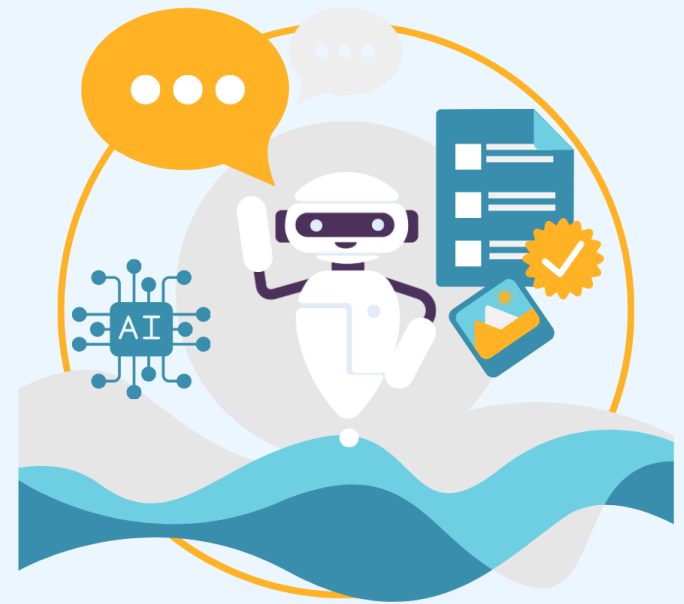


.ableRAG

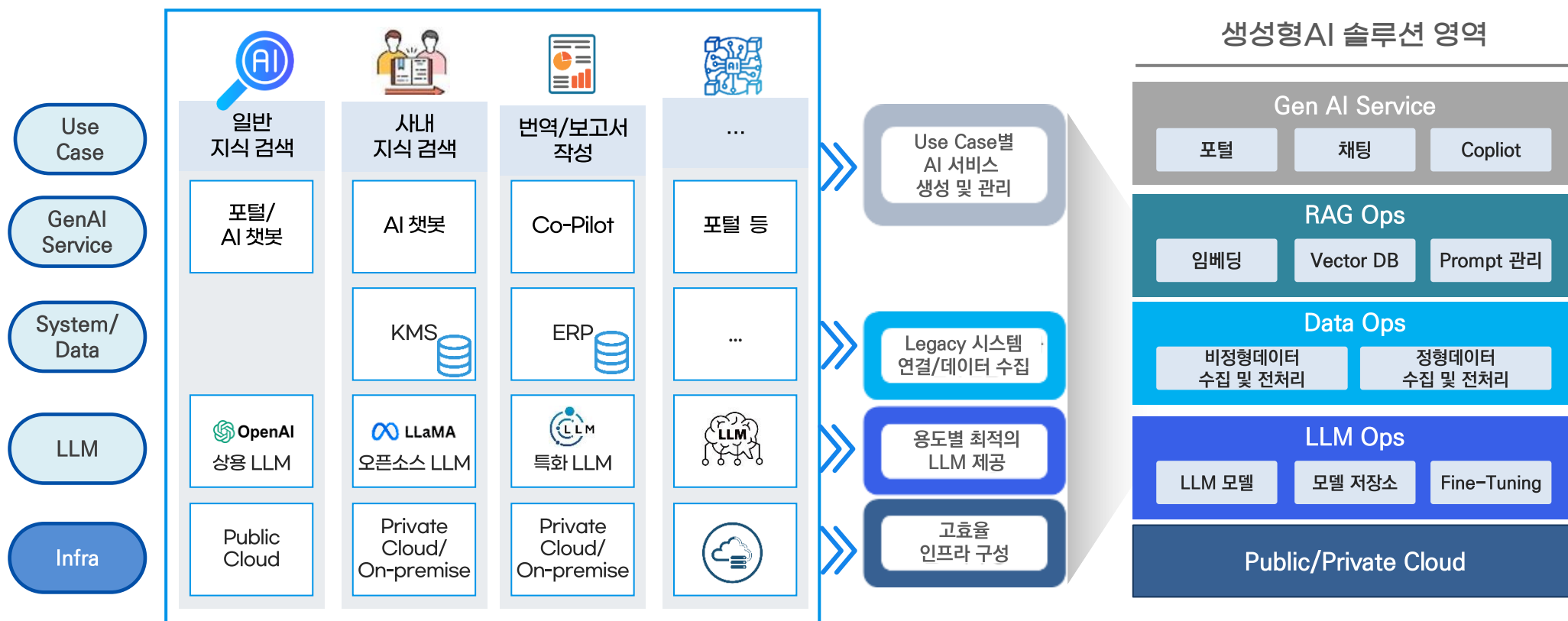
엔터프라이즈 AI Agent 솔루션

AbleRAG 제품 소개



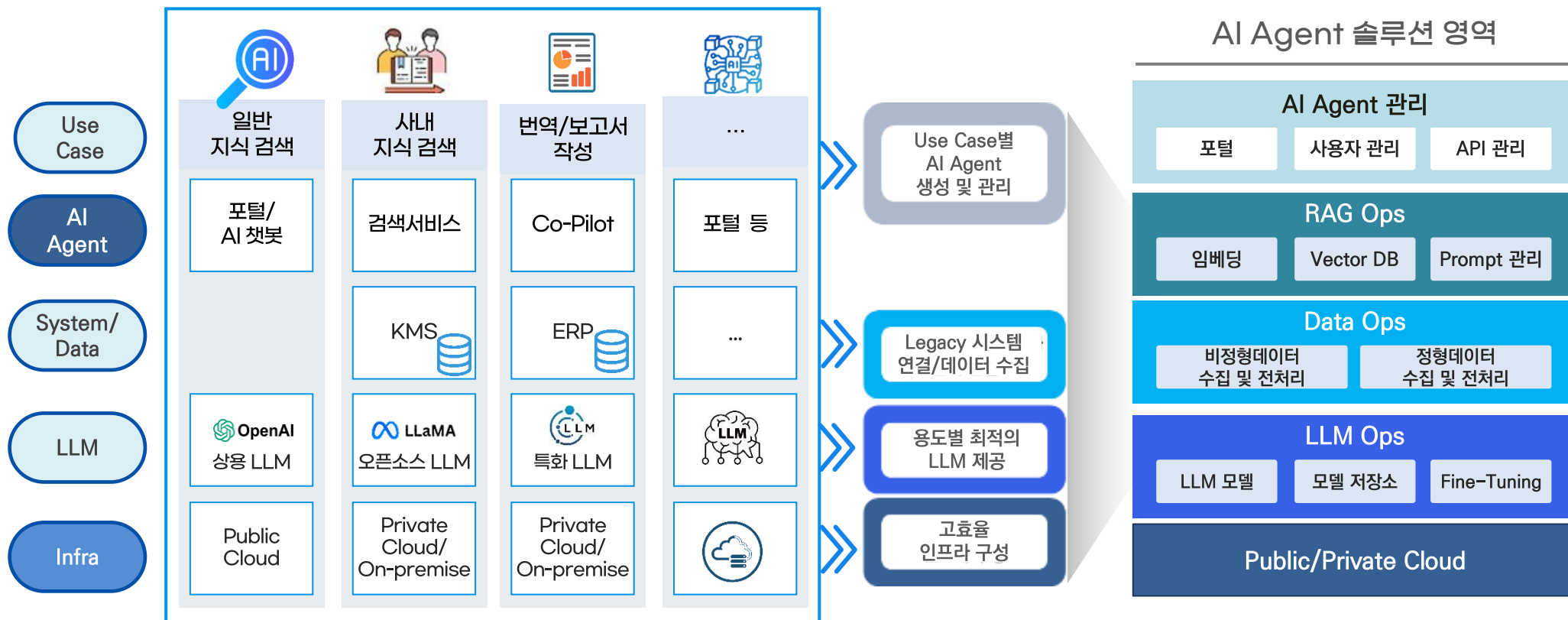
// 다양한 Use Case별 필요한 요소 기술 구현 및 생성형AI 솔루션 제공 영역

