

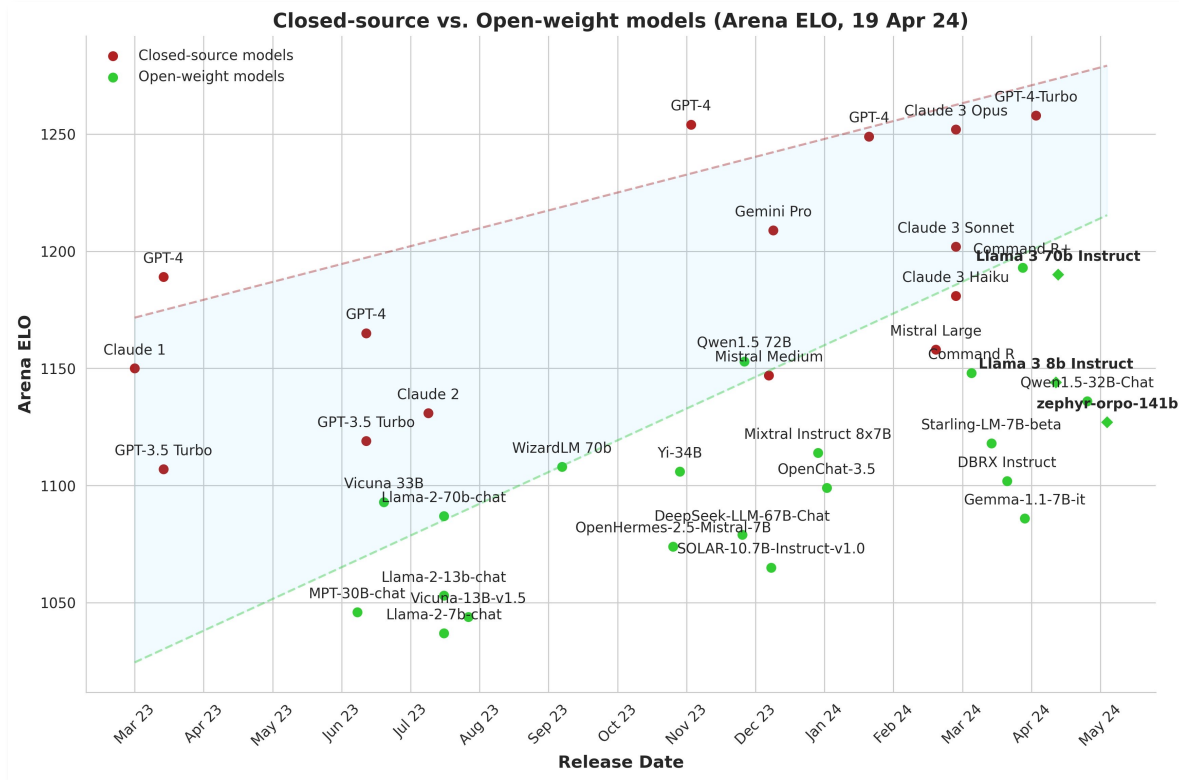
코리아 핀테크 위크 2024
Korea Fintech week 2024

생성형 AI 시대 금융사(보험사) 대응 전략

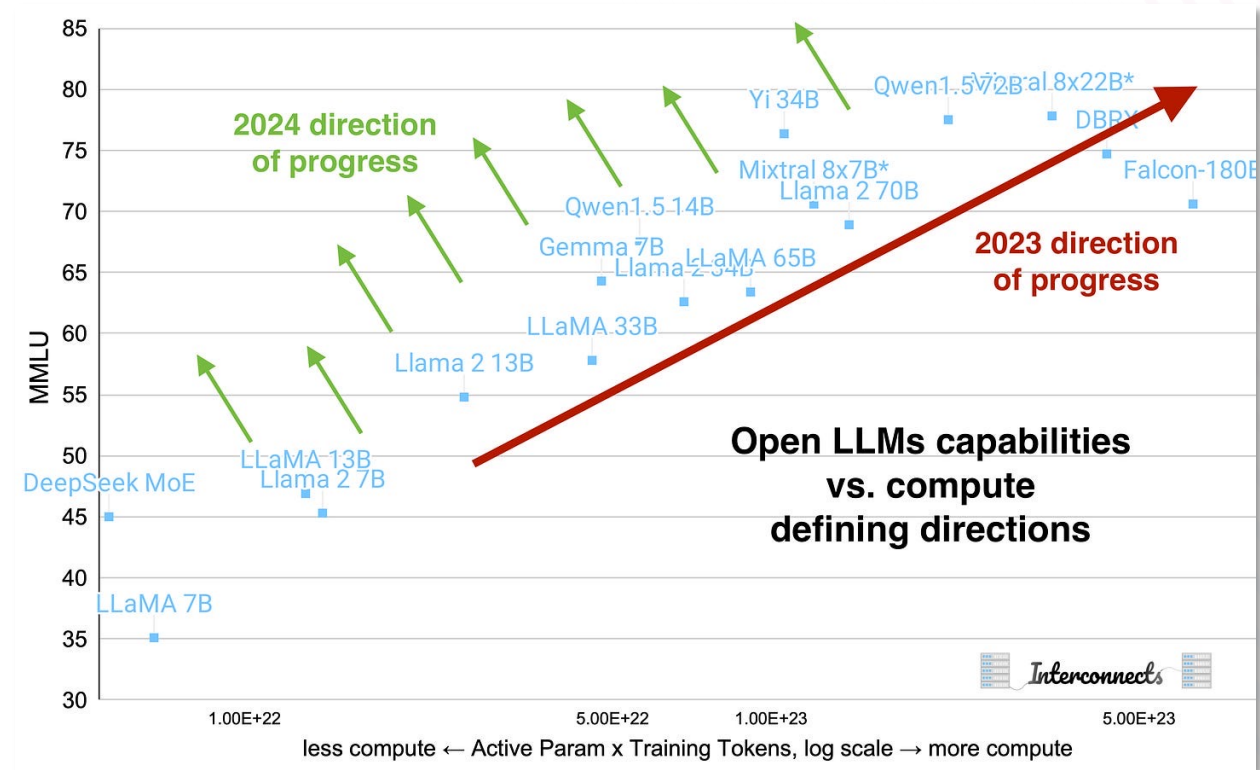
MINDs@company 고석태 대표

Chat GPT 이후 빠르게 진화하고 있는 기술 변화 트렌드

SOTA 모델 출시와 LLM 성능 변화 트렌드



LLM 기술 발전 방향



금융사 생성형 AI 도입 시 주요 고려 사항 및 추진 전략

주요 고려 사항

추진 전략

1

생성형 AI 추진 전략 및 방안

- 단위 기능 개선 중심의 AI 과제 접근으로 AI 혁신의 성공 경험 축적이 미흡
- 대규모 예산 투입 대비 AI 도입을 통한 실질적인 기대효과 불확실성 존재

추진방안 1

- 전사 관점에서 생성형 AI를 도구적으로 활용할 수 있는 혁신 전략 수립
- AI agent 등 금융사에 실질적인 도입효과를 기대할 수 있는 application 개발에 집중

2

금융사 특성을 고려한 생성형 AI Infra 구축

- 금감원 규제, 금융 개인정보 보호 이슈 등 외부 API를 적극적으로 활용할 수 없는 한계
- GPU 서버 등 LLM 학습 및 서빙을 위한 인프라 구축 비용 부담

추진방안 2

- 오픈 소스 생태계의 기술 진화 혜택을 공유
- 외부의 빠른 기술 변화를 효과적으로 활용할 수 있는 유연한 체계 구축
- 다양한 application을 효율적으로 생성하고 관리할 수 있는 인프라 구축(LLM DevOps)

