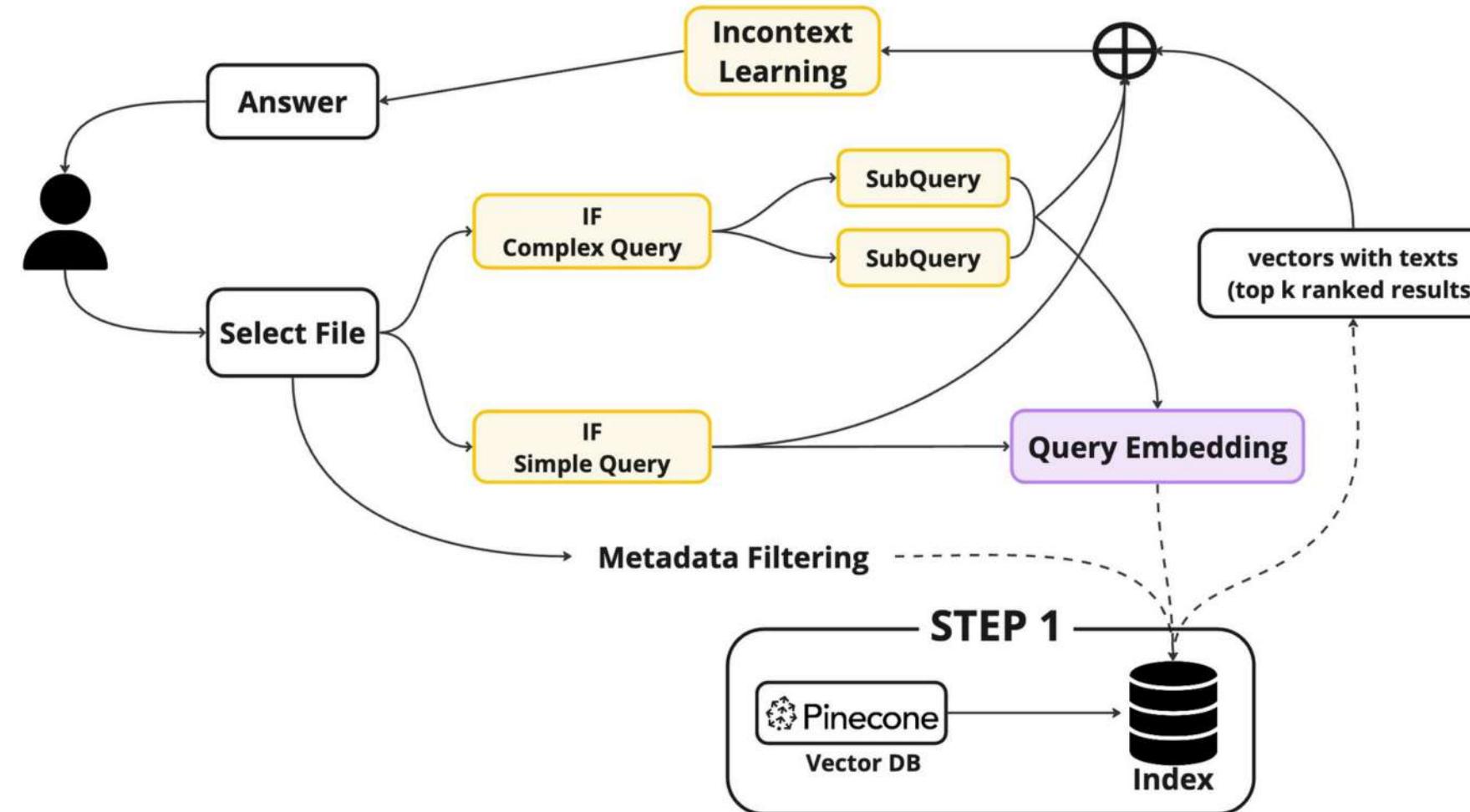


STEP 2

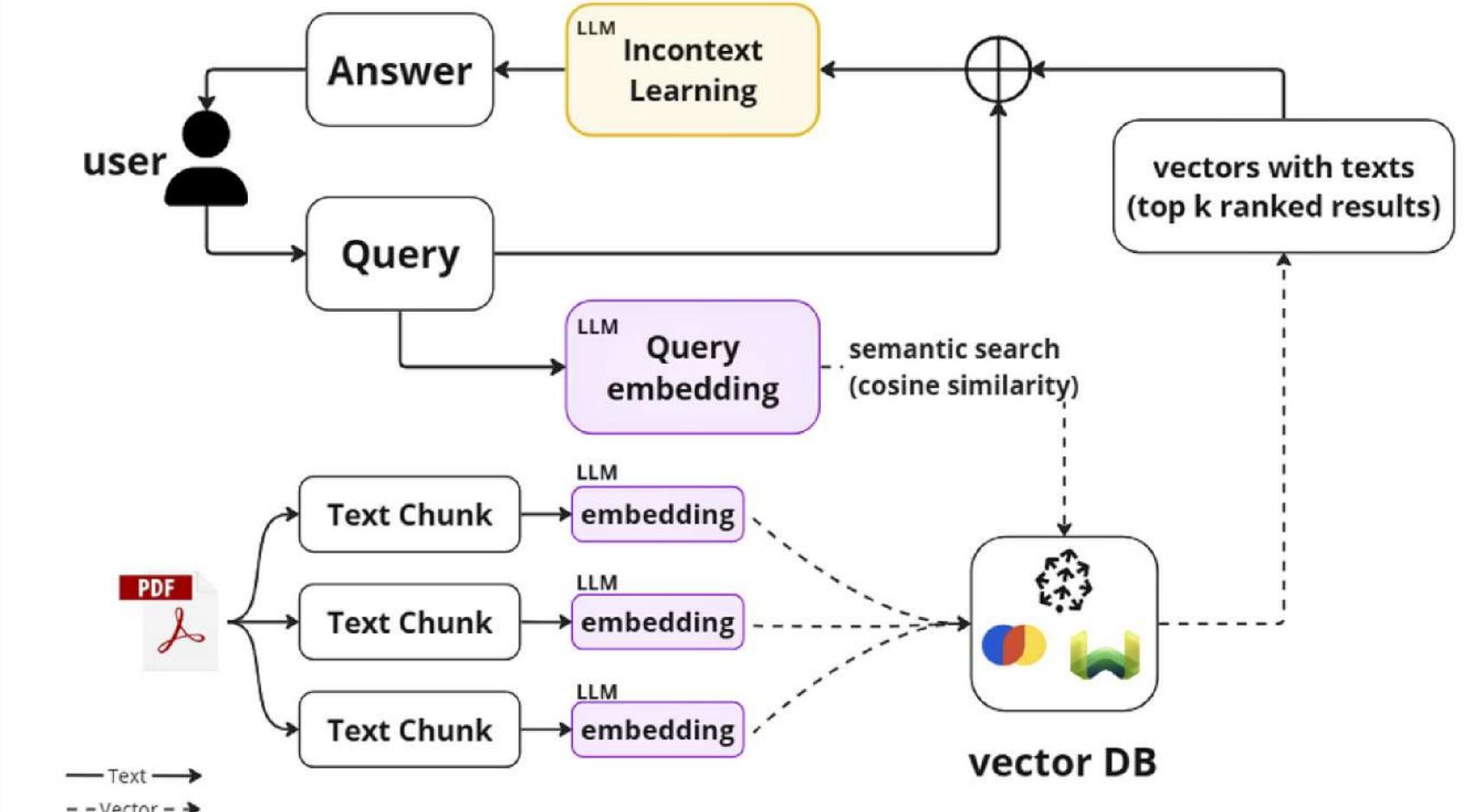




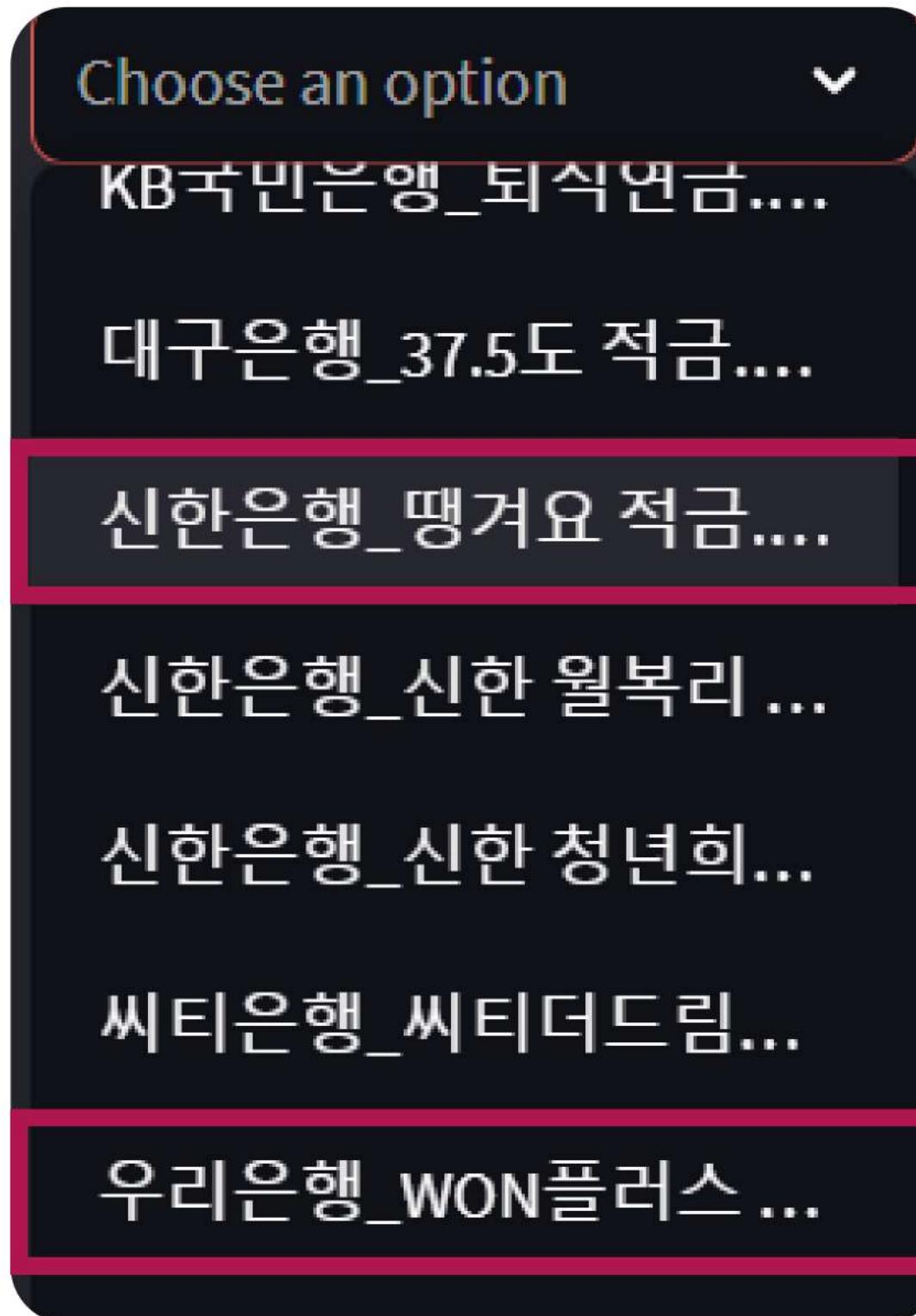
STEP 2



Chat PDF - How it works



STEP 2 - Select File & Metadata Filtering



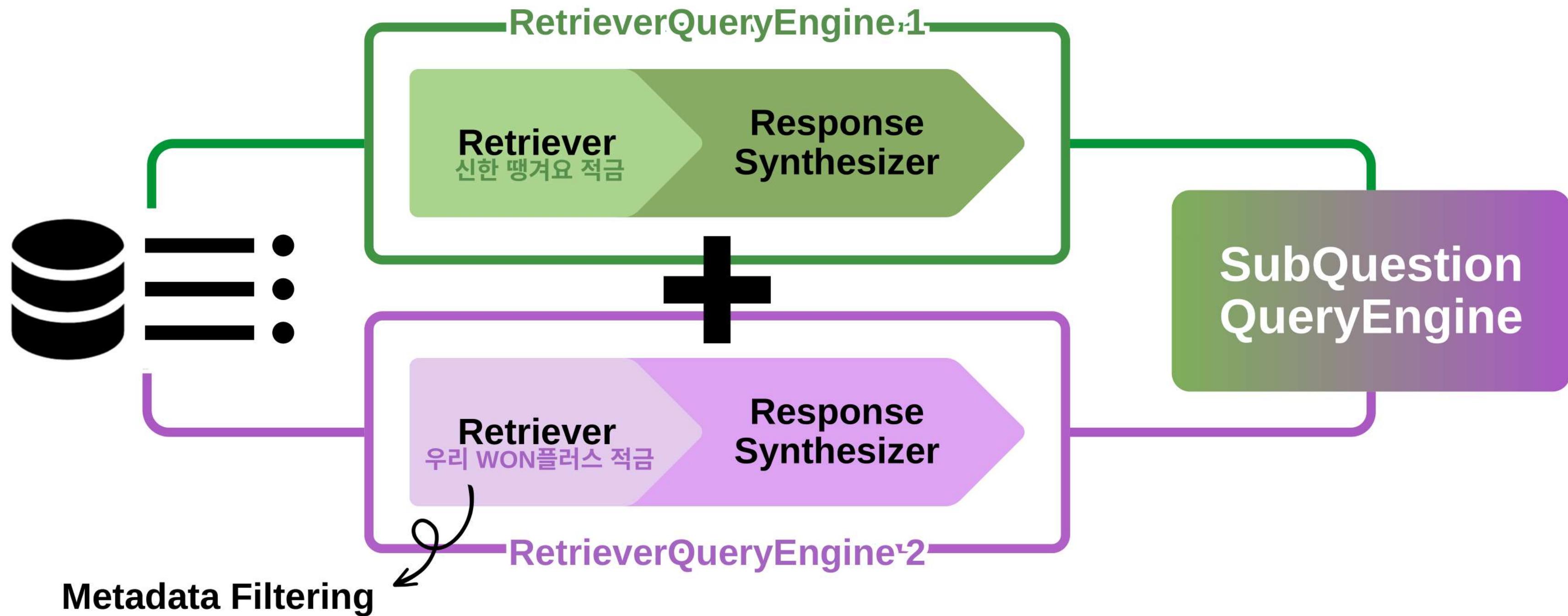
Metadata Filtering



각각의 파일에 대해서만 질문할 수 있도록 **filtering**

Challenge 2 - Mechanism for SubQuery