생성형 AI Use Case별 영역 및 기술 요소

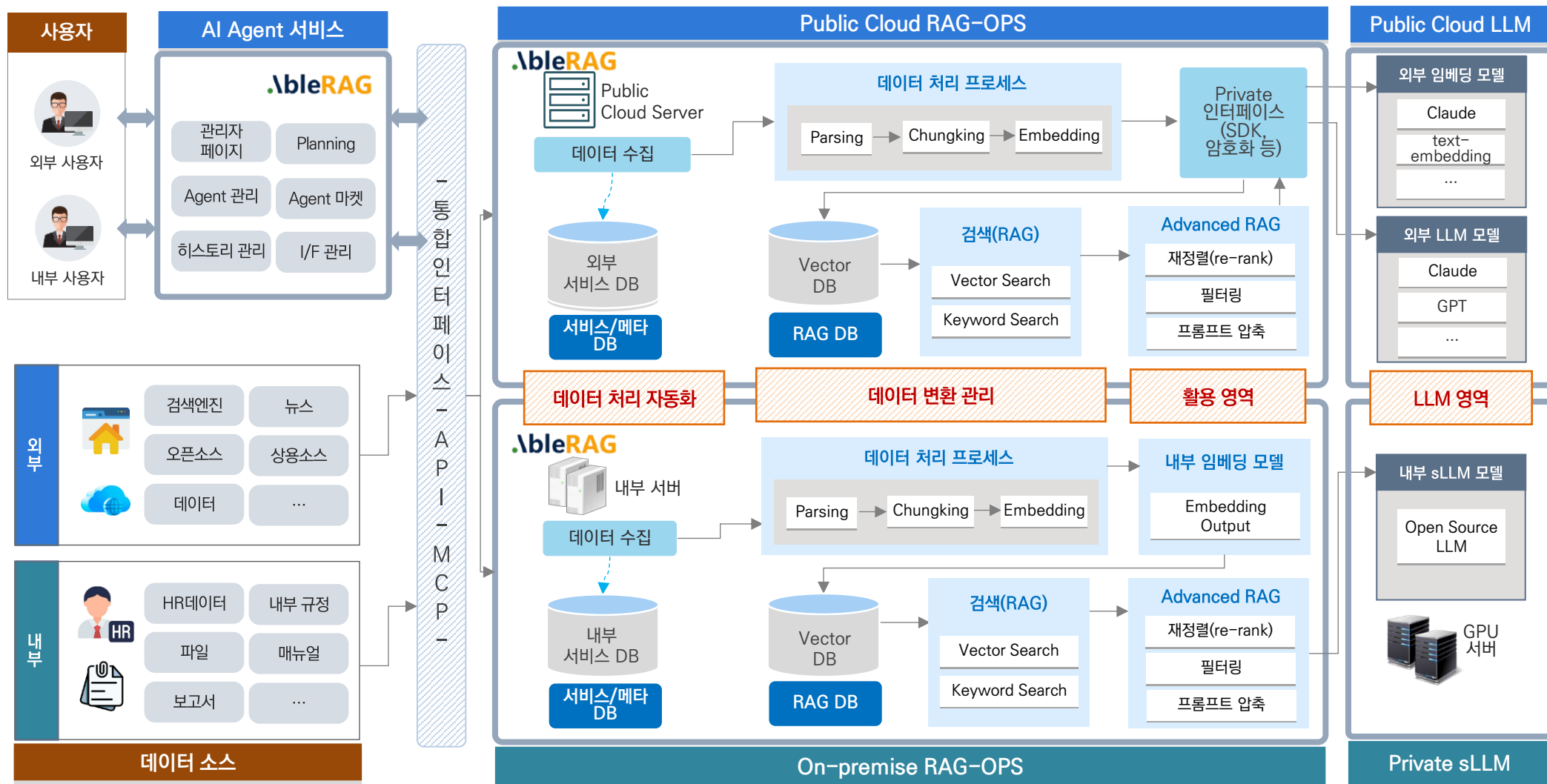


// 다양한 Use Case별 필요한 요소 기술 구현 및 생성형AI 솔루션 제공 영역

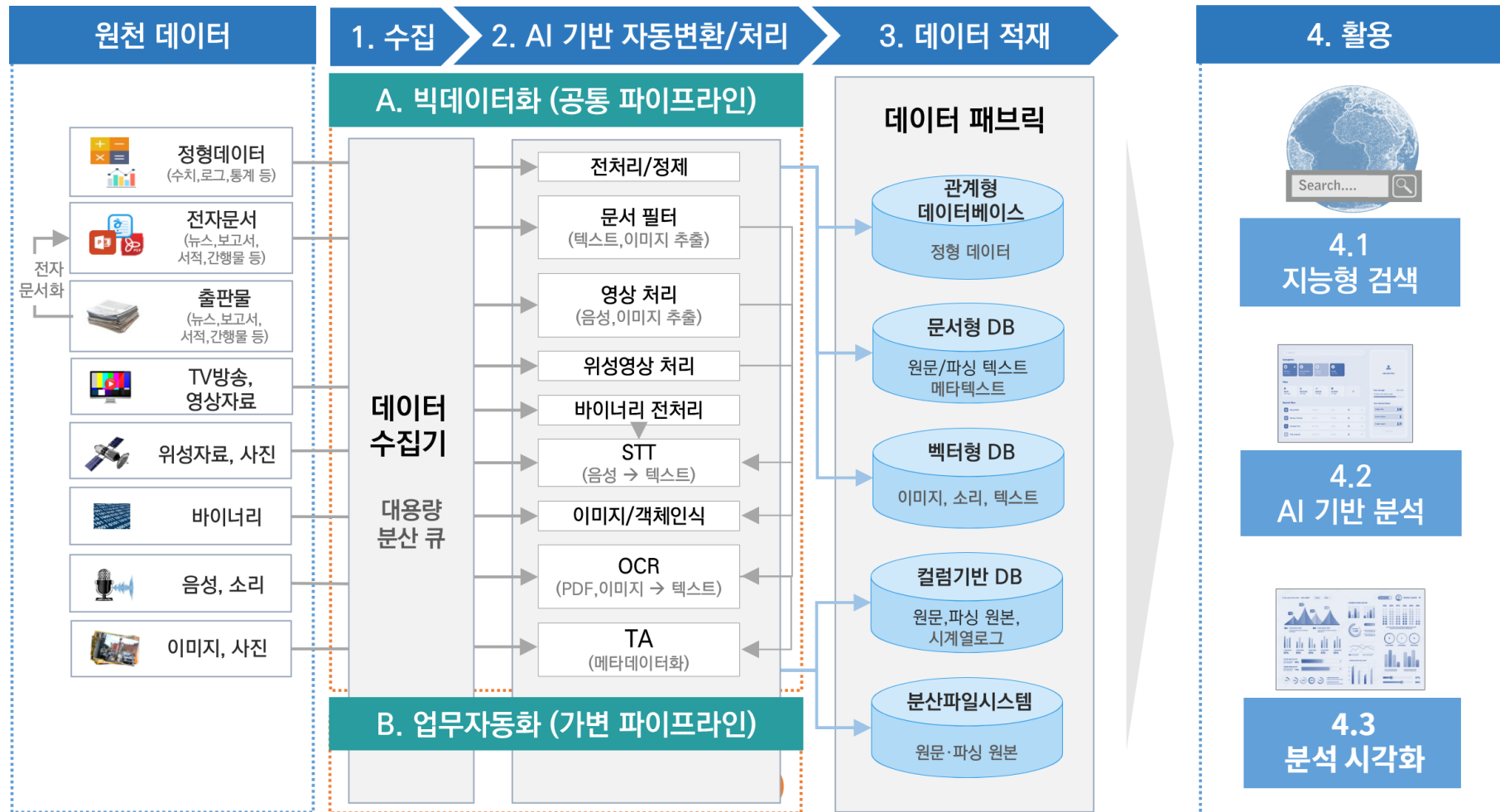
생성형 AI Use Case별 영역 및 기술 요소



Public Cloud 와 On-premise 환경에서 내/외부 데이터 연계 수집, 처리 및 LLM 활용



내/외부 원천 데이터 연계 수집 및 전처리 프로세스 자동화 구축



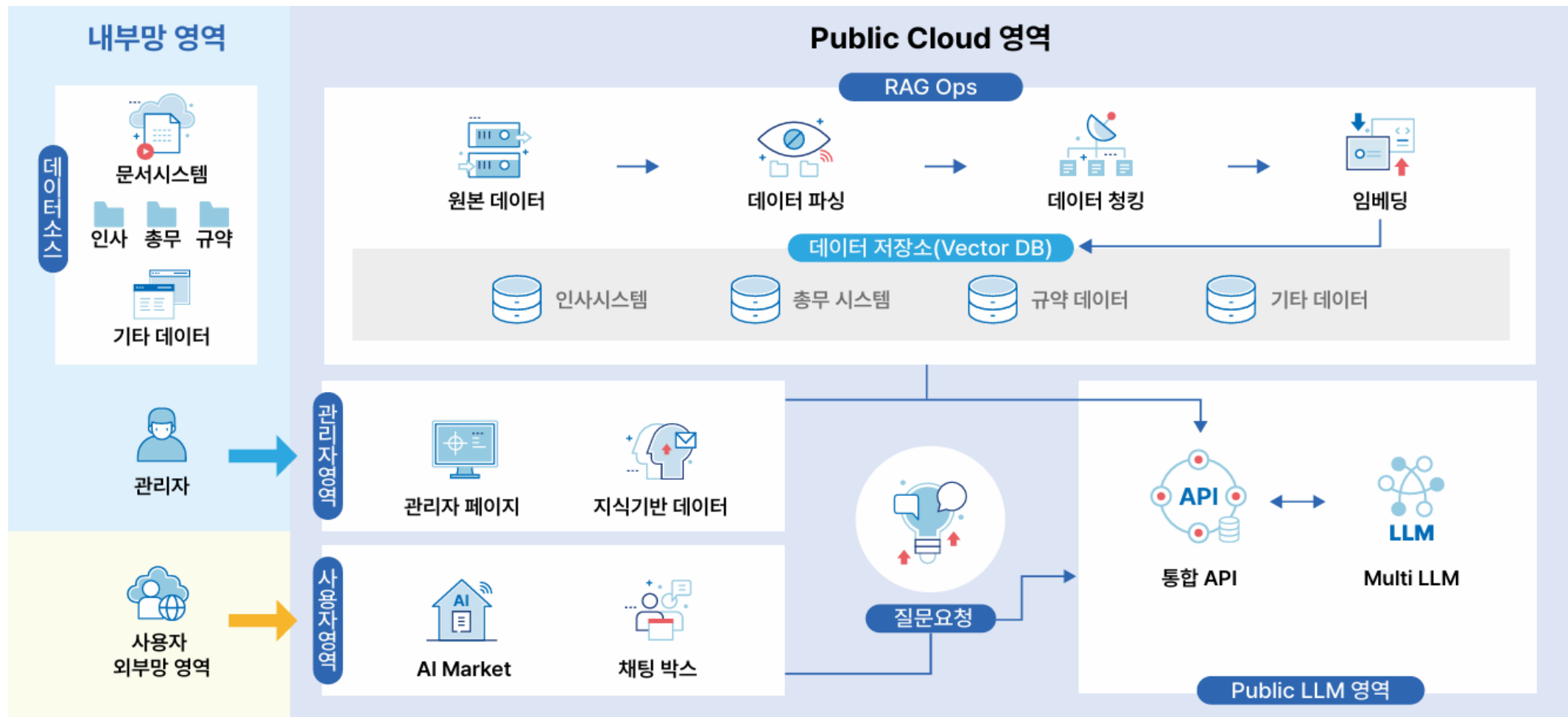
/// 맞춤 대화형 AI 챗봇 통해 헬프 데스크를 비롯한 각종 내부 직원 문의 응대 업무 자동화