3

금융사 내부 생성형 AI 추진 역량 내재화

- AI 전문가 인력 확보의 어려움
- 비즈니스와 기술을 매개해서 문제를 해결하는 경험 부족
- 외부 협력사에 대한 기술 의존성 우려

추진방안 3

- COE 조직은 AI Engineer 중심의 역량 강화
- 도메인 지식이 있는 현업 담당자들이 생성형 AI를 도구적으로 활용할 수 있는 변화 관리 추진

LLM은 데이터 유형과 AI 과업별로 분산되었던 혁신 기회에 기반 기술 역할을 수행하며 전략적 확산이 가능

LLM 등장 이전 AI Transformation 과제

- NLP 영역 과제들도 데이터, 모델 전문성, 주무 사업부서에 따라 다양한 기술 요소, 제품들을 산발적으로 적용
- 또한 실무 적용을 위한 Application 개발, 일하는 방식 변화 등의 실행 비용이 성공을 가로막는 사례 속출

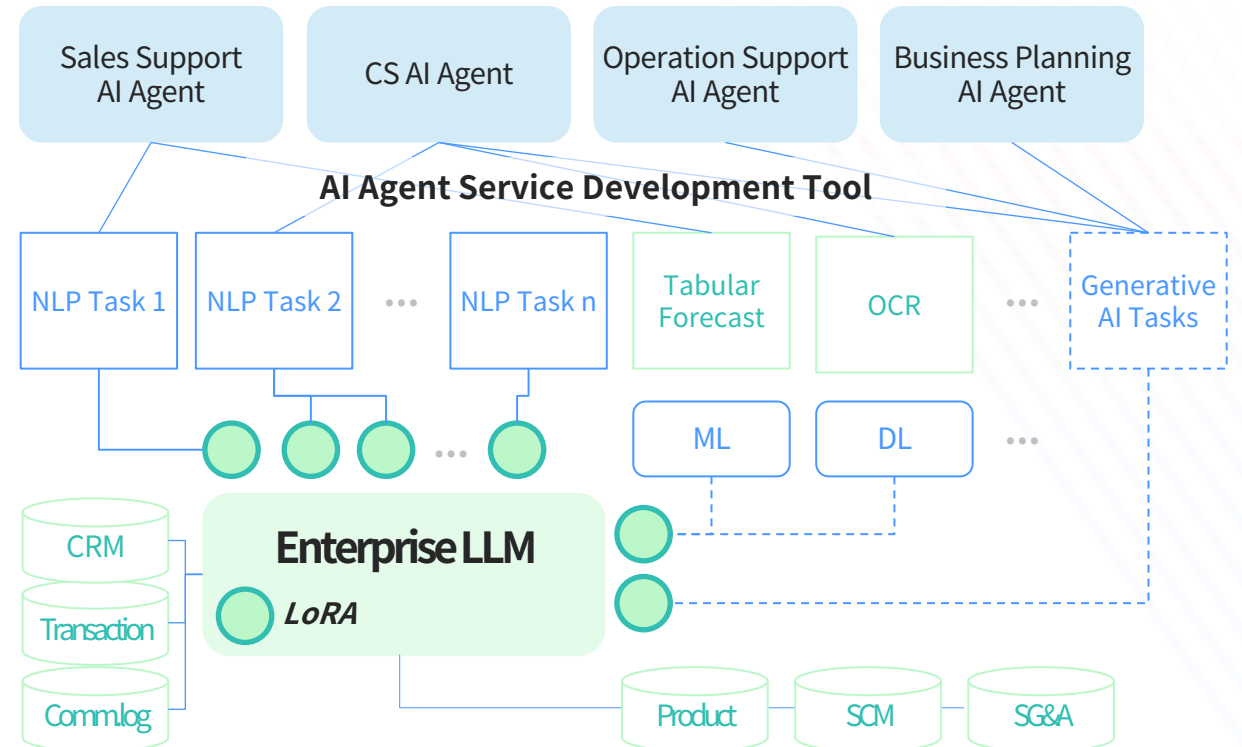
고객사 NLP 진단 사례



LLM 기반 전사 AI 혁신 추진 가능

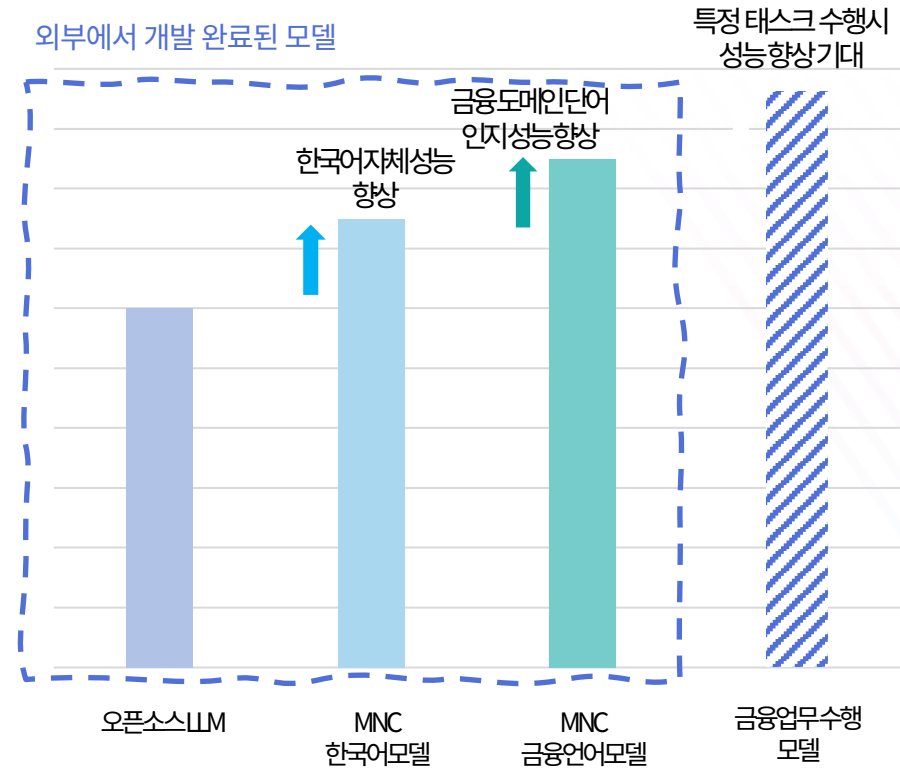
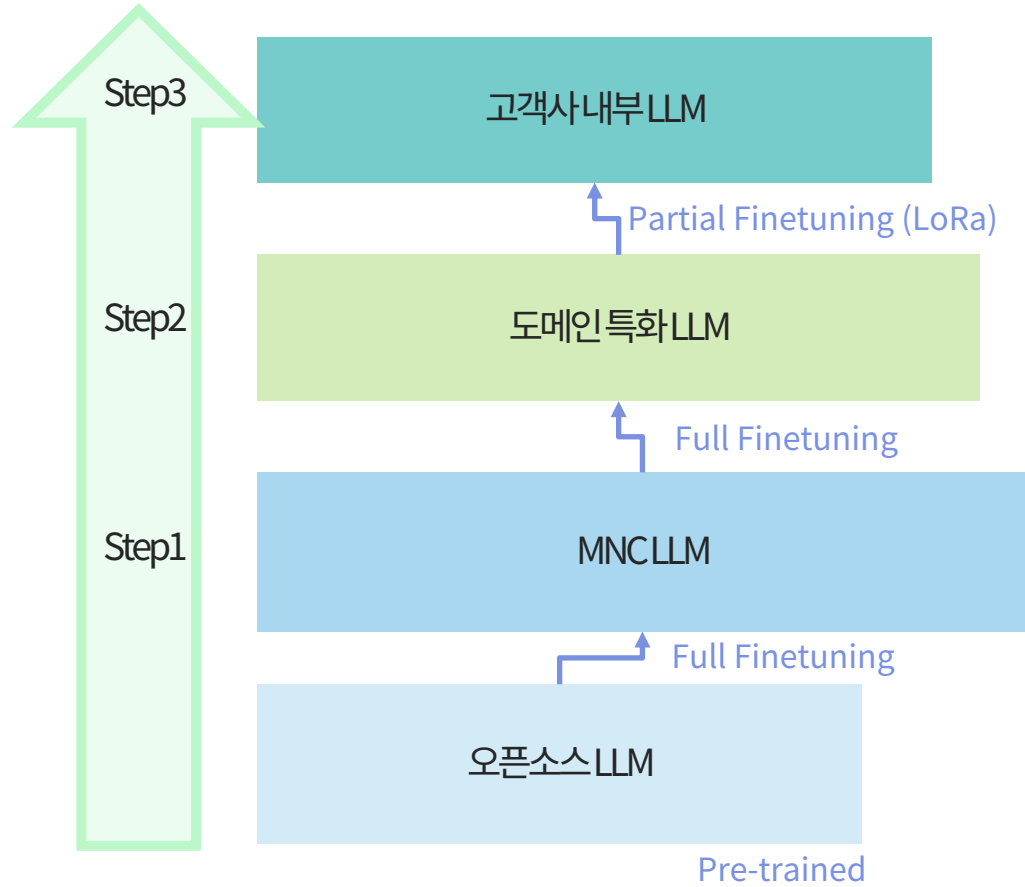