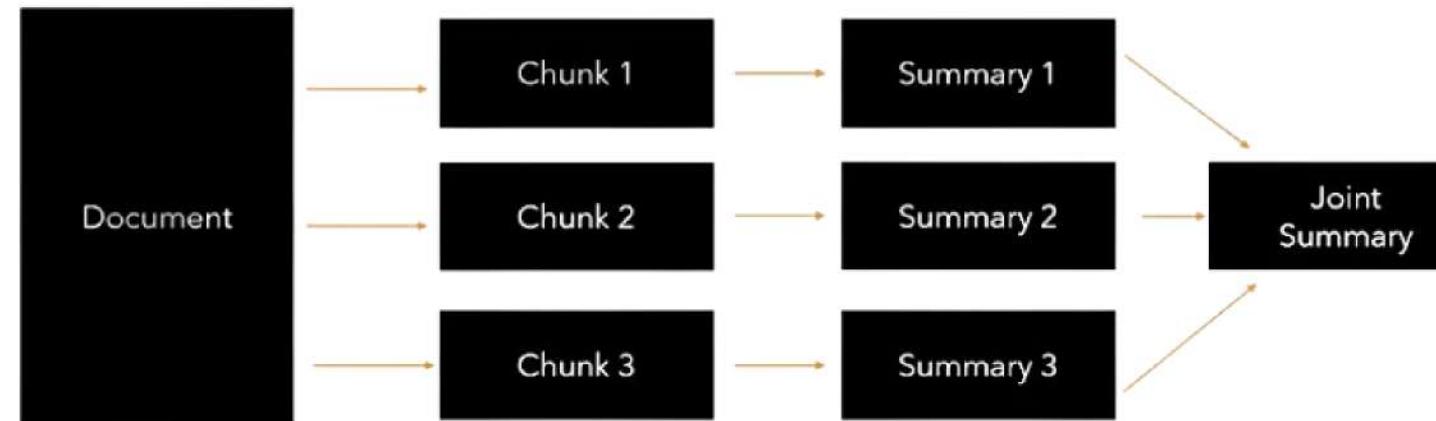
Multi PDF chat 구현하기



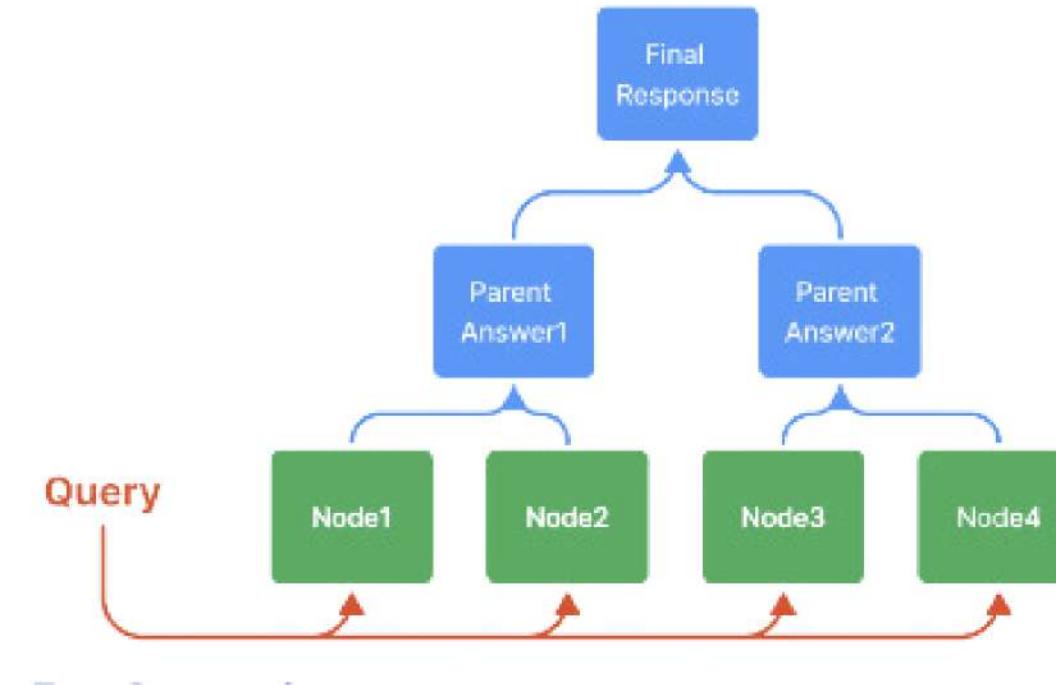
Challenge 3 - Summary for Description

기존 summary 방식들

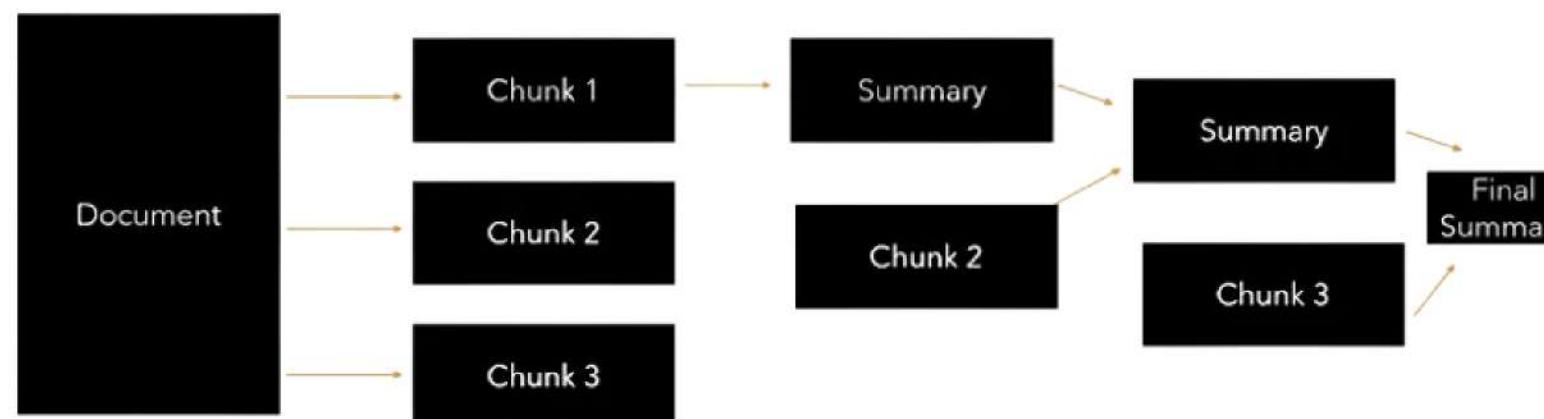
Map Reduce



Tree



Refine

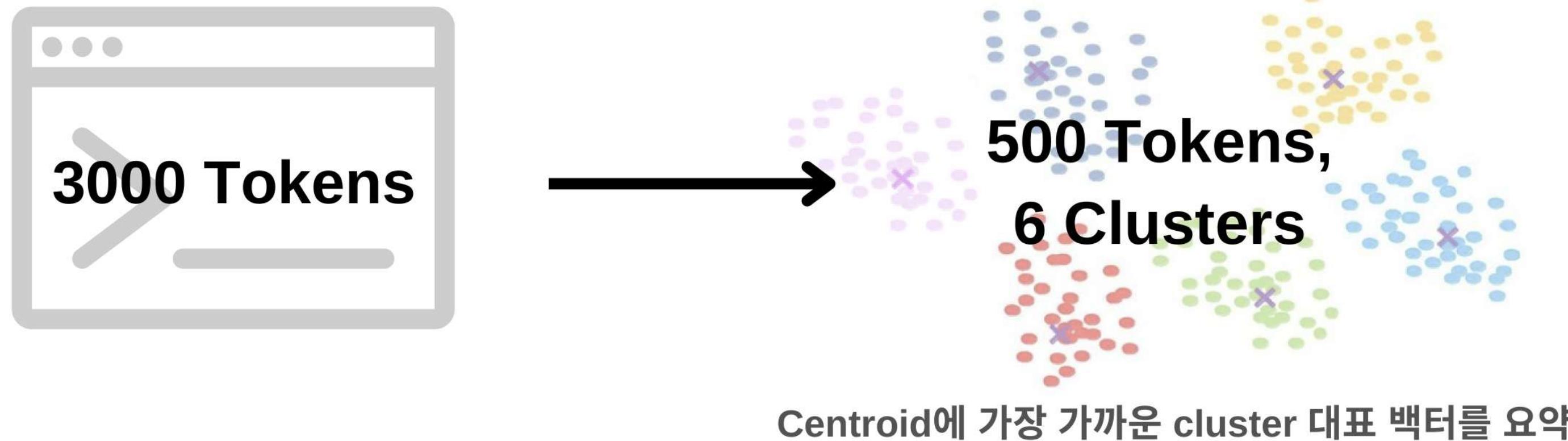


모두 매우 긴 시간이 소요되어
새롭게 K-means 방식을 고안

Challenge 3 - Summary for Description