// AI 챗봇, 비서, 인사 등 기업 내부 업무 자동화 Agent 기능 구현



/// 내부 데이터를 벡터화해 저장하고 이를 기반으로 외부 LLM과 연동해 정확한 답변 제공



/// 다양한 AI 앱을 빠르게 탐색하고 미리보기와 모델 정보 확인 후 클릭 한 번으로 AI 앱 실행

The image shows a web application interface for 'AbleRAG'. At the top, there's a blue header with the text '앱 마켓' (App Market). Below it, a navigation bar contains a '+ 새로 만들기' (New) button, a search bar with the text '찾고싶은 AI' (AI I want to find), and the 'AbleRAG' logo. The main area displays a grid of app cards. The first row includes '기본 앱' (Basic App) with a description '자유로운 주제로 언어모델과 대화할 수 있습니다.' and a '미리보기' (Preview) button; '명언 추천기' (Quote Recommendation) with '명언 데이터 기반으로, 원하는 느낌의 명언을 추천해줍니다...' and a '미리보기' button; and '미국 영화 DB' (US Movie DB) with '1900년도부터 2023년 사이의 미국 영화 데이터 정보를 조회할 수 있습니다...' and a '미리보기' button. The second row includes '영어 번역' (English Translation) with '영어 문장을 자연스럽게 정확하게 번역해줍니다.' and a '미리보기' button; '중국어 번역' (Chinese Translation) with '중국어 문장을 한국어로, 또는 그 반대로 정확하게 번역해줍니다.' and a '미리보기' button; and '일본어 번역' (Japanese Translation) with '일본어 문장을 매끄럽게 번역하고, 의미를 이해하기 쉽게 풀어줍니다.' and a '미리보기' button. On the right side, there are two detailed panels. The top panel, titled '기본 앱', shows '언어모델' (Language Model) as 'Claude 3.5 sonnet', '임베딩 모델' (Embedding Model) as 'Titan Text Embeddings V2', and '지식기반' (Knowledge Base) as 'OpenSearch', with a '사용 중' (In Use) button. The bottom panel, titled '명언 추천기', shows the same model and knowledge base information, but with a '사용하기' (Use) button. At the bottom of the interface, there's a pagination bar with '<<' '<' '1' '>' '>>'.

/// 다양한 모델 선택과 문서 임베딩을 통해 목적에 맞는 AI 앱 쉽게 생성 및 간편한 설정

* 언어모델

Claude 3 Haiku

Claude 3 Haiku

Claude 3.5 Sonnet

Claude Haiku 3.5

Claude Sonnet 3.7

Claude Sonnet 4 (2025-5-14)

Claude-V3-Sonnet

Gemini 2 Flash

Gemini 2 Flash lite

Gemini 2.5 Pro preview 3/25

GPT 4.1

GPT 4.1 mini

GPT 4o

* 임베딩 모델

Titan Text Embeddings V2

Titan Text Embeddings V2

Text Embedding 3 small

Text Embedding 3 large

새 AI 앱 만들기

1

기본 설정

2

프롬프트 테스트

3

공개 여부

AI 템플릿 이름

설명

AI 템플릿에 대한 간단한 설명을 입력하세요 #해시태그

* 언어모델

Claude 3 Haiku

* 임베딩 모델

Titan Text Embeddings V2

* 지식기반

OpenSearch

임베딩 진행중

임베딩 중... 일반적으로 5분 내외 소요됩니다.
파일 크기에 따라 시간이 더 오래 걸릴 수 있습니다.

임베딩 파일 선택

추가

파일명

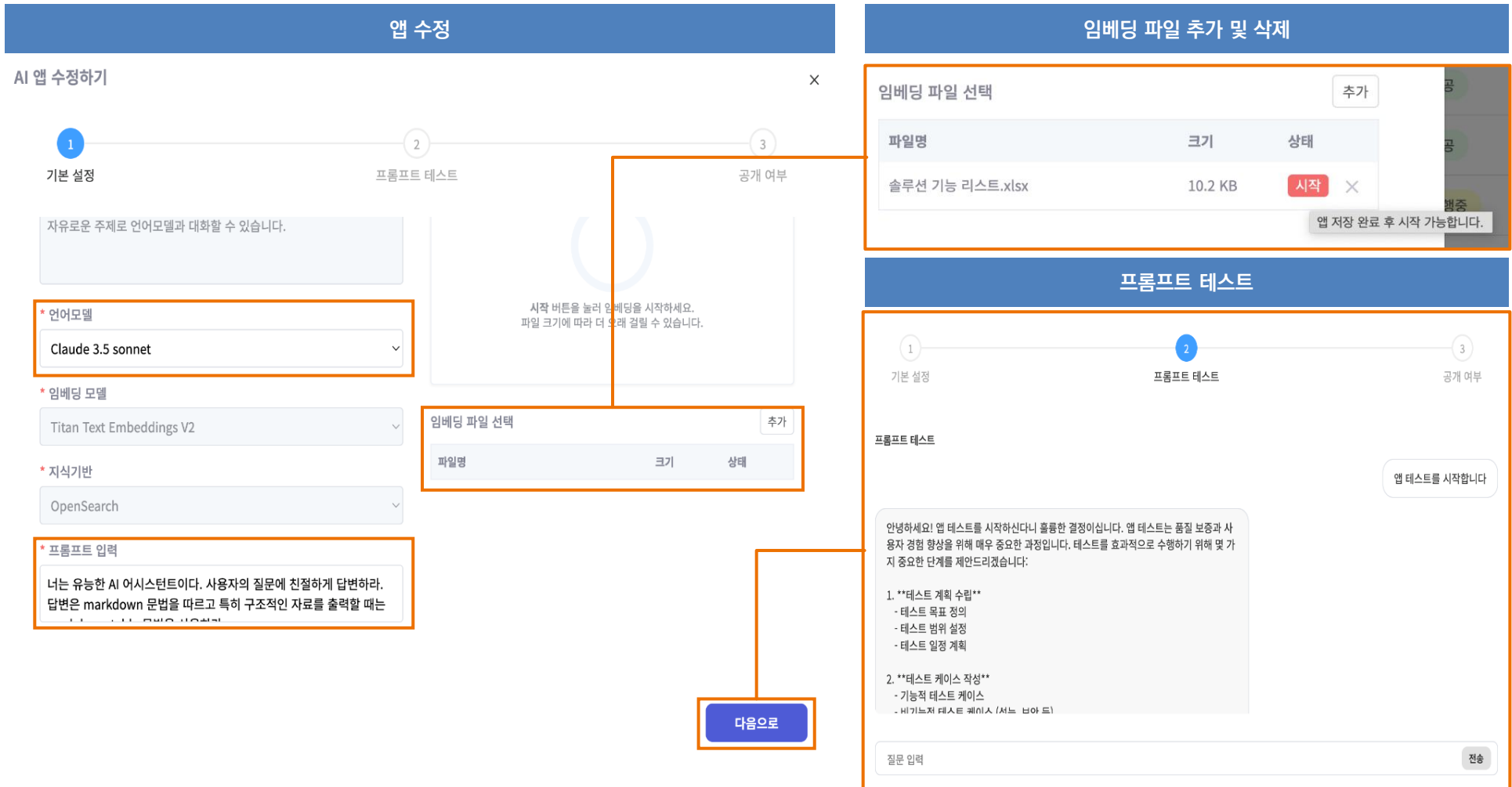
크기

상태

임시저장

다음으로

// 임베딩 파일 추가와 단계별 설정이 직관적으로 이루어져 손쉬운 앱 개선 가능



/// 실행, 권한, 수정 등 주요 기능 직관적으로 제어할 수 있는 최적화된 UI 제공 및 작업 진행 상황 확인

전체 ▾

앱 이름	앱 Ver.	언어 모델	임베딩 모델	생성일	작업 현황	관리
한국어 번역	1	anthropic.claude-3-5-sonnet-20240620-v1:0	amazon.titan-embed-text-v2:0	2025-02-22 11:12:06	실패	▶ 📁 ⚙️ 🗑️
미국 영화 DB	1	anthropic.claude-3-5-sonnet-20240620-v1:0	amazon.titan-embed-text-v2:0	2025-02-22 11:12:06	진행중	▶ 📁 ⚙️ 🗑️
일본어 번역	1	claude-sonnet-4-20250514	amazon.titan-embed-text-v2:0	2025-02-21 11:12:06	성공	▶ 📁 ⚙️ 🗑️
중국어 번역	1	gpt-4.1	text-embedding-3-large	2025-02-20 11:12:06	성공	▶ 📁 ⚙️ 🗑️
영어 번역	1	gemini-2.0-flash	amazon.titan-embed-text-v2:0	2025-02-18 11:12:06	성공	▶ 📁 ⚙️ 🗑️

관리

▶ 📁 ⚙️ 🗑️

▶ 📁 ⚙️ 🗑️

▶ 📁 ⚙️ 🗑️

바로가기

권한 관리

AI 앱 권한 관리

구분	이름	삭제
사용자	현재용	🗑️
사원사	전용자	🗑️

🔍 Search 시뮬레이션 부서검색

부서	선택
AI사업팀	추가
번역부드사업팀	추가
테스트	추가

이전 1 다음

수정하기

AI 앱 수정하기

1 기본 설정 2 프롬프트 텍스트 3 공개 여부

기본 설정

이름: 기본 앱

설명: 지능로운 주제를 언어모델과 대화할 수 있습니다.