- 당사의 데이터로 최적 학습된 LLM을 기반으로,
- LLM의 다재다능함을 활용하여 다양한 AI Task를 처리하고,
- 이를 사용자들에게 다양한 AI Agent Service로 제공할 수 있음

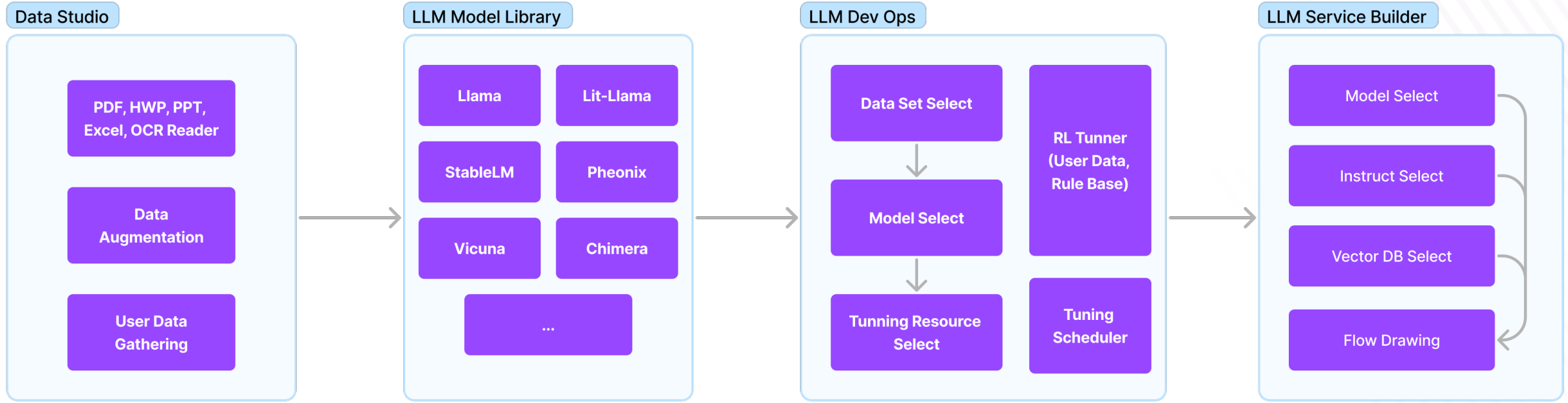


금융사 폐쇄망 환경에서 효과적인 오픈소스 LLM 모델 활용 방안

Llama2기반실험결과추이



LLMops 기반 모델 학습 및 배포 프로세스 효율화



- 효율적인 학습 데이터 구축 지원
- 도메인 전문가 Seed data(500~1,000개) 제작 → LLM 모델을 활용한 데이터 추가 생성

- 신규 SOTA 모델에 대한 빠른 업데이트 제공
- 오픈소스 진영의 빠른 기술 변화에 대응
- 짧은 시간안에 신규 모델 재학습 및 모델 교체 가능

- 학습 데이터 및 기존 파이프라인을 활용해 효율적인 재학습 가능
- Full-finetuning/LoRA 방법을 적용해 모델 성능 최적화
- User Feed Back을 반영한 성능 개선

- 학습 완료된 모델을 특정 Task에 활용하기 위한 Service Flow 구성
- 서비스별 LLM 모델 연계 Pipeline을 생성하고 컨테이너 단위 배포 및 서빙 관리

금융사 생성형 AI 단계적 도입 전략

Phase 1. 생성형 AI 활용 기반 마련

Phase 2. AI Agent 확대

LLM 활용
Application

- 현업 사용자들이 AI application을 쉽게 구성하고 활용할 수 있는 과제 정의(RAG 서비스 프레임 워크)
- 기존 업무 프로세스를 개선하고 업무 담당자를 지원할 수 있는 AI Assistant 과제 정의

역량
내재화

- 전사 생성형 AI 추진 및 운영 전담 조직 구성
 - AI Engineer 중심으로 전사 CoE¹ 조직 운영
- 현업 AI Power User 육성
 - 도메인 날리지를 가지고 있는 현업 사용자의 AI 활용 확대

AI 플랫폼/
인프라

- 모델 및 서비스 배포 효율화를 위한 LLM Ops 인프라 구축
- application을 효율적으로 생성할 수 있는 서비스 관리 도구 도입
- 사내 지식 자산화를 위한 Vector DB 구축
- GPU 서버 클러스터 구축

- 다양한 Task를 완결 할 수 있는 AI Agent 도입
- 사내 정형 데이터, 모델 등 기존 시스템 기능과 연동해서 업무를 수행할 수 있는 Agent Builder 고도화

- AI 전담 조직 확대
 - 다양한 과제를 동시에 추진할 수 있는 매트릭스 조직 형태로 확대
- 현업 사용자 역량 내재화
 - 현업 사용자들이 AI를 도구적으로 활용할 수 있도록 역량 강화

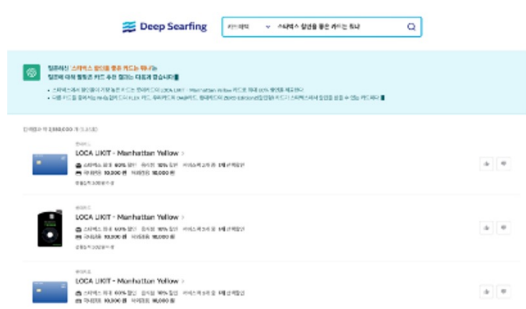
- AI Agent Builder 솔루션 고도화
- 다양한 내외부 Action 연계할 수 있는 API I/F 확대
- GPU 서버 클러스터 자원 확대