● K-means

Embedding이 잘 되어있다는 전제 하에 비슷한 의미를 지닌 chunk들은 가깝게 위치하고 있을 것



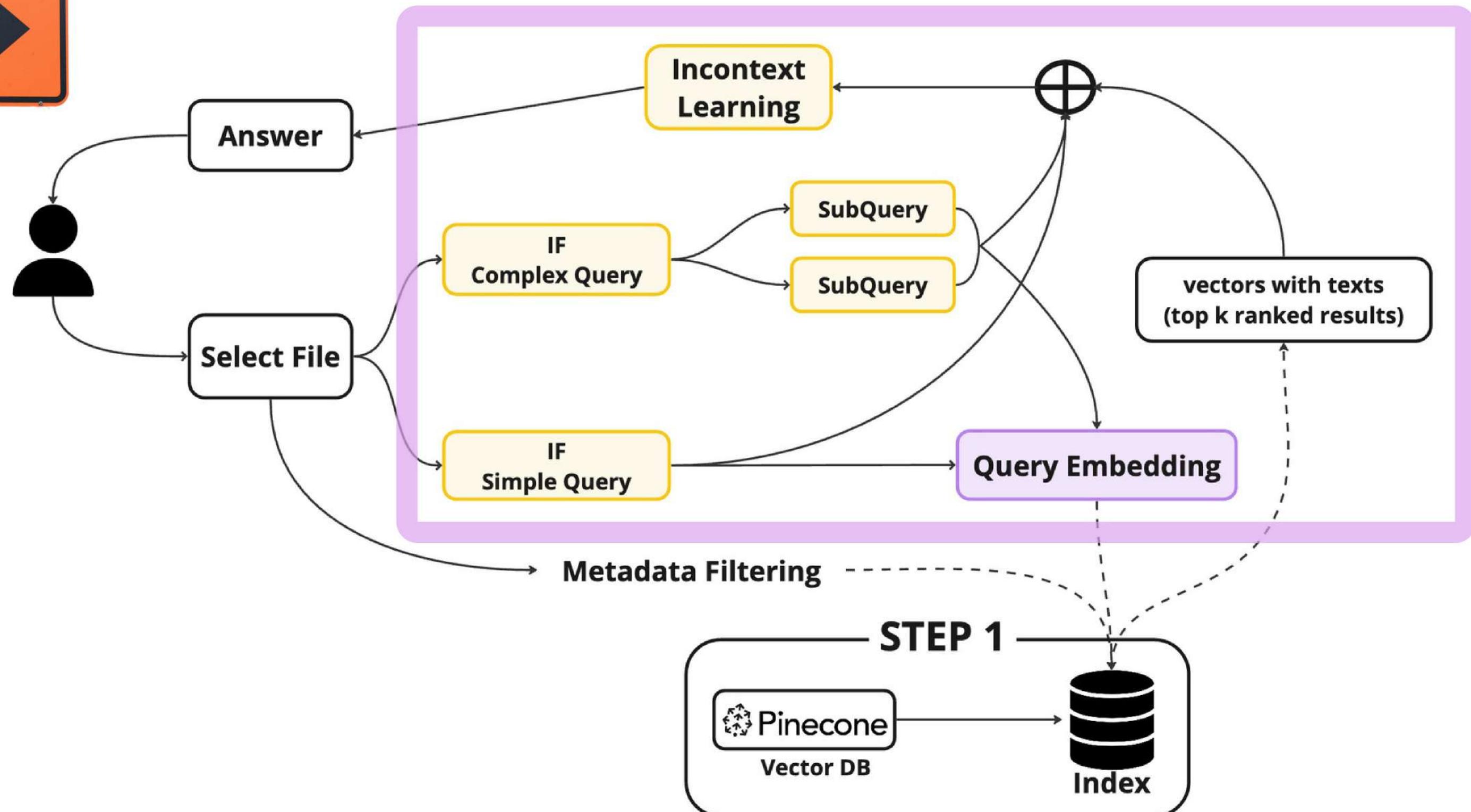
Challenge 3 - Summary for Description

● K-means

완성된 Summarization

'신한은행_땡겨요 적금.pdf': 'Provides information about 신한은행_땡겨요 적금.pdf, and the summary of the information is enclosed in triple backticks(```)\n```이 문서는 신한은행의 적금 상품에 대한 설명서이다. 