언어모델: Claude 3.5 sonnet

임베딩 모델: Titan Text Embeddings V2

저색기법: OpenSearch

임베딩 현황

임베딩 대기중

시작 버튼을 눌러 임베딩을 시작하세요. 시퀀스 크기에 따라 몇 초에 걸릴 수 있습니다.

임베딩 파일 선택

파일명 크기 상태

추가 삭제

다음으로

삭제

삭제

// 선택한 앱 기반으로 실시간 대화를 수행하며, 드롭 다운과 필터로 앱 전환과 기록 조회 제공

The screenshot displays the AbleRAG application interface. At the top, there is a header bar with the .ableRAG logo and the title 'AbleRAG 솔루션 기능 - 앱 사용'. Below the header, a navigation bar shows the selected app '기본 앱 (ver.1)' and the user 'ADMIN 이만경'. The main chat area features a large '.ableRAG' logo and a text input field with a placeholder '메시지를 작성하세요'. To the right of the chat area is a filter panel with a dropdown menu '기본 앱 (ver.1)' and a list of filters: '중국어 번역 (ver.1)', '기본 앱', and '중국어 번역'. The chat history on the left shows two messages: '안녕하세요' (2025-06-22) and '안녕하세요 를 번역해줘' (2025-06-20). The chat area shows a response '안녕하세요 를 번역해줘' (오전 10:41) and a follow-up message '죄송합니다. 중국어 번역과 무관한 작업 요청에는 답변해드릴 수 없습니다.' (오후 9:17).

// 사용자 또는 부서 단위로 앱 접근 권한 설정 기능 제공 및 앱별 사용자 목록으로 체계적인 운영 지원

AI 앱 권한 관리

기본 앱

구분	이름	삭제
사용자	황태웅	
사용자	정동진	

Search

사용자검색

부서검색

부서	선택
AI사업팀	
클라우드사업팀	
테스트	

이전 1 다음

취소

저장

demo

사용자검색

부서검색

부서	사용자	선택
test	demo	
test	demo	

이전 1 다음

Search

사용자검색

부서검색

부서	사용자	선택
AI사업팀	test	
AI사업팀	여지혜	
AI사업팀	김종호	
AI사업팀	문영은	
AI사업팀	박성호	

이전 1 2 3 다음

AI APP 수정 시 버전관리를 통한 Roll back 기능 및 업데이트 기능 제공

앱 관리

전체

앱 이름	언어 모델	임베딩 모델	생성일	앱 ver.	앱 공개 여부	작업 현황	관리
인사시스템	GPT 4o	Titan Text Embeddings V2	2025-07-24 14:46:08	2	N	성공	   

인사시스템	GPT 4o	Titan Text Embeddings V2	2025-07-24 14:45:22	1
-------	--------	--------------------------	---------------------	---

앱 관리

전체

앱 이름	언어 모델	임베딩 모델	생성일	앱 ver.	앱 공개 여부	작업 현황	관리
인사시스템	Claude Sonnet 4 (2025-5-14)	Titan Text Embeddings V2	2025-07-24 16:54:28	3			
인사시스템	GPT 4o	Titan Text Embeddings V2	2025-07-24 14:46:08	2			
인사시스템	GPT 4o	Titan Text Embeddings V2	2025-07-24 14:45:22	1	N	성공	   

AI 앱 수정하기

1 기본 설정 2 프롬프트 테스트 3 공개 여부

* 이름
인사시스템

설명
25년 혁신 인사 데이터 및 규정 데이터

* 언어모델
Claude Sonnet 4 (2025-5-14)

* 임베딩 모델
Titan Text Embeddings V2

* 지식기반
OpenSearch

* 프롬프트 입력
다음 내용에 대해서 회사 내규 사내 도우미로서 대화를 진행해줘

임베딩 현황

임베딩 대기중

시작 버튼을 눌러 임베딩을 시작하세요.
파일 크기에 따라 더 오래 걸릴 수 있습니다.

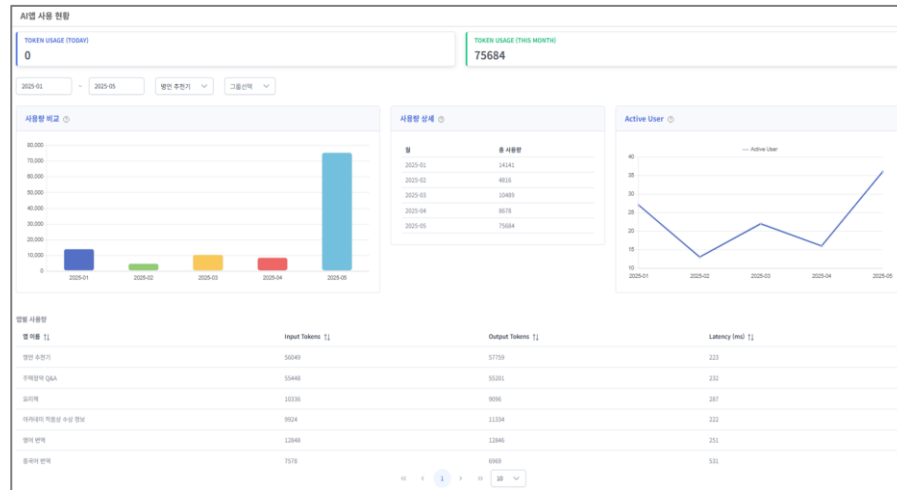
임베딩 파일 선택

파일명	크기	상태
1_기초자료+재용기능+성공안내용.ppt	263.3 MB	선택
IV. 프로젝트 수행 및 관리 방안_V1.1.pptx	4.5 MB	삭제
III. 사업 추진 방안_V1.0.pptx	18.0 MB	삭제
(무)고보라물직장인보원_20250619_12001_...	2.7 MB	삭제
(무)고보라물연금보험(유니버설)_20250619_...	1.4 MB	삭제