금융사 폐쇄망 환경에서 다양한 생성형 AI 상용화 사례를 구축하고 있음

A카드사

의미기반 검색

의미기반 검색 및 검색 결과 요약 솔루션을 통한 카드 상품 혜택 검색



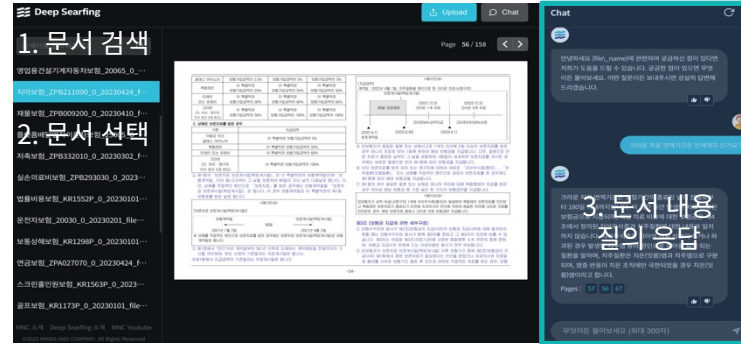
B증권사

C카드사

D은행

Document Question & Answering

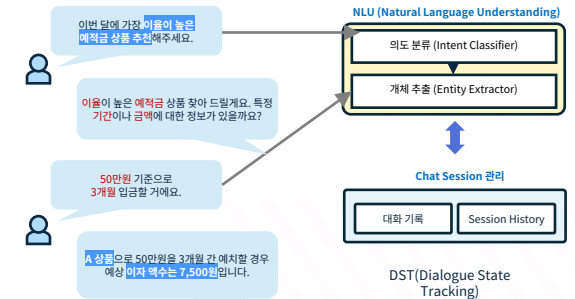
의미기반 검색 및 검색 결과 요약 솔루션을 통한 카드 상품 혜택 검색



D은행

Task Oriented Dialogue

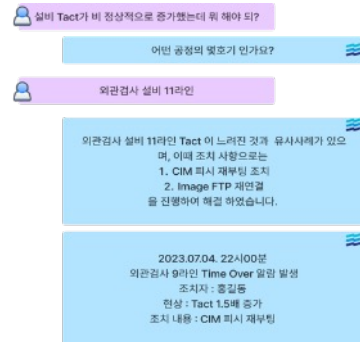
금융 상품 가입 등 복잡한 Task 수행하는 AI Banker (AI 뱅커) 구축



E제조사

Task Oriented Dialogue

고장 원인 추천 대화 모델 구현



- 멀티턴 기반 특정 Task를 완결할 수 있는 대화모델 구현
- 질문의 의도를 파악하고 사내 문서에서 적절한 지식을 검색 후 답변
- 과거 유사한 Case를 검색 후 신규 고장 원인에 대한 적절한 Action을 추천