이 설명서는 금융소비자의 권리 보호와 예금상품에 대한 이해를 증진하기 위해 작성되었다. 설명서에는 예금상품의 주요 내용이 쉽게 이해할 수 있도록 작성되어 있다. 설명서를 제대로 이해하지 못한 상태에서도 설명을 이해했다는 서명이나 녹취 기록을 남기면 추후에 해당 내용과 관련한 권리구제가 어려울 수 있다고 되어있다. 또한 인터넷뱅킹 등 비대면 채널을 통해 가입하거나 통장 미 발행 요청 시에는 통장이나 증서가 교부되지 않는다고 되어있다. 이자율은 최저금리 연 0.10%이며, 만기일에 따라 기간을 구분하여 변경된 이자율이 적용된다고 되어있다. 또한 5년 이내의 법위반 사실을 안 날로부터 1년 이내에서면 등으로 해당 계약의 해지를 요구할 수 있으며, 금융소비자의 요구가 정당하다고 판단될 경우 수수료와 위약금 등 계약해지와 관련한 추가 비용 부담 없이 계약 해지가 가능하다고 되어있다. 이용제한 사항으로는 예금잔액증명서 발급 당일에는 잔액 변동이 불가능하며, 예금 해지 시에는 중도해지이자가 적용된다고 되어있다.'``',

현재 알 수 없는 원인으로 오류 발생

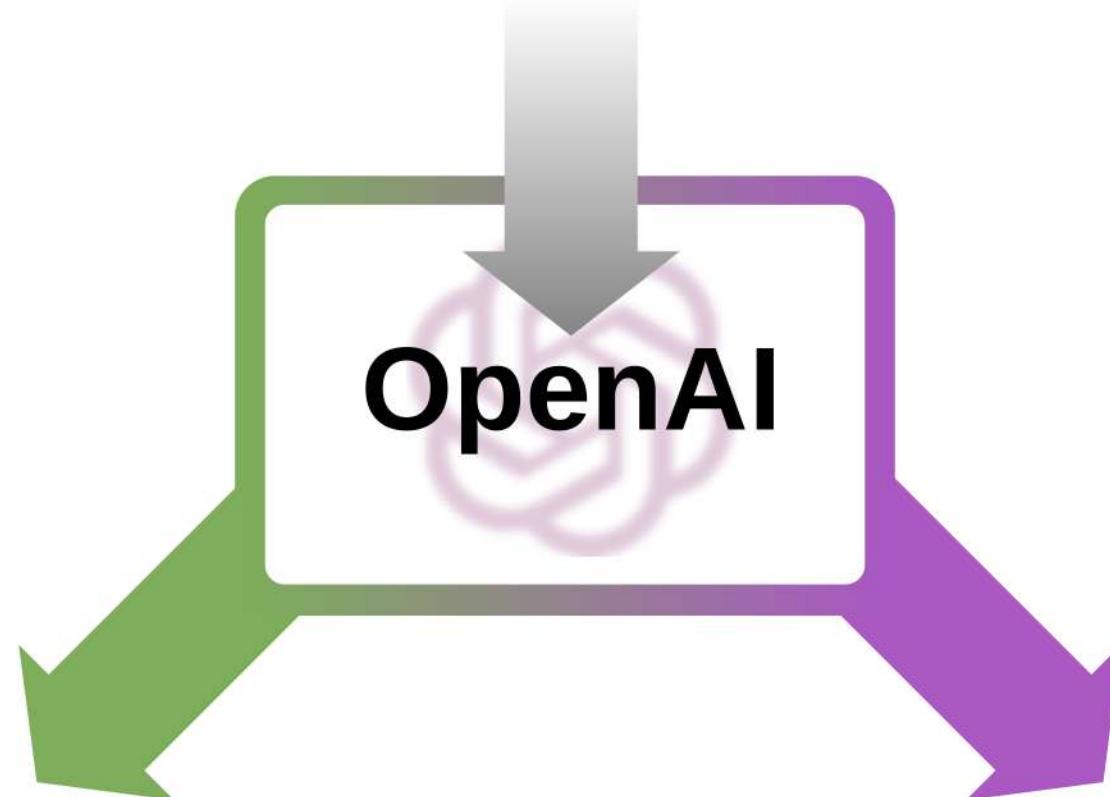


STEP 2 - Querying

Complex Query

Query

신한은행과 우리은행 상품의 가입대상자를 비교해서 알려줘



SubQuery

신한은행 상품의 가입대상자는 무엇인가요?