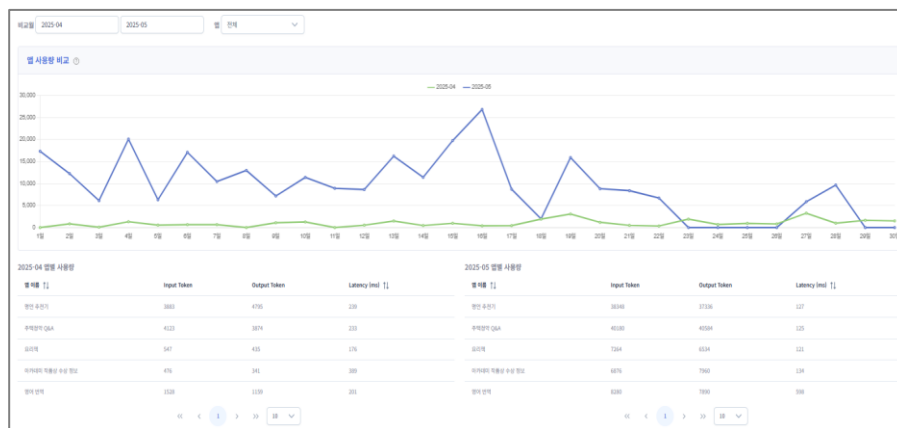
다음으로

AI 서비스의 사용량, 응답 시간, 사용자 활동 등 대시보드 시각화로 실시간 모니터링 지원

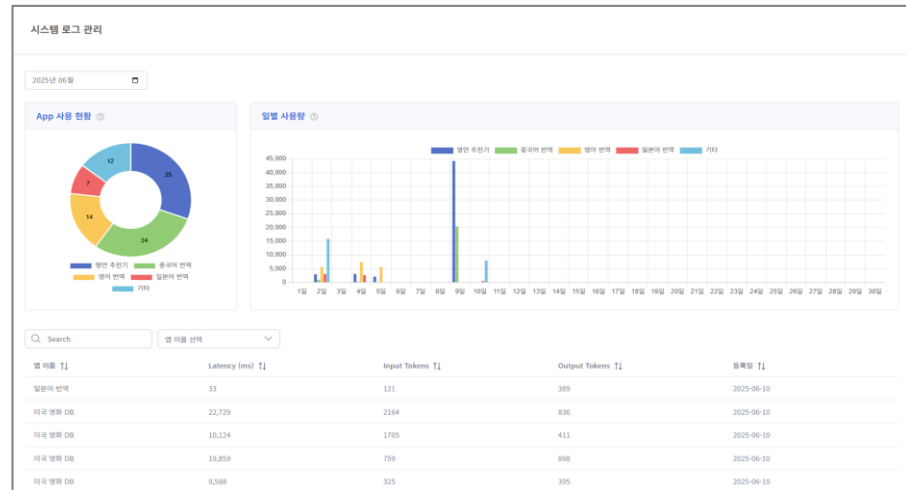
대시보드



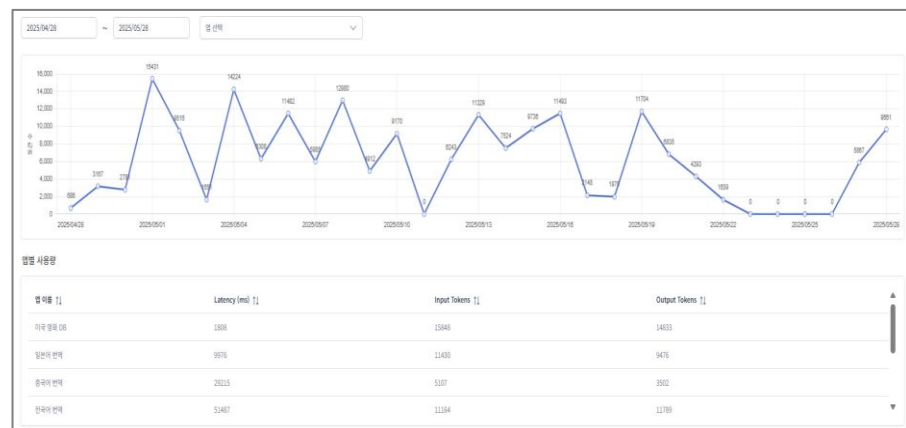
월별 사용량 비교



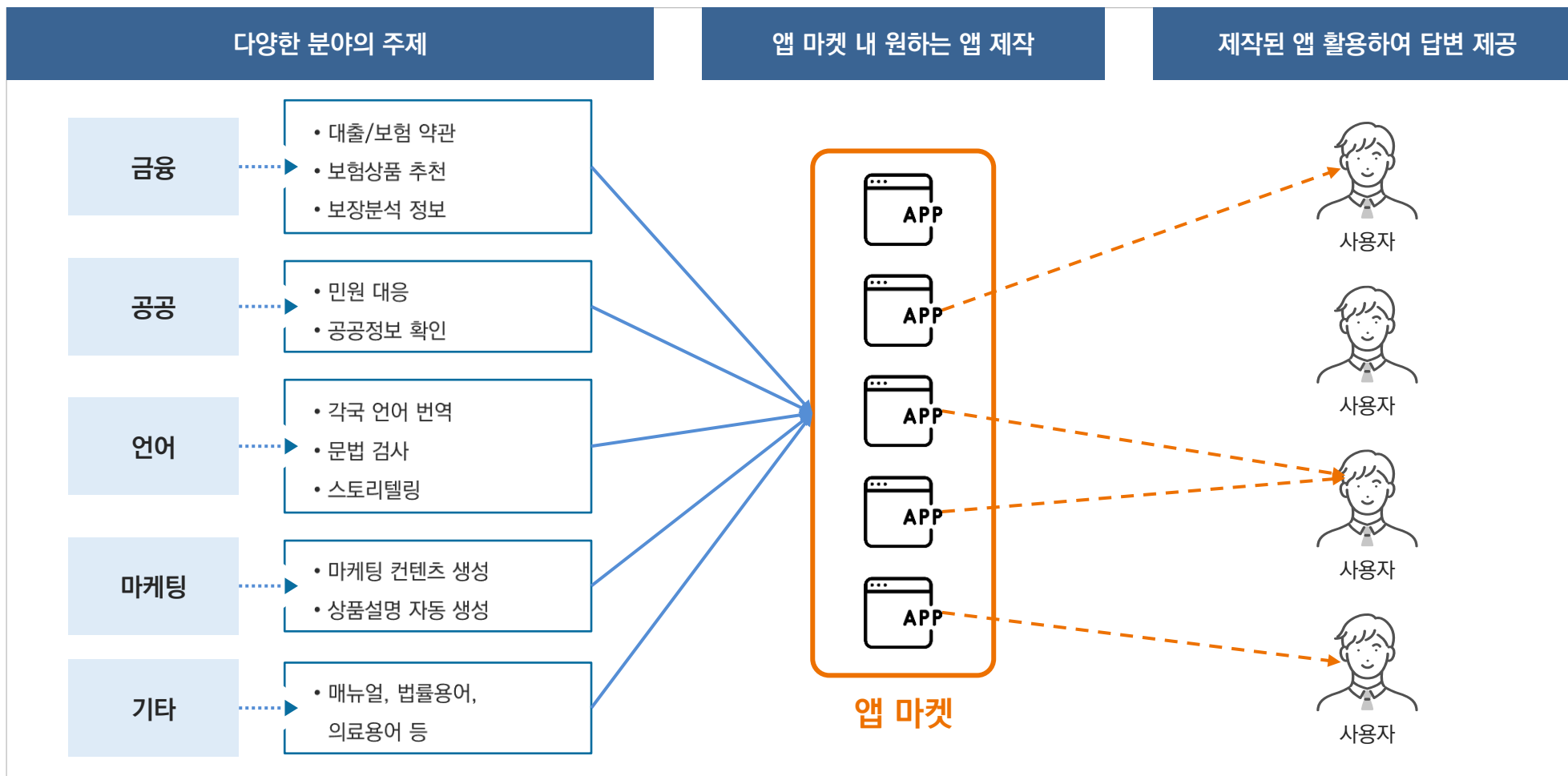
시스템 로그 관리



기간별 사용 현황 비교



// 금융, 공공 등 다양한 분야의 AI 앱을 제작하고 활용할 수 있는 앱마켓 기능 제공



/// 사용자의 의도를 명확하게 파악하기 위해 AI AGENT 추론기능

AI Agent 기능 영역



고객사 도메인 특정 데이터 사용

- 의도 파악 및 정밀도 향상
- 사용자 질문이나 발화가 모호할 경우 AI 반복 추론을 통한 다각도 문맥해석

답변 목표 구체화

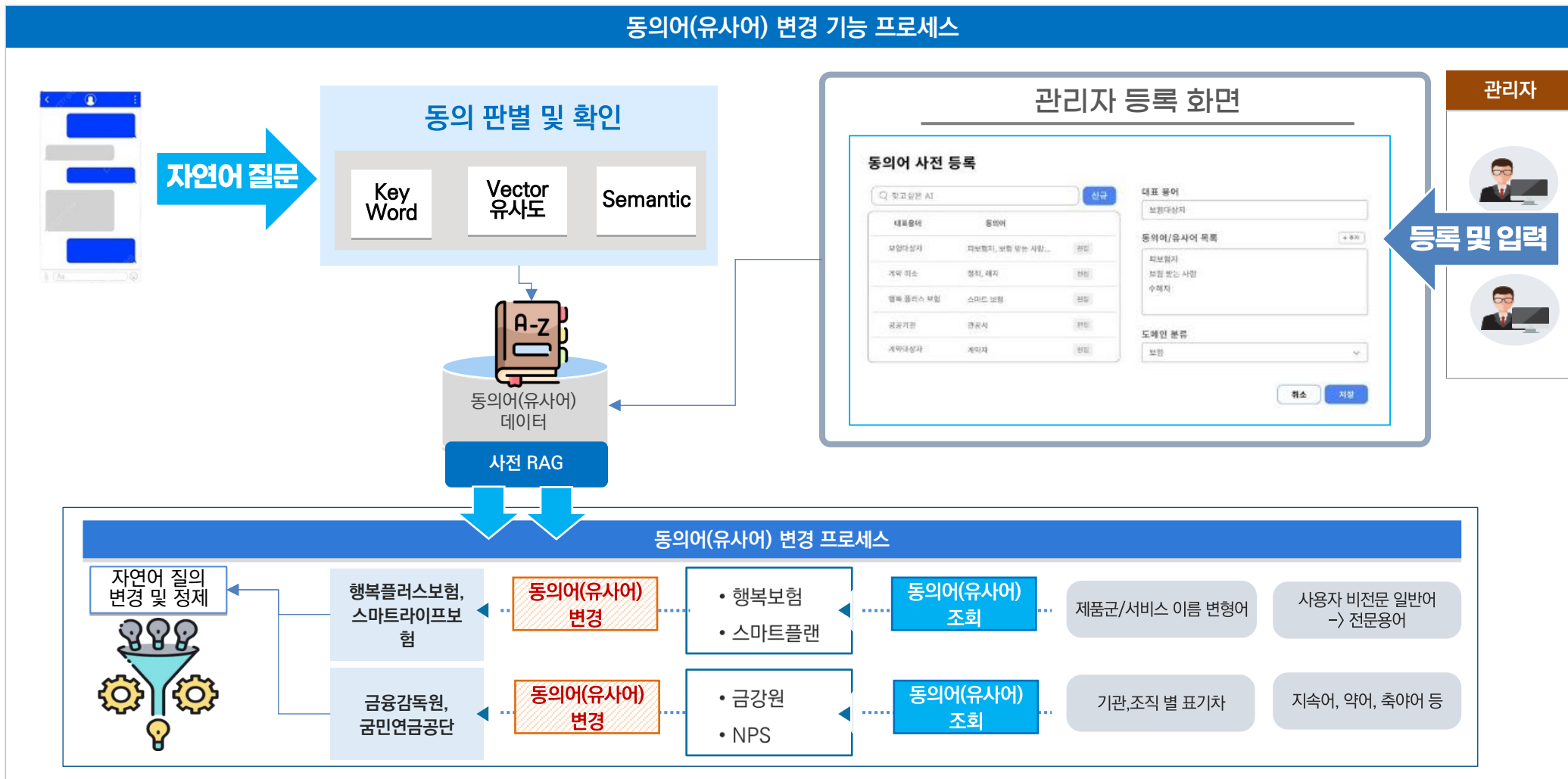
- 일반적인 초기 답변을 LLM 반복 검토 루프를 통해 답변의 충분 여부 판단
- 부족한 경우 목표에 부합하는 방향성 재구성 및 재답변 생성