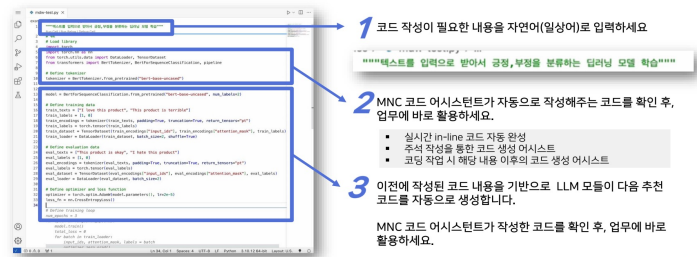
G제조사

F은행

Code Assistant

코드 자동 생성 AI 어시스턴트 솔루션

한국어로 코드 작성이 필요한 내용을 입력만 해주셔요
한 줄의 개발자 주석만으로도 코드를 생성합니다

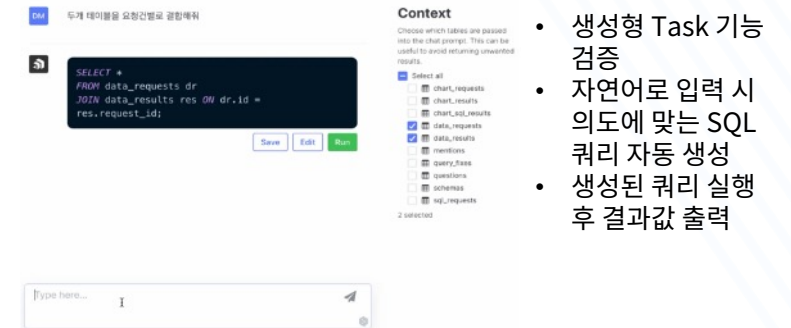


C카드사

E제조사

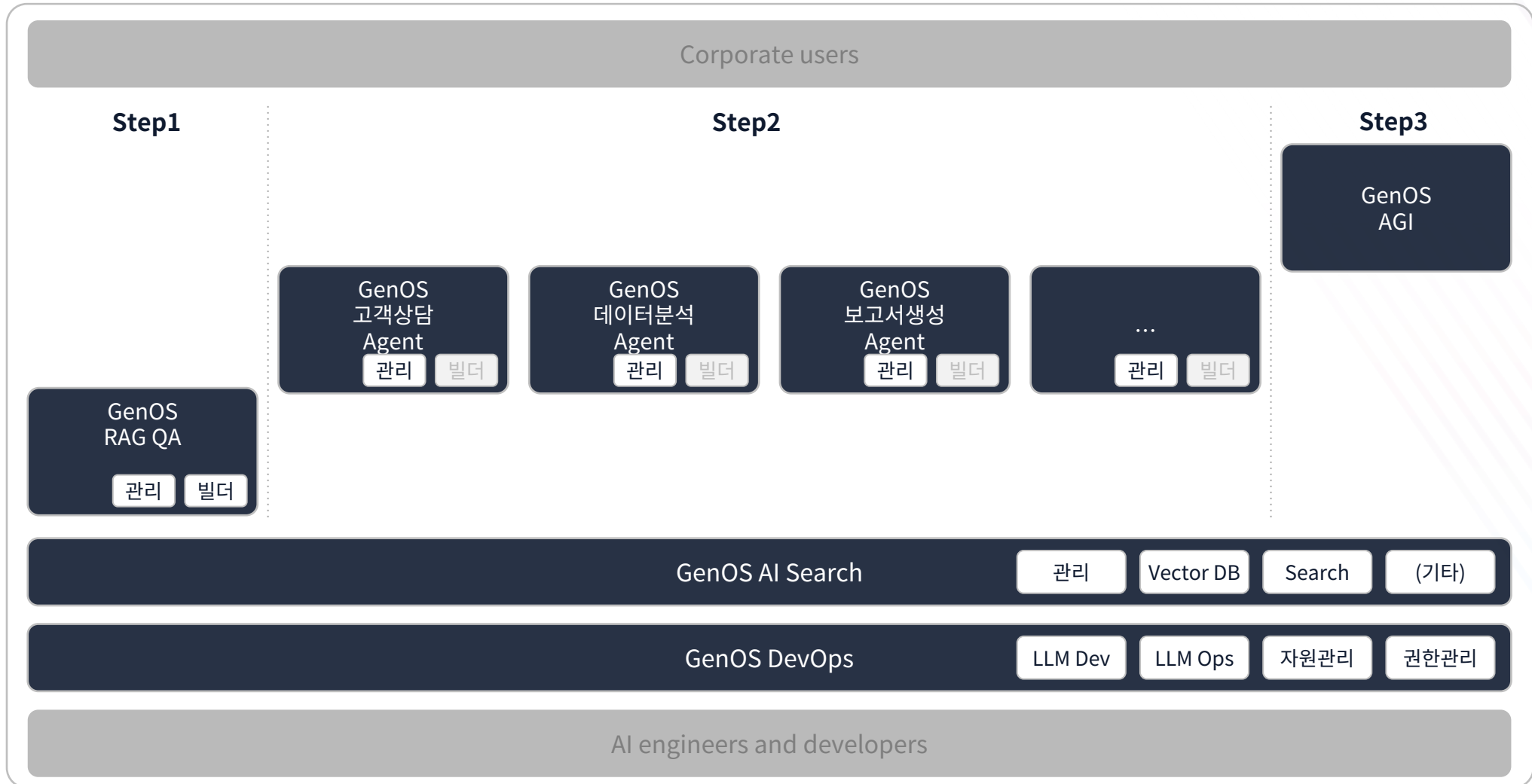
SQL Assistant

SQL 쿼리 자동생성 솔루션

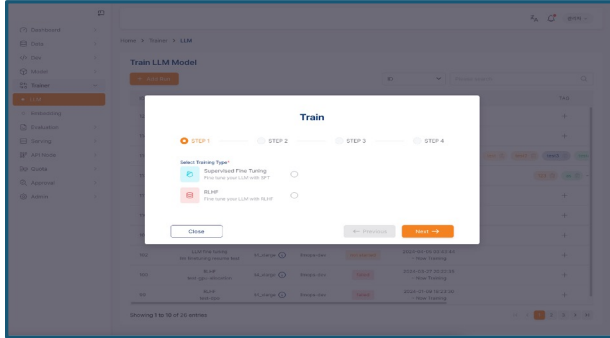


- 생성형 Task 기능 검증
- 자연어로 입력 시 의도에 맞는 SQL 쿼리 자동 생성
- 생성된 쿼리 실행 후 결과값 출력

생성형 AI 플랫폼 “GenOS” 중심의 Tech Roadmap



생성형 AI를 도구적으로 활용하기 위한 플랫폼 주요 기능

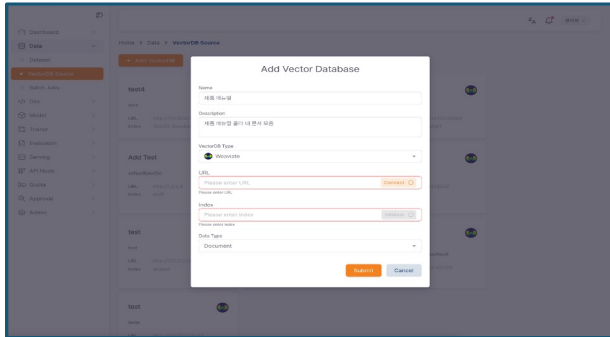


GenOS LLM Ops

- 생성형 AI 서비스를 위한 LLM Ops
- LLM의 학습용 데이터 관리를 위한 데이터 스튜디오 기능
- LLM 학습관리 및 배포, 모니터링 기능



오픈소스 LLM 혁신을
지속 수혈할 수 있는
Future-proof 구조 확보

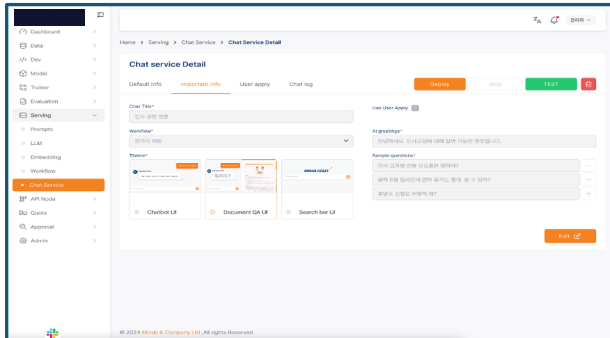


GenOS AI Search (Deep Searfing)

- VectorDB, 전처리 및 임베딩을 포함한 AI Search 프레임워크
- AI Search 프레임워크 기반 의미기반검색 솔루션



지식 문서 변화, 확대에
지속 대응할 수 있는
검색증강생성 제품



GenOS 서비스 빌더

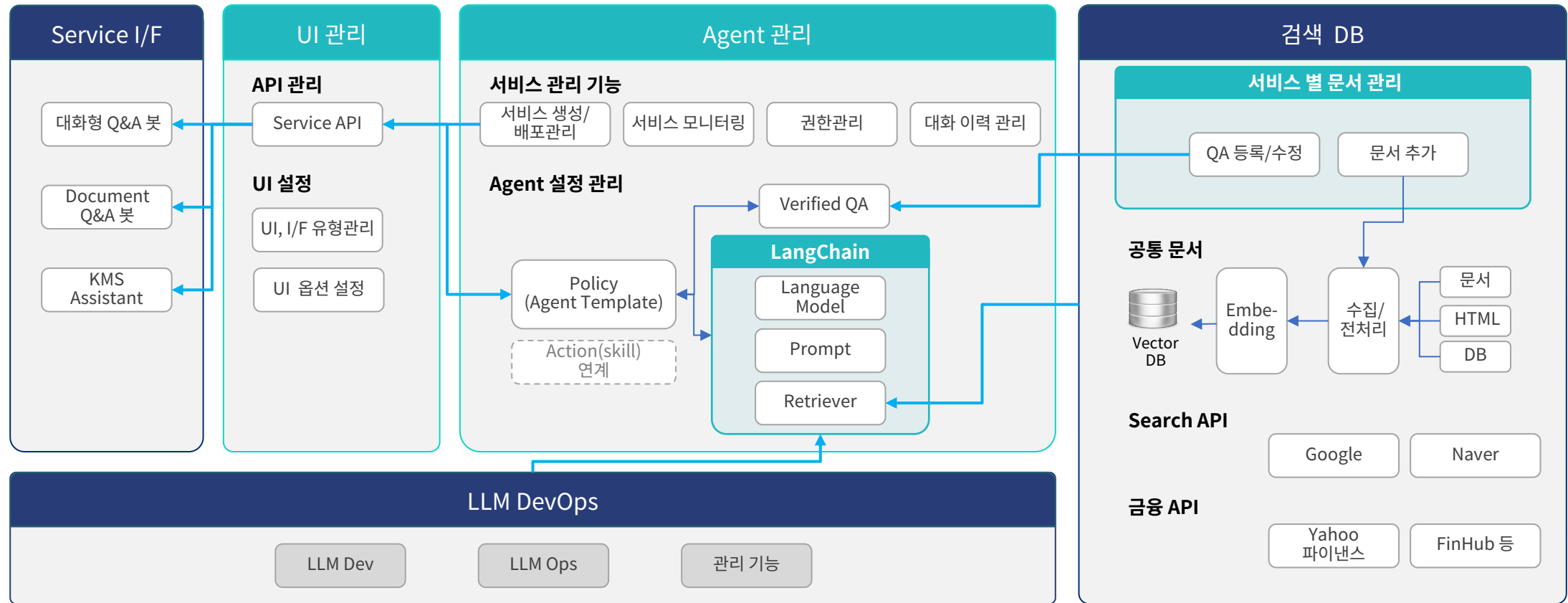
- RAG 기반 문서검색 Assistant를 노코드로 생성·관리
- 최종 서비스 UI, 임베딩 모델 등을 Pre-set에서 손쉽게 선택
- 서비스 배포 및 서비스를 통한 대화관리 기능