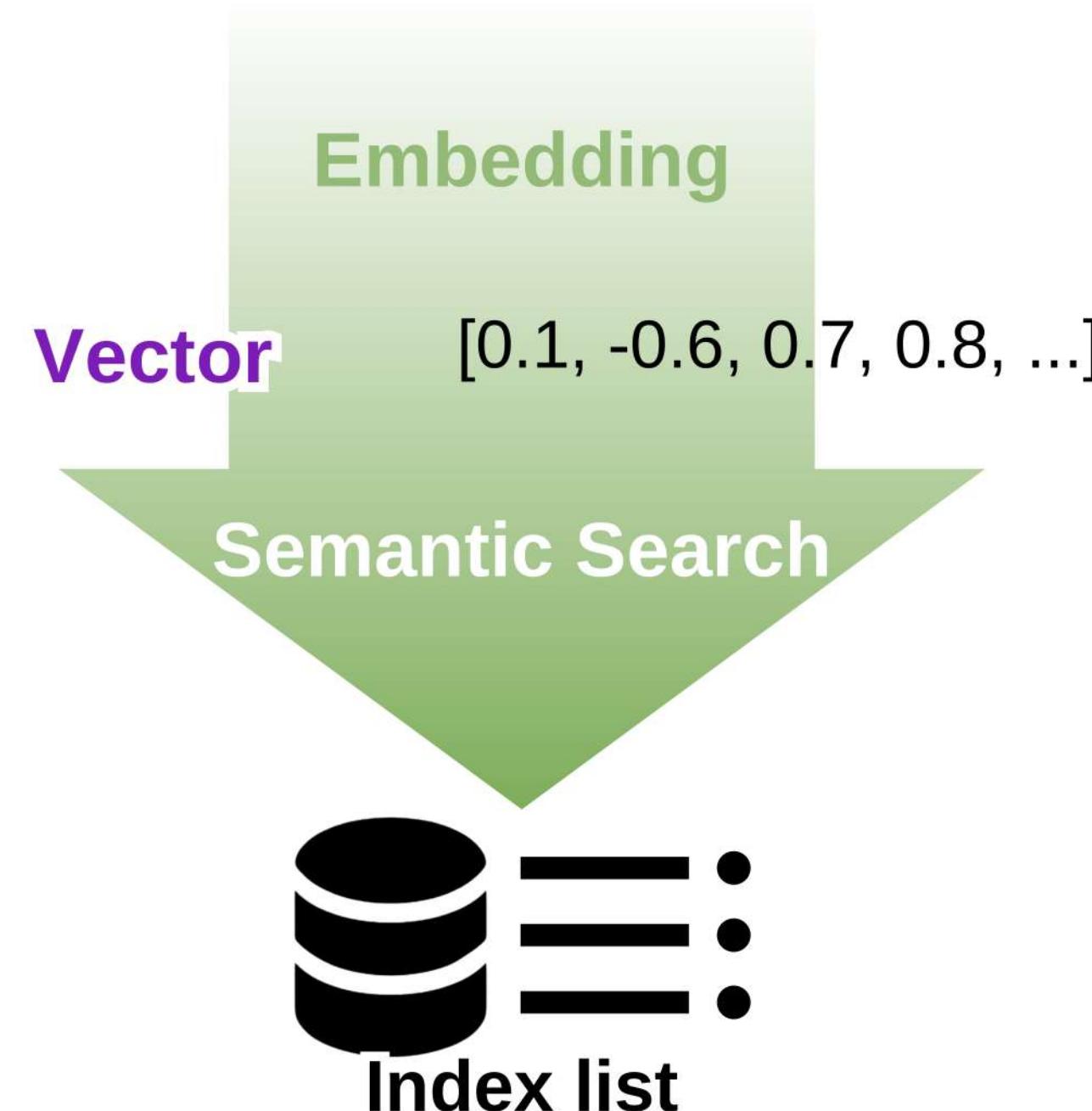
SubQuery

우리은행 상품의 가입대상자는 무엇인가요?

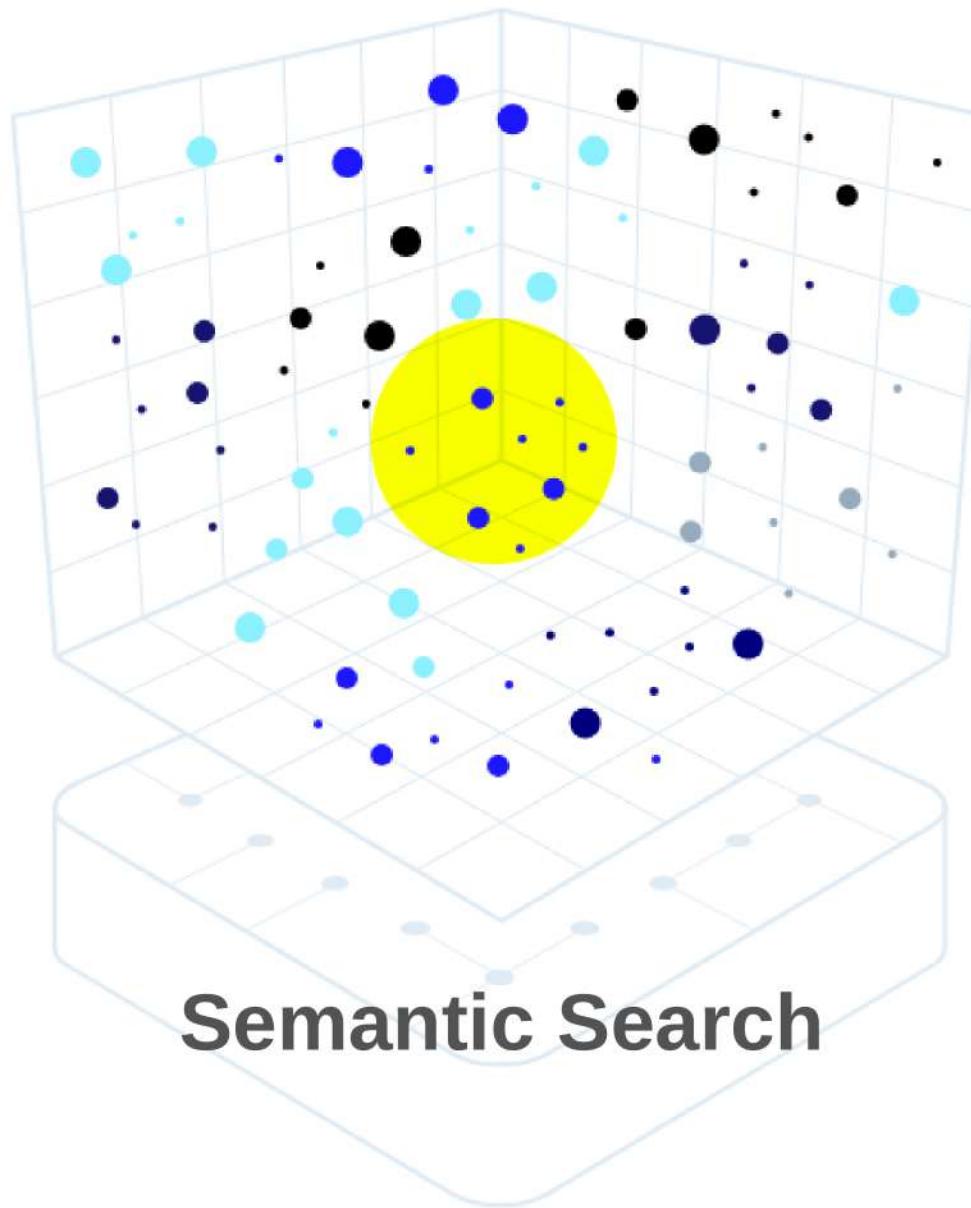
STEP 2 - Embedding & Semantic Search

SubQuery

신한은행 상품의 가입대상자는 무엇인가요?



STEP 2 - Embedding & Semantic Search



Cosine similarity top k=3 ranked results

- 1 {id_='6efb4ffa-ecb3-4d39-9358-26dc7d6d4a18', 'page_label': '4',
'file_name': '신한은행_땡겨요 적금.pdf'}
- 2 {id_='8867cdee-2f57-4170-8630-e9ea265cc1bb', 'page_label': '2',
'file_name': '신한은행_땡겨요 적금.pdf'}
- 3 {id_='3bfd9cb8-2a26-40a7-97dc-3bea6c817b7f', 'page_label': '6',
'file_name': '신한은행_땡겨요 적금.pdf'}

STEP 2 - Answering

Generated 2 sub questions.

[신한은행_땡겨요 적금.pdf] Q: 신한은행 상품의 가입대상자는 무엇인가요?

[신한은행_땡겨요 적금.pdf] A: 신한은행 상품의 가입대상자는 모든 성인이며, 신한은행 배달앱 “땡겨요” 서비스를 이용하거나 주단위 적금 납입 미션 달성 고객이라면 우대금리를 제공받을 수 있습니다. 또한, 민원 상담이 필요하시면 영업점, 고객상담센터(1577-8000) 또는 인터넷 홈페이지(www.shinhan.com)을 통해 문의하실 수 있으며, 금융소비자는 본 상품에 대하여 당 은행과 분쟁이 있는 경우에는 금융소비자보호법 제 36 조 제 1 항에 따라 금융감독원(국번없이 1332)에 분쟁조정을 신청할 수 있습니다. 또한, 신규일로부터 6개월 이내의 계약기간 동안 중도해지할 경우 약정한 이자율보다 낮은 중도해지 이자율이 적용됩니다.