분산 데이터 소스 탐색

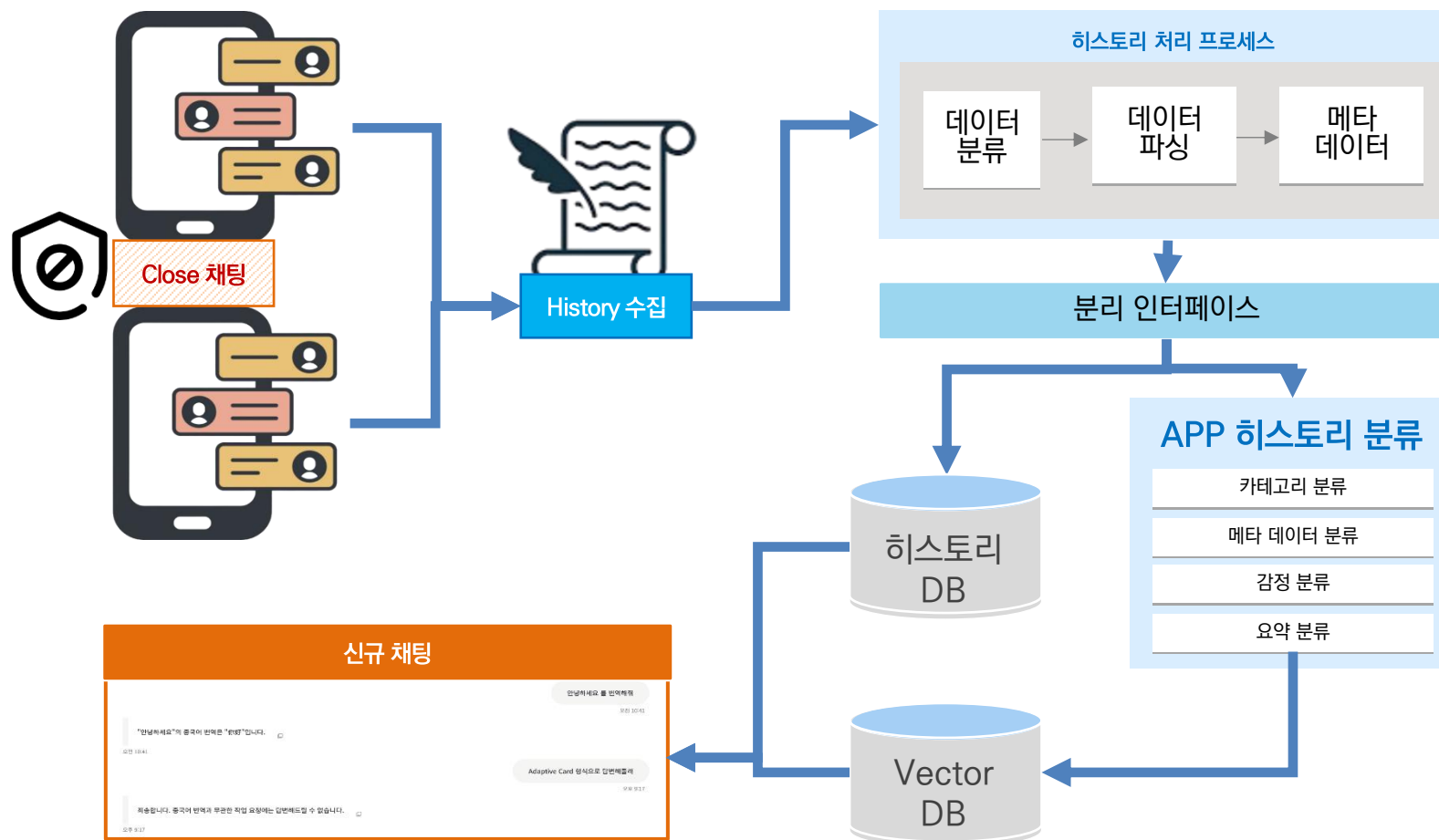
- 사용자 질문 및, 키워드, 문맥 등을 기반으로 가장 높은 데이터 소스 후보 분류
- 후보 데이터 대상으로 유사도 및 내부 규칙 기반 필터링 수행
- 답변 검증 및 루프를 통한 잘못된 데이터 소스 배제

AI 챗봇, 비서, 인사 등 기업 내부 업무 자동화 Agent 기능 구현

동의어(유사어) 변경 기능 프로세스



History RAG를 통하여 APP별 특화 답변 및 사용자 의도파악 향상



의도파악 및 정밀도 향상

- 신규 대화에서 추출된 의도와 문맥을 기반으로 과거 유사 대화를 탐색하고, 해당 APP 전용 RAG에 연결함으로써 APP의 도메인 대응력과 응답 일관성을 높임

종료 대화 자산화 및 재사용

- 폐쇄된 대화 기록을 벡터 및 키워드 기반으로 구조화하여 APP 단위로 축적하고, 이후 유사 질문 발생 시 의도 예측 및 응답 품질을 지속적으로 개선함

// 금융, 공공 등 다양한 분야의 AI 앱을 제작하고 활용할 수 있는 앱 마켓 기능 제공

RDB 형식의
Vector DB 구성
(PostgreSQL)



RDB 형식의 데이터
저장 및 조인 기능



데이터베이스의
성능, 확장성, 가용성의
확보 가능

Vectorization

Data record 1

Data record 2

Data record 3



vector		
[0,27, 0,68, ..		
[0,53, -0,44, ..		
[0,01, -0,85, ..		

POSTGRESQL

OpenSearch
Vector DB
구성



하이브리드 검색 지원
(벡터 검색, 키워드검색)



고성능 및 대규모
Vector 유사성 검색 지원

```
PUT my-index
{
  "settings": { "index.knn": true },
  "mappings": {
    "properties": {
      "my_vector1": { "type": "knn_vector", "dimension": 2 },
      "my_vector2": { "type": "knn_vector", "dimension": 4 }
    }
  }
}

GET my-index/_search
{
  "size": 2,
  "query": {
    "knn": {
      "my_vector2": {
        "vector": [2, 3, 5, 6],
        "k": 2
      }
    }
  }
}
```

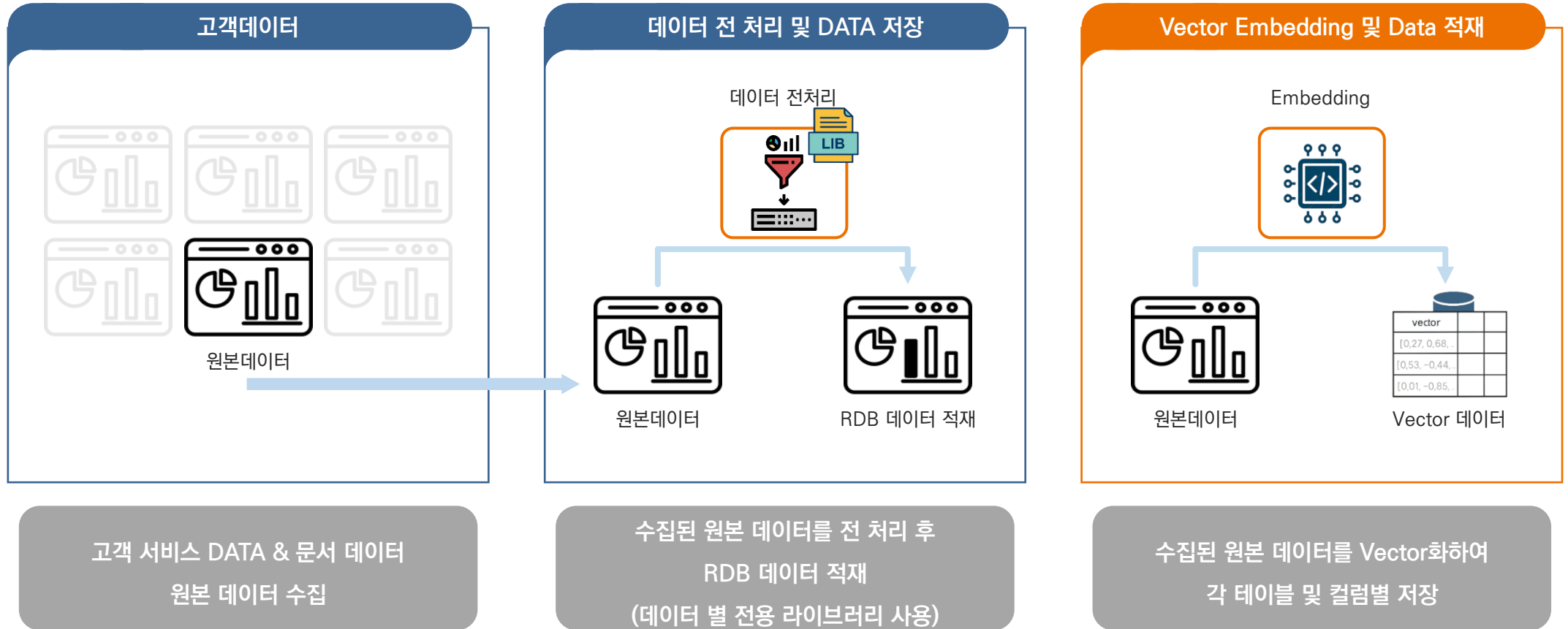
```
POST _bulk
{ "index": { "_index": "my-index", "_id": "1" } }
{ "my_vector1": [1, 1, 2, 3], "price": 10.2 }
{ "index": { "_index": "my-index", "_id": "2" } }
{ "my_vector1": [2, 1, 3, 3], "price": 7.1 }
{ "index": { "_index": "my-index", "_id": "3" } }
{ "my_vector1": [3, 1, 4, 3], "price": 12.9 }
{ "index": { "_index": "my-index", "_id": "4" } }
{ "my_vector1": [5, 1, 6, 3], "price": 1.2 }
{ "index": { "_index": "my-index", "_id": "5" } }
{ "my_vector1": [6, 1, 5, 3], "price": 2.7 }
{ "index": { "_index": "my-index", "_id": "6" } }
{ "my_vector2": [1, 1, 5, 5, 4, 5, 6, 6], "price": 10.3 }
{ "index": { "_index": "my-index", "_id": "7" } }
{ "my_vector2": [2, 1, 3, 5, 6, 6, 7], "price": 5.5 }
{ "index": { "_index": "my-index", "_id": "8" } }
{ "my_vector2": [5, 1, 5, 6, 7, 3, 7], "price": 4.4 }
{ "index": { "_index": "my-index", "_id": "9" } }
{ "my_vector2": [1, 1, 5, 4, 5, 6, 6], "price": 1.9 }
```

```
GET my-index/_search
{
  "size": 2,
  "query": {
    "knn": {
      "my_vector2": {
        "vector": [2, 3, 5, 6],
        "k": 2
      }
    }
  }
}
```