빠른 챗봇 서비스 개발과
유지보수를 위한
Low Code 개발도구

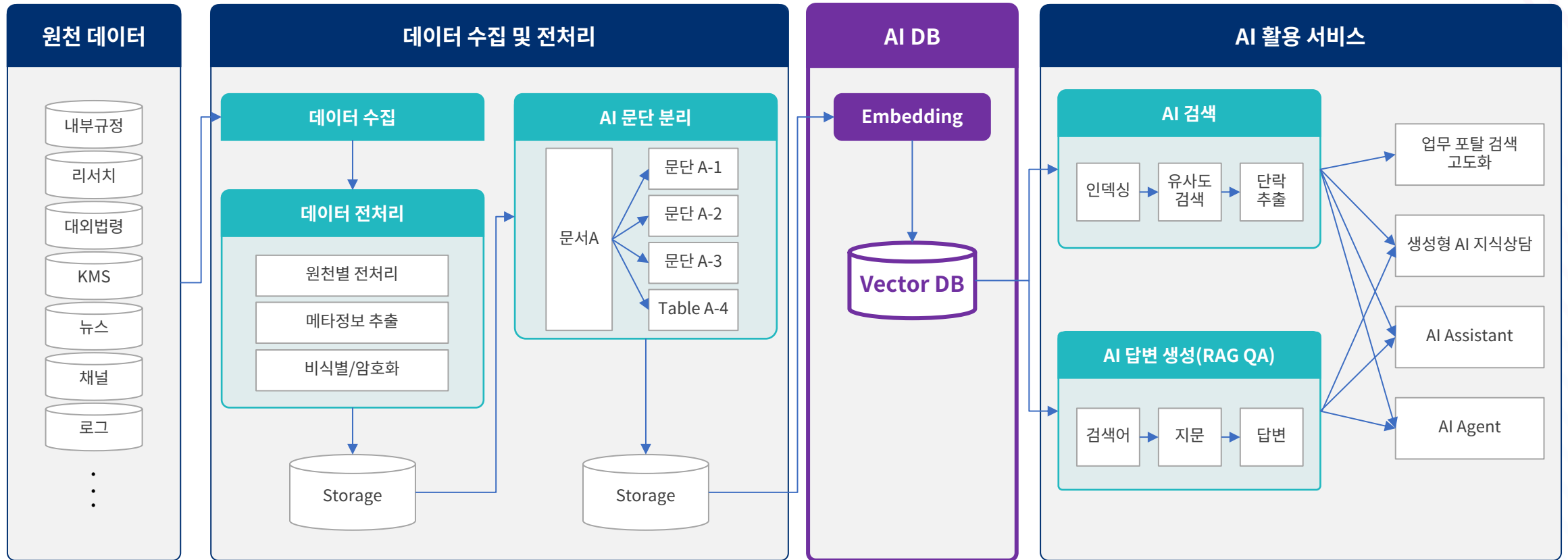
GenOS 서비스 관리 도구 layer - 다양한 서비스를 생성하고 관리 가능한 기능 구현

- 서비스 관리 도구 기능 개요 및 프로세스 구성도



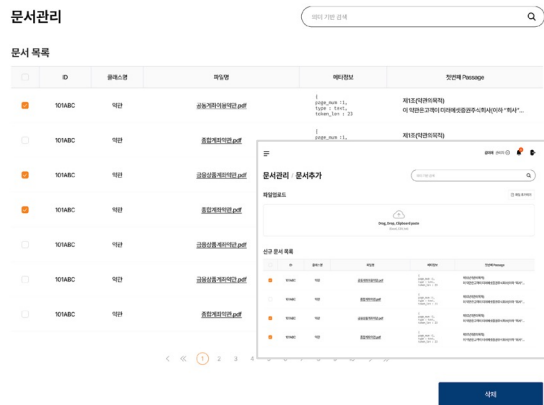
GenOS AI Search layer - 다양한 사내 지식 데이터를 AI가 활용할 수 있도록 벡터 DB로 구축하고 향후 다양한 AI 서비스에 활용