[우리은행_WON플러스 예금.pdf] Q: 우리은행 상품의 가입대상자는 무엇인가요?

[우리은행_WON플러스 예금.pdf] A: 우리은행 상품의 가입대상자는 만 18세 이상의 개인고객이며, 금융소비자 보호에 관한 법률 및 동 법 시행령에 따라 금융상품에 관한 중요한 사항을 이해하고 이에 동의하는 사람들이며, 계좌당 1만원 이상의 금액을 가입하고, 1개월 이상 ~ 36개월 이하 (월, 일 단위 가입 가능) 기간으로 가입하는 사람들이며, 잔액의 95% 범위 내 가능하며, 예금자보호는 원금과 이자 포함 1인당 최고 5천만원까지 가능합니다. 또한, 계좌통합관리서비스 이용시 불필요한 계좌를 정리할 수 있으며, 휴면계좌통합조회 서비스 이용시 휴면계좌 현황을 파악할 수 있습니다. 상품 가입 전 반드시 특약 및 상품설명서를 참고하시고, 궁금하신 점은 영업점 직원 또는 우리은행 고객센터 (☎1588-5000)로 문의하시기 바랍니다.

STEP 2 - Answering

Response 1

[신한은행_땡겨요 적금.pdf] A: 신한은행 상품의 가입대상자는 모든 성인입니다. 적금 납입 미션 달성 고객이라면 우대금리를 제공받을 수 있습니다. 또한, 민원 또는 인터넷 홈페이지(www.shinhan.com)을 통해 문의하거나 상담이 있는 경우에는 금융소비자보호법 제 36 조 제 1 항에 따라 금융감독원(=국일로부터 6개월 이내의 계약기간 동안 중도해지할 경우 약정한 이자율보다

Response 2

[우리은행_WON플러스 예금.pdf] A: 우리은행 상품의 가입대상자는 만 18세 이상의 성인입니다. 법 시행령에 따라 금융상품에 관한 중요한 사항을 이해하고 이에 동의하는 사람 ~ 36개월 이하 (월, 일 단위 가입 가능) 기간으로 가입하는 사람들이며, 잔액 1인당 최고 5천만원까지 가능합니다. 또한, 계좌통합관리서비스 이용 시 블록 이용 시 휴면계좌 현황을 파악할 수 있습니다. 상품 가입 전 반드시 특약 및

OpenAI

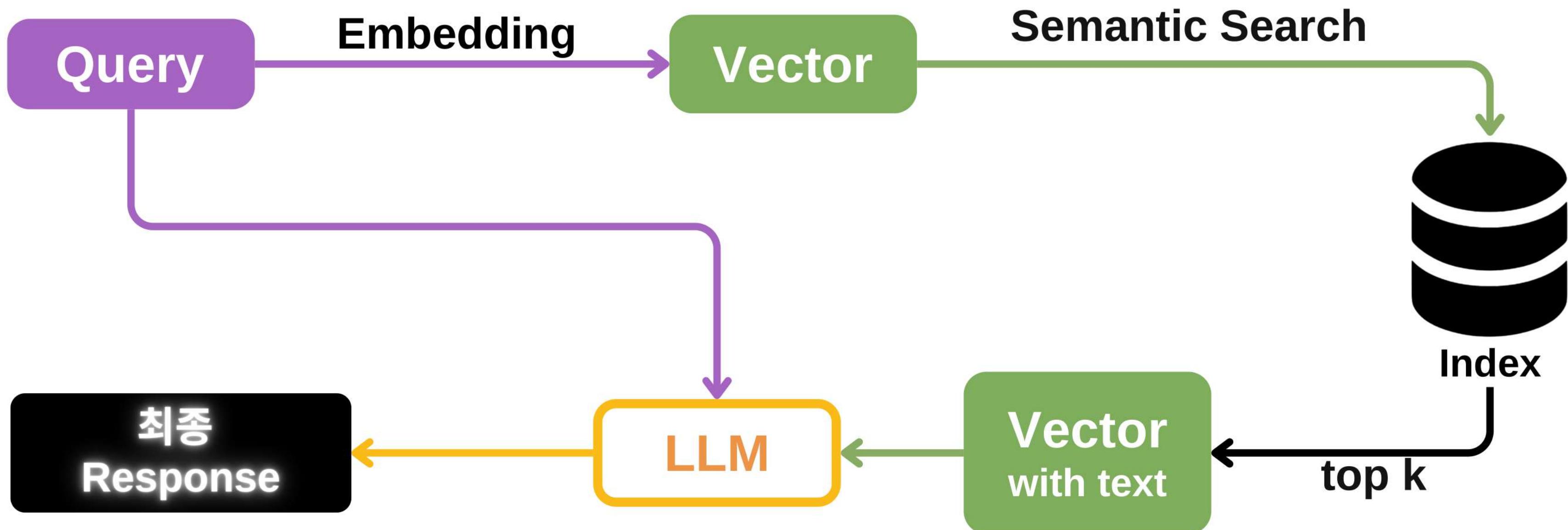
Query

최종 Response

신한은행과 우리은행 상품의 가입대상자는 모두 만 18세 이상의 성인입니다. 다만, 신한은행 상품의 경우 배달앱 “땡겨요” 서비스를 이용하거나 주단위 적금 납입 미션 달성 고객이라면 우대금리를 제공받을 수 있습니다. 또한, 우리은행 상품의 경우 계좌당 1만 원 이상의 금액을 가입하고, 1개월 이상 ~ 36개월 이하 (월, 일 단위 가입 가능) 기간으로 가입하는 사람들이며, 잔액의 95% 범위 내 가능하며, 예금자보호는 원금과 이자 포함 1인당 최고 5천만원까지 가능합니다. 또한, 신한은행과 우리은행 모두 금융소비자는 본 상품에 대하여 당 은행과 분쟁이 있는 경우에는 금융감독원(국번없이 1332)에 분쟁조정을 신청할 수 있습니다.