혼합 검색 (Hybrid search)

```
GET my-index/_search
{
  "query": {
    "bool": {
      "must": [
        { "knn": {
          "my_vector2": {
            "vector": [2, 3, 5, 6],
            "k": 2
          }
        }
      ],
      "filter": {
        "range": {
          "price": {
            "gte": 5,
            "lt": 10
          }
        }
      }
    }
  }
}
```

/// 원본 데이터를 정제해 RDB에 저장, 임베딩하여 Vector DB에 적재함으로써 검색 기반 강화



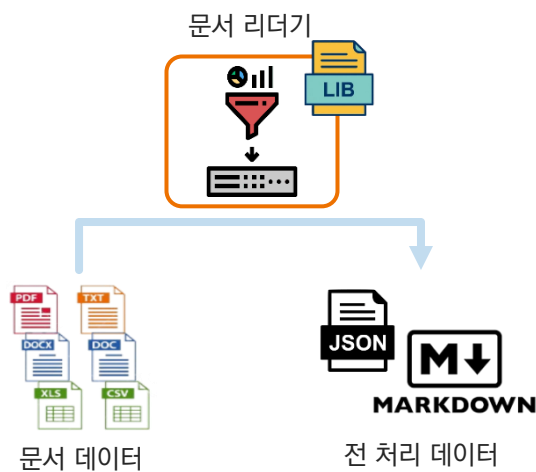
/// 다양한 문서 포맷을 자동 전처리, JSON/Markdown 형태로 저장 및 벡터 데이터 저장

다양한 문서 포맷



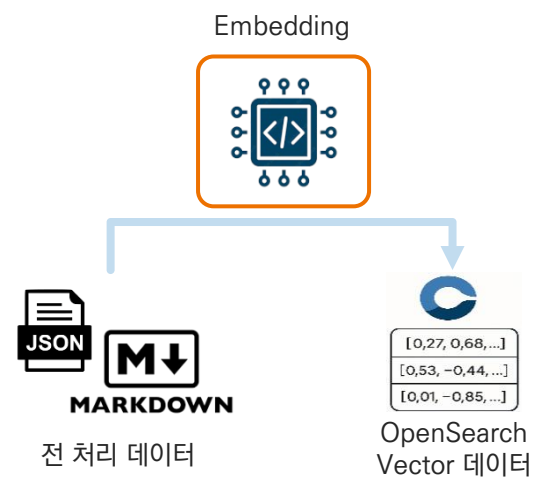
다양한 형식의 문서 DATA 수집

데이터 전 처리 및 DATA 저장



수집된 원본 데이터를 전 처리 후
RDB DATA 적재
(문서형식 별 전용 라이브러리 사용)

Vector Embedding 및 Data 적재

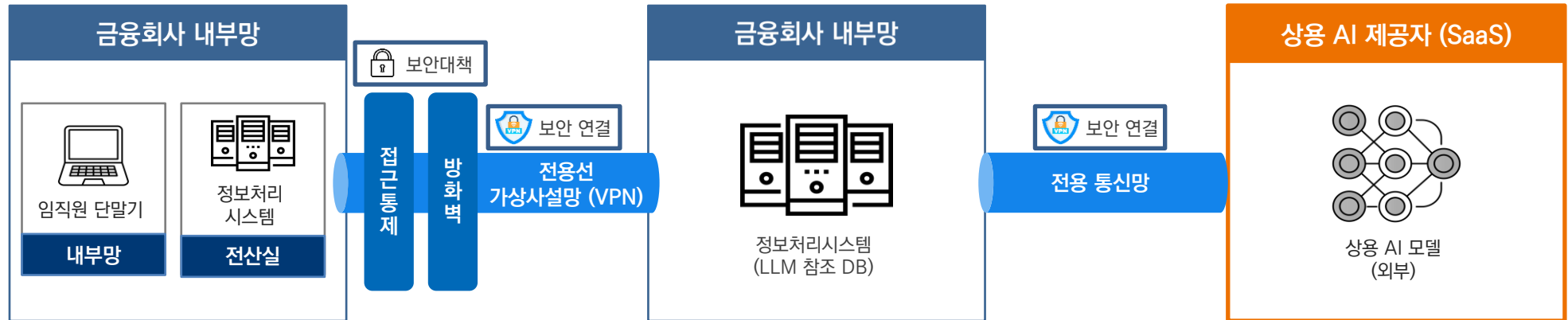


전 처리된 데이터 Vector 및
원본 적재

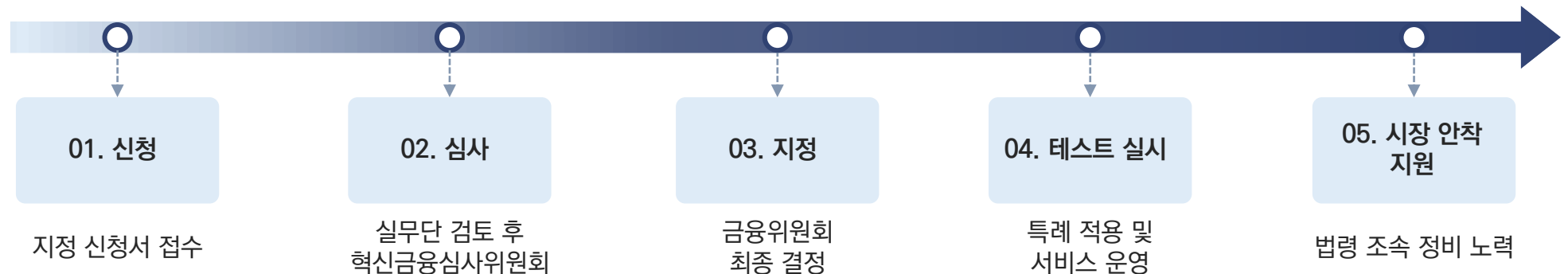
[첨부] 금융규제 샌드박스 - 혁신금융서비스 신청 지원

교보DTS는 금융사에서 새로운 금융서비스 도입을 위한 혁신금융서비스 신청을 지원합니다.