- Vector DB 구축 후 다양한 AI 서비스에 활용



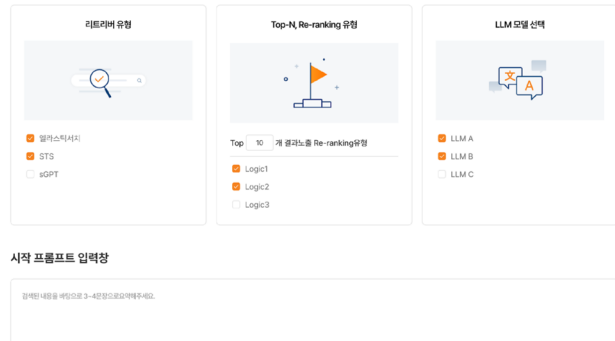
현업 업무 담당자들이 코딩 없이 쉽게 서비스를 생성하고 배포 가능

● RAG 서비스 구성 및 배포 Flow 예시



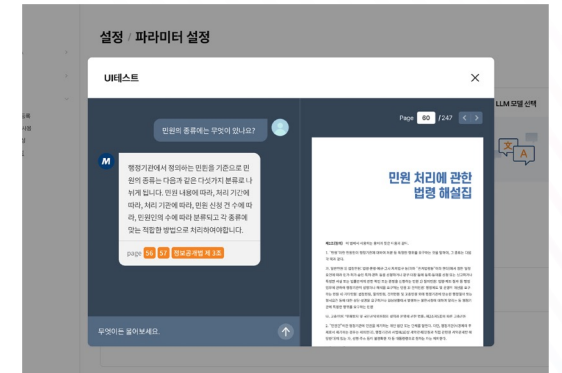
설정 / 파라미터 설정

Document QA UI



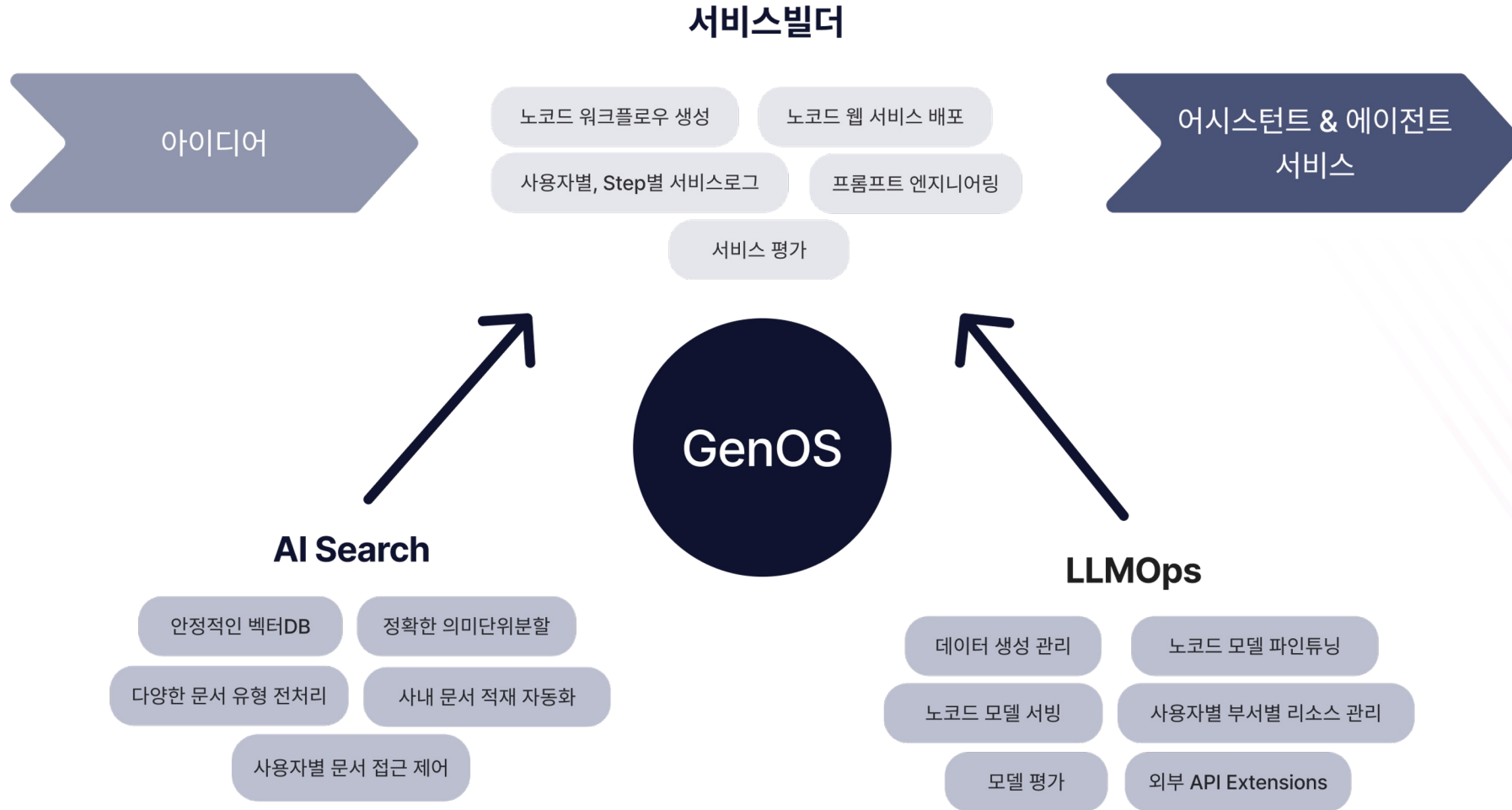
설정 / 사용자 웹 구성

Document QA UI



현업 업무 담당자가 쉽고 빠르게 AI 서비스 구성하고 운영할 수 있는 시스템 지원

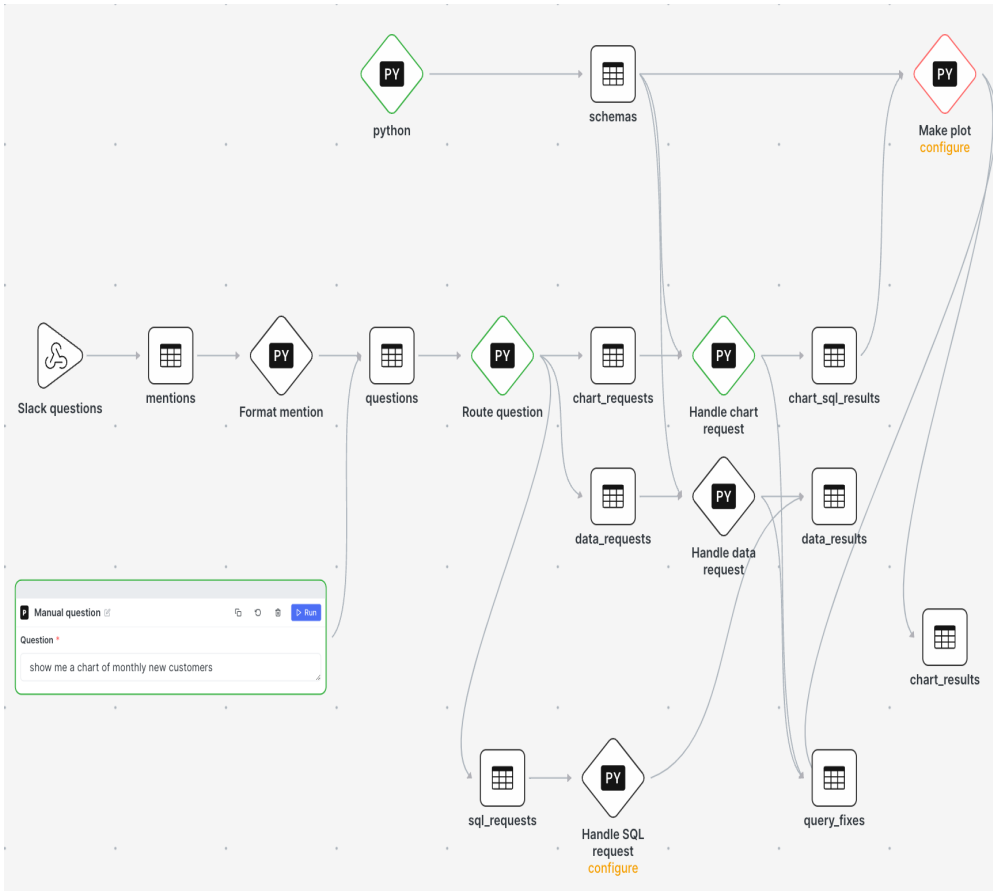
아이디어를 빠르게 서비스로 구현하는 GenOS 서비스빌더



GenOS Features

생성형 AI 플랫폼을 활용해 다양한 Application을 효율적으로 생성 및 관리

AI Agent Service Flow



예시) 자연어로 쿼리, 표, 차트 생성

User Applications

AI Chatbot

- 일반적 질의응답, 문서 요약, 특정 문서에 대한 질의응답 등의 LLM Chatbot
- 맥락을 유지한 연속대화 제공
- Multi-modal 데이터 처리까지 발전 가능

AI Search

- 의미기반 자연어 처리로 문서 검색을 먼저 수행
- 유의미한 결과 제시 및 자연스러운 답변 또는 요약문 제시
- 검색 결과 표시 인터페이스 제공

AI Assistant

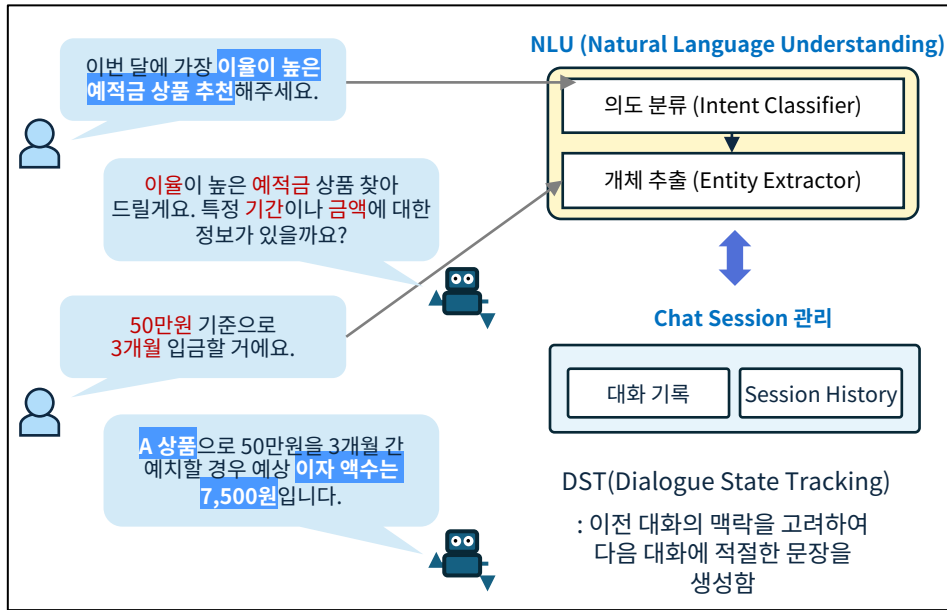
- MS Co-pilot과 같이 사용자의 입력을 실시간으로 처리하여 Task 수행 결과 제공
- 추천 코드, 추천 답변 제시, 자동완성 기능

Implications

- LLM 인터페이스는 다양한 서비스에 수정 없이 적용 가능함
- 기존 AI Model의 현업 적용 시 별도 앱 또는 애플리케이션을 개발하거나, 기존 시스템 화면 개발에 드는 공수 절약
- 사용하는 유저들의 변화관리 장벽을 현저히 낮출 수 있음

금융상품 상담 AI Agent - 우리은행 AI Banker 구축 사례

Hallucination을 최소화하고 목표 과업을 수행하는데 집중한 LLM AI Banker



우리나라 4대 은행인 D은행은 마인즈앤킴퍼니와 함께 ‘국내 금융권 최초’로 생성AI를 활용한 고객 금융상담 서비스의 구축을 시작했습니다. ‘AI Banker Agent’는 은행원의 대고객 예·적금 상품 상담업무를 대체할 수 있도록 해줄 것이며, 이를 위해서는 고객이 요구하는 Task의 맥락에서 벗어나지 않으면서 자연스러운 대화를 이어갈 수 있는 기술이 필요합니다.