STEP 2 - Querying

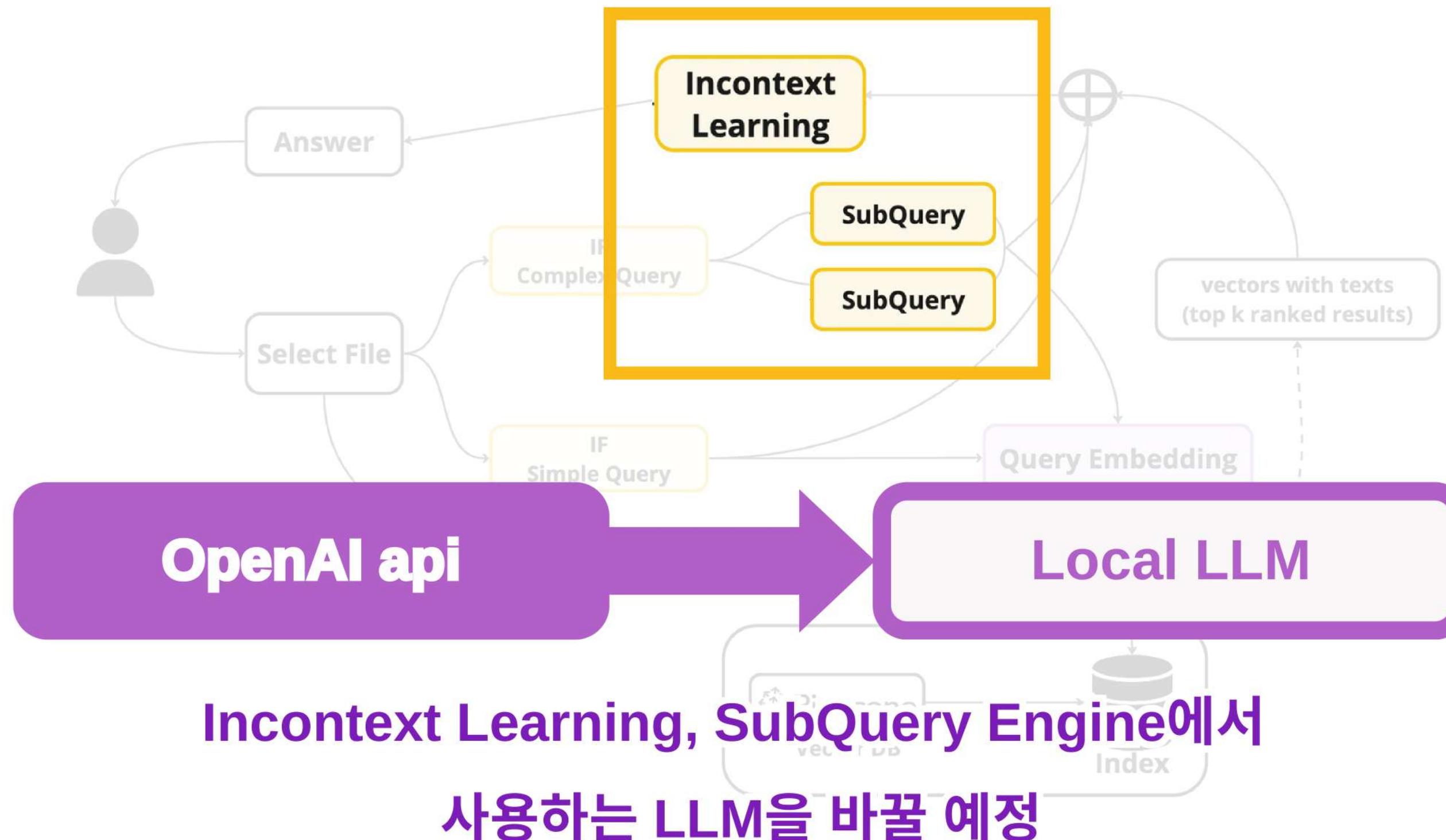
Simple Query



후반기 계획

사용 LLM 변경
Fine-tuning

후반기 계획



후반기 계획

OpenAI의 압도적 성능에도 Open Source LLM으로 바꾸는 이유

정보 유출 우려



대부분의 고객사는
(e.g. 금융사와 정부기관)
정보 유출에 민감

비용 문제



API 사용할 때마다
비용 발생

후반기 계획

Local LLM

Polyglot-Ko

Open source LLM - 다국어 거대 언어 모델 프로젝트

TuNiB AI에서 수집한 1.2TB 규모의 한국어로 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	-
Open subtitles	< 1.0	opus.nlpl.eu/OpenSubtitles.php
AIHub various tasks datasets	< 1.0	aihub.or.kr
Standard Korean language dictionary	< 1.0	stdict.korean.go.kr/main/main.do

후반기 계획

Fine-tuning

LoRA

Low-Rank Adaptation

원래 parameter는 freeze 시키고,

Transformer architecture의 각 layer마다 학습 가능한

Rank decomposition matrices를 추가하는 방식

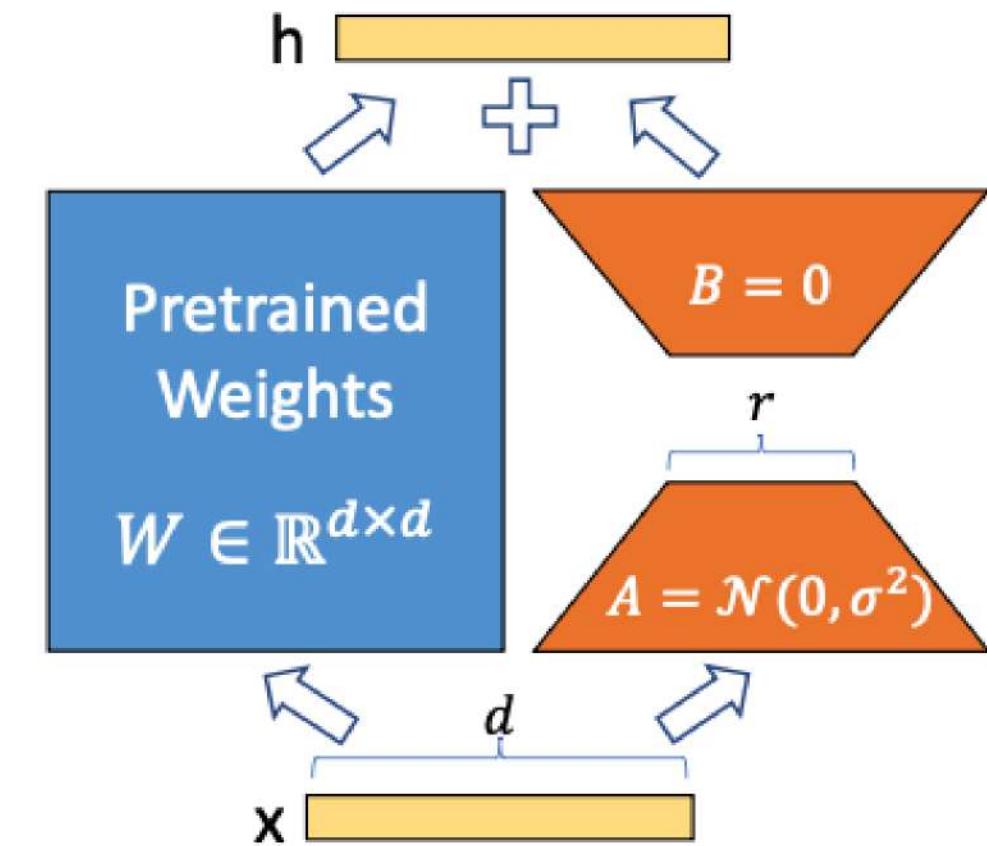


Figure 1: Our reparametrization. We only train A and B .

예전에 도메인에 맞는 Q&A 데이터셋을 생성해서 Fine-tuning 할 예정

THANK
YOU