혁신 금융 서비스 신청 대상



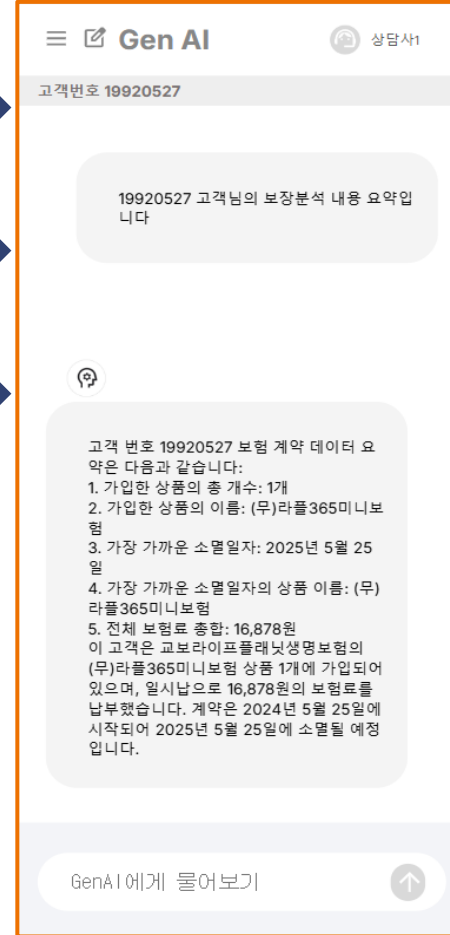
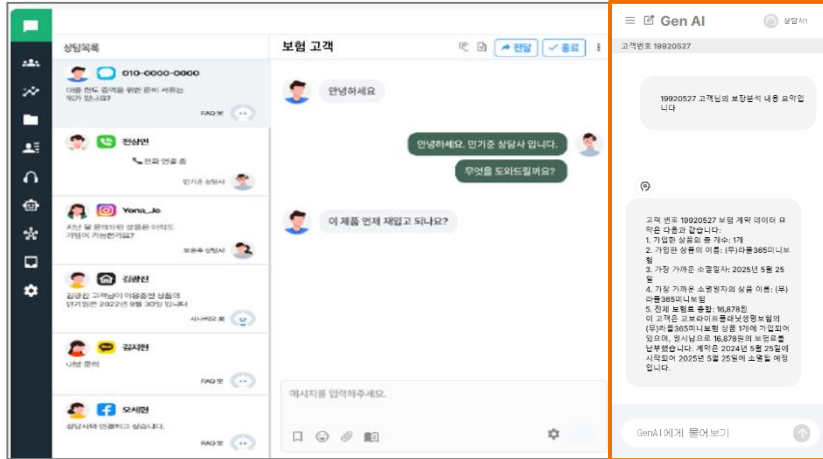
신청 절차 프로세스



[첨부] 구축 사례 - 교보라이프플래닛생명보험 AI Assistant



상담 고객에 대한 보장 분석 등 관련 정보를 신속하게 생성, 제공하여
일관성 있는 고객 상담 품질 및 추가 세일즈 기회 제공



AI 어시스턴트

- 자연어 기반의 상담사 요청에 따른 피드백 제공
- 상담중인 고객을 식별하여 정보 조회
- 상담사 요청내용을 학습된 데이터를 통하여 빠르게 식별 및 제공
- 상담내용(채팅 및 유선) 요약 기능제공

상담사

- 상담 중에 필요한 필요 정보를 간단하게 채팅으로 요청
- AI 답변을 활용하여 고객에게 빠른 응대 수행
- 상담사 표준 응대력 향상

관리자

- 상담 이력 및 상담 내용 추세 확인
- 고객 니즈 및 추이 확인

고객

- 상담사로부터 빠른 피드백을 통한 의사소통

대고객 상담 채팅 시스템
API 연동 및 개발을 통해
신속한 상담 업무 지원

- 기존 상담 내용 요약, 음성 상담
문자 전환(STT)

고객 데이터 활용을 위한
전처리 및 메타데이터화

- 상품별 보장 정보 분석을 위해 고객
데이터 전처리 및 개인정보 등 민감
데이터 제외하여 안전하게 고객
식별 기능 제공

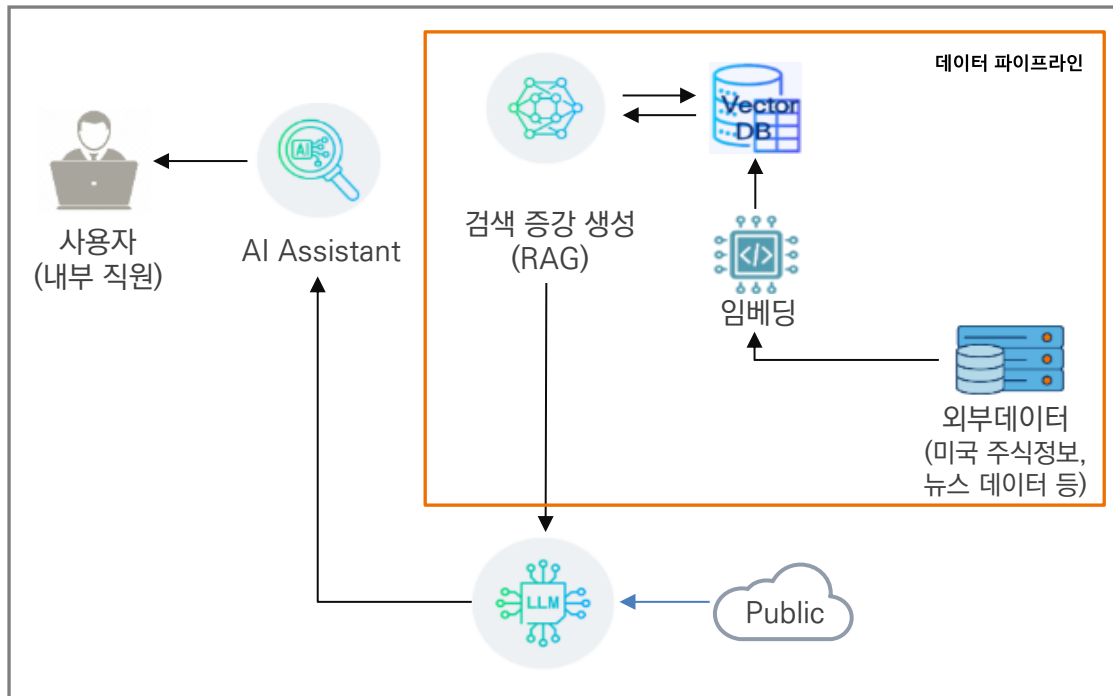
맞춤형 상품 추천 및
보장 내역 제공

- 알고리즘 활용하여 보험 상품 추천
및 상품 관련 세부 정보 및 보장
내역 제공

[첨부] 구축 사례 - 교보증권 해외주식정보 제공 PoC



해외 주식 정보에 대한 빠르고 정확한 제공을 통해
해외 주식거래 고객의 만족도를 증가시키고, 업무 효율화 및 편의성 제공



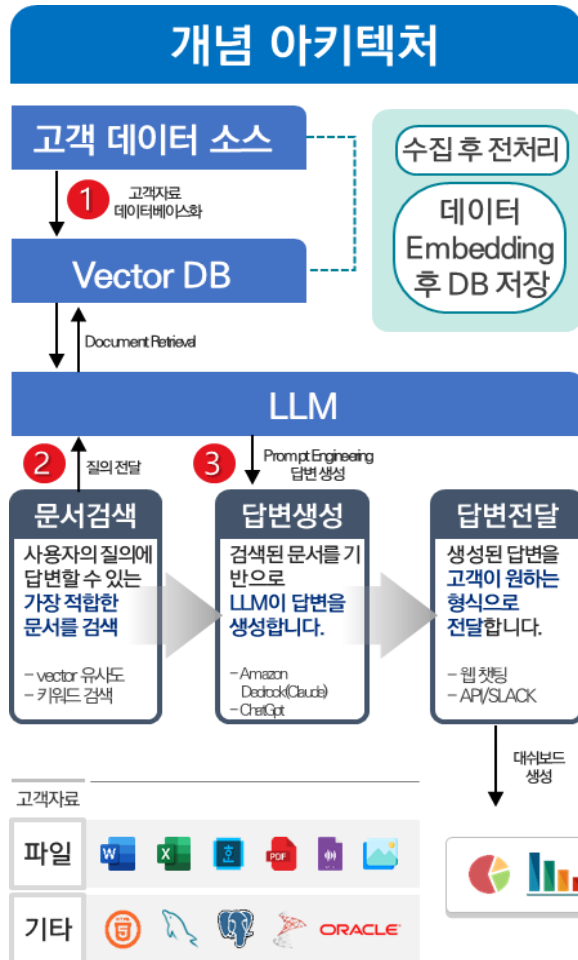
데이터 파이프라인 구축

- 해외주식 어닝콜, 뉴스 데이터 수집 및 RAG화
- SEIBro 순매수 상위종목, S&P500종목

생성형AI 기반 해외주식 정보 제공

- 해외주식 어닝콜 자료를 수집하여 번역 및 핵심내용 요약
- 다수 뉴스 매체의 뉴스를 수집 후 뉴스 내용 요약

RAG(Retrieval-Augmented Generation)는 적절하게 임베딩 된 벡터DB에서 관련 문서를 검색(Retrieval)한 뒤, LLM이 이를 바탕으로 답변을 생성(Generation)하는 방식의 AI 기술입니다.

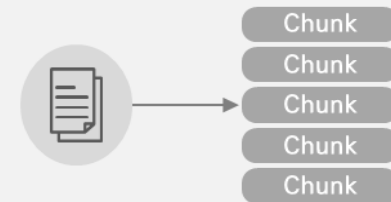


문서 chunk & Linking

내용이 긴 문서 분할

- 적절한 chunk size를 설정
- 데이터의 내용에 따라 기준 설정

Point 단순 텍스트와 표를 구분하여 chunk!



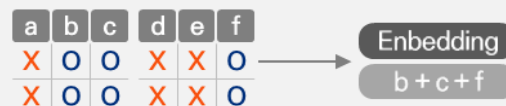
Document Retrieval

데이터 전처리

의미없는 데이터는 삭제

- 무의미하거나 중복된 데이터 삭제
- RAG에 필요한 데이터만 merge

Point 먼저 고객 자료를 분석 후, 데이터 전처리 프로세스를 자동화



프롬프트 엔지니어링

Prompt에 따라 답변 질 결정

- LLM을 효과적으로 제어
- 다양한 LLM마다 프롬프트를 적절하게 변경

Point Context와 질문의 의도를 명확히 전달!

```
<context>
.....
</context>
<function>
.....
</function>
Human:
```

End of Documents