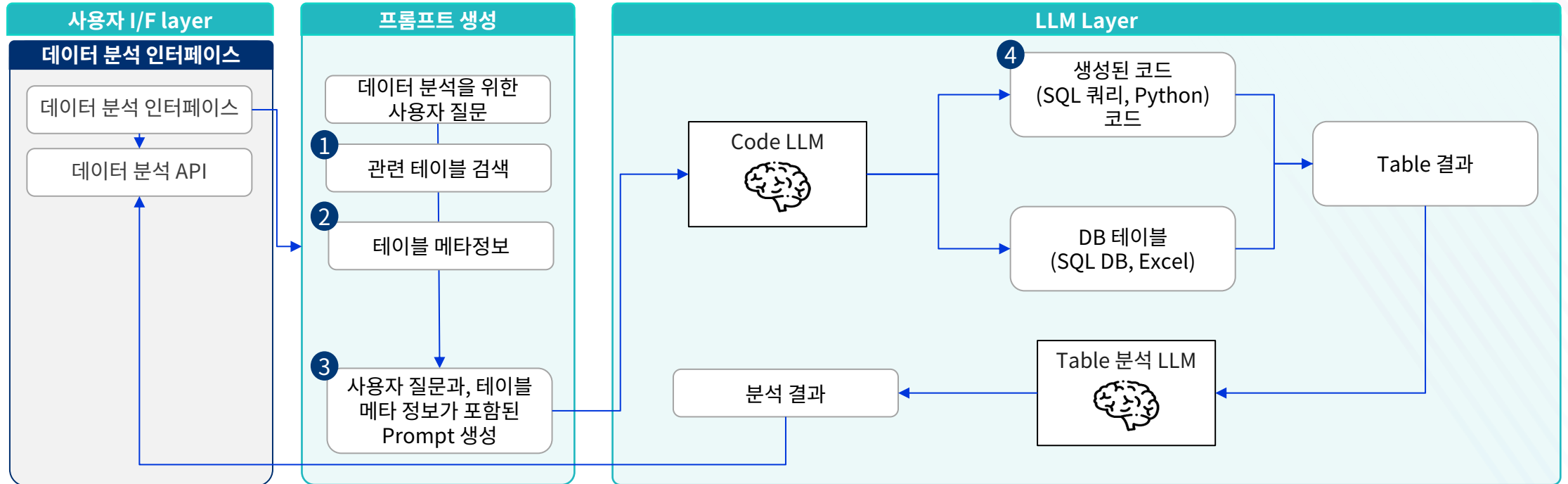
기업을 위한 마인즈앤킴퍼니의 솔루션 제안

고객이 요구하는 Task의 맥락에서 벗어나지 않고 자연스러운 대화를 이어갈 수 있도록 TOD¹⁾ 및 DST²⁾ 등의 기술 적용한 AI Banker Agent 고객이 말하는 의도를 정확히 파악할 수 있도록, 다양하게 표현되는 단위(예: ‘만 원’, ‘10,000원’)를 동일하게 인식하는 Entity 추출 활용

- LLM챗봇
- TOD
- Multi-turn
- DST²⁾

데이터분석 Agent 시스템 아키텍처 구성도

- 생성형 AI 모델을 기반으로 데이터 분석 기능이 구현이 가능하도록 시스템 아키텍처 수립



1 테이블 메타정보

- Code LLM이 사용자의 자연어 질의를 이해하고 테이블에서 정확한 정보를 추출하는 쿼리를 생성하기 위한 메타 정보 작성

2 관련 테이블 검색

- 테이블 검색용 벡터 DB에서 사용자 쿼리와 관련된 테이블 검색

3 프롬프트 생성

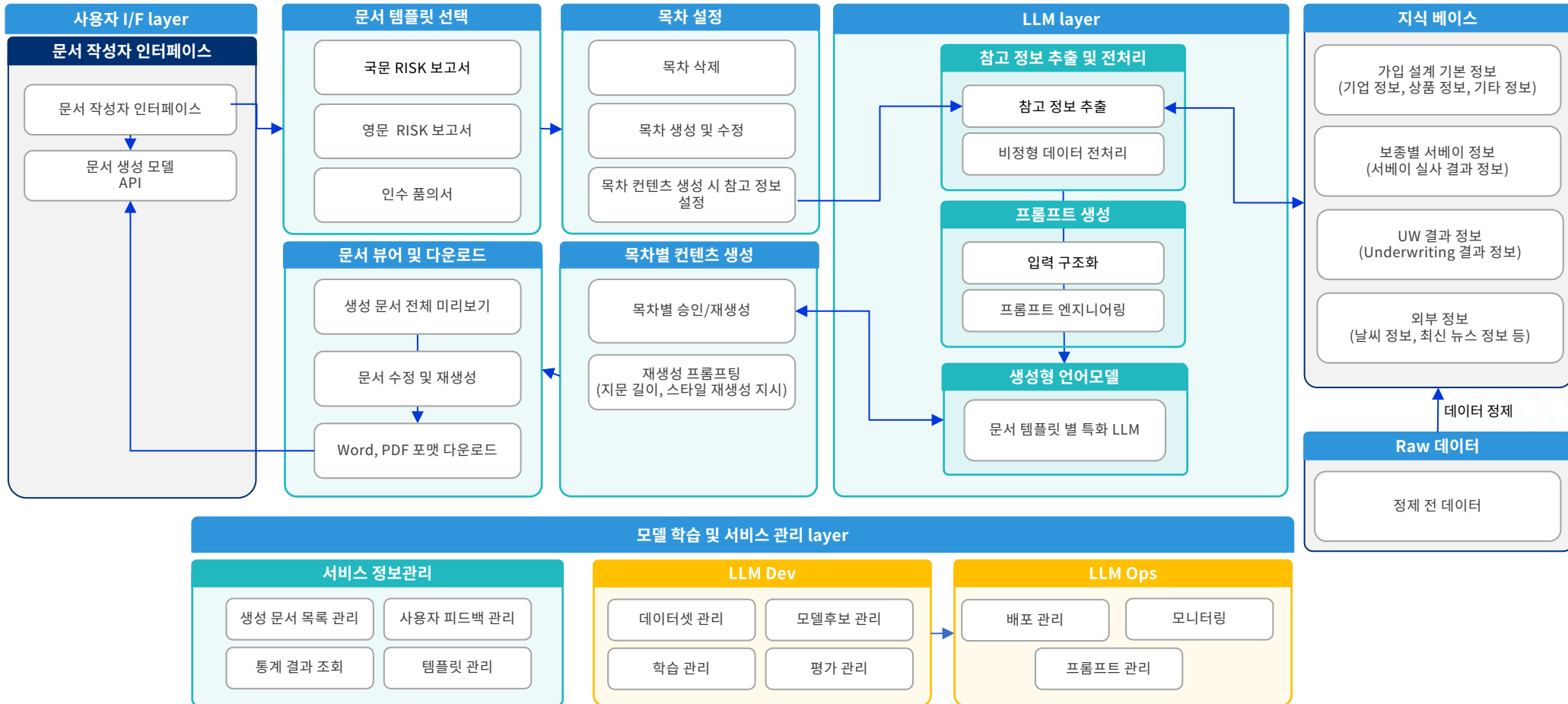
- Prompt Engineering을 통해 다양한 사용자의 질문에 Code LLM이 정확한 코드를 생성할 수 있는 prompt를 분석

4 데이터 분석 코드

- SQL 쿼리를 활용해 DB에서 관련 데이터를 불러옴
- Python 코드로 불러온 데이터를 처리함

보고서 생성 Agent 시스템 아키텍처 구성도

- 생성형 AI 모델을 기반으로 데이터 분석 기능이 구현이 가능하도록 시스템 아키텍처 수립



감사합니다

MINDs@company