

Leap, Leap ahead

Konan LLM

제품소개서



CONTENTS

01	시장 동향	-----	p.03
02	제품 개요	-----	p.07
03	제품 상세	-----	p.11
04	특장점	-----	p.15
05	도입/운영	-----	p.25



01

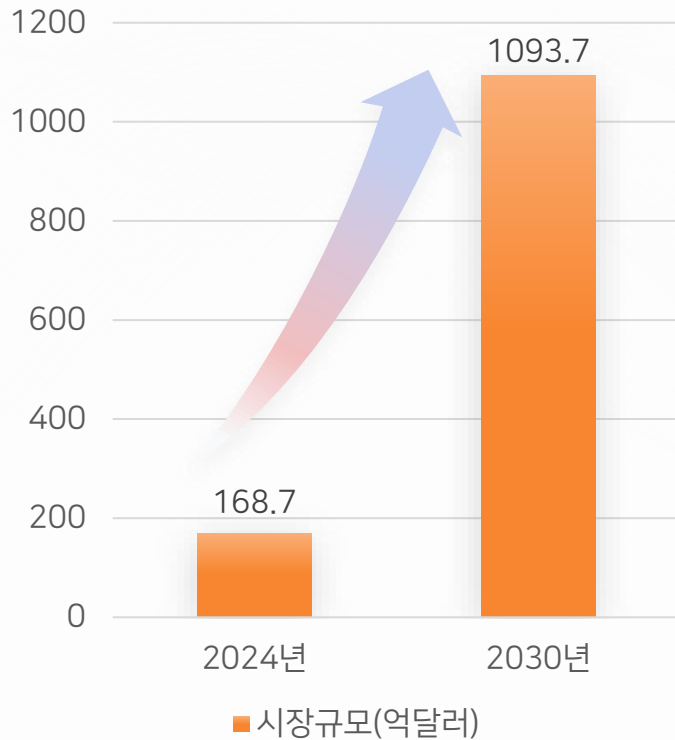
시장 동향

1. 시장 전망
2. 고객의 기대
3. 고객의 고민

생성형 AI 시장 전망

생성형 AI 기술은 기존 데이터에서 패턴을 학습하여 새로운 결과를 생성하는 기술입니다.
전 세계 생성형 AI 시장규모는 2030년까지 연평균 30% 이상 성장률을 기록하며 빠르게 성장할 것으로 전망됩니다.

성장률 연 평균 37.6% 성장 전망



자료: 그랜드뷰 리서치

● 딥러닝 기술의 발전

- 딥러닝, 강화학습 등의 알고리즘의 발전으로 인해 생성형 AI 모델의 성능이 크게 향상

● 하드웨어의 발전

- GPU(그래픽 처리 장치)의 발전은 대규모 딥러닝 모델의 학습을 가능하게 함.

● 데이터의 증가

- 인터넷과 모바일의 보급으로 인해 데이터의 양이 급증

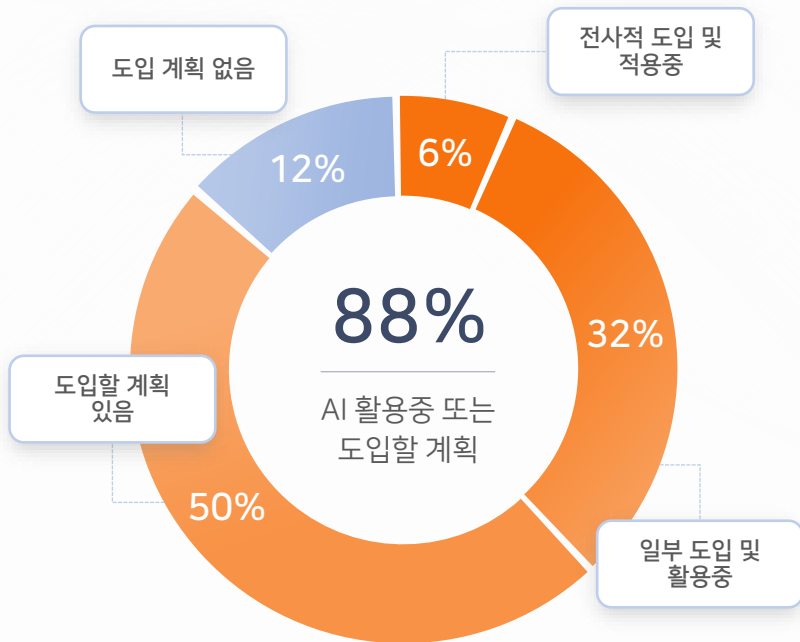
● 연구개발 투자의 증가

- 기업과 정부의 연구 개발 투자가 증가함에 따라 생성형 AI 기술의 발전이 가속화

고객의 기대

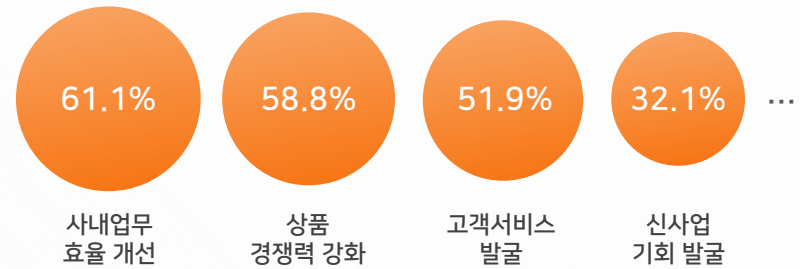
국내 기업 10곳 중 9곳은 'AI기술을 도입하였거나 도입할 계획이 있다.'고 조사되었고, 기업 및 기관은 '사내 업무 효율 개선', '상품경쟁력 강화', '고객서비스 강화'를 기대하고 있습니다.

Q. AI기술을 도입했거나 도입할 계획이 있다.

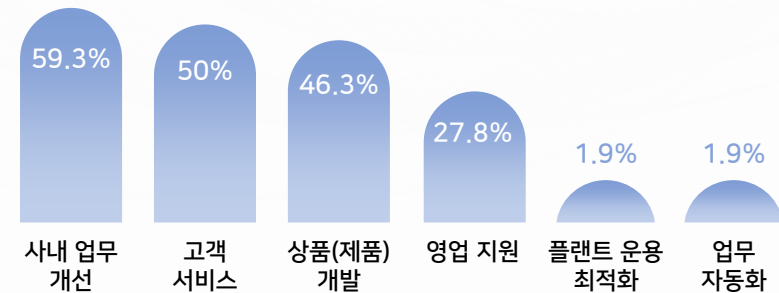


자료: EY한영, CEO스코어

Q. AI 업무 적용 후, 기대 효과는?



Q. AI 신규 도입해 적용할 업무분야는?



고객의 고민

고객들은 생성형 AI 도입을 통해 업무혁신을 기대하면서 동시에 도입과정에서 기술/보안/비용 등 다양한 고민거리를 안고 있습니다.



기술 측면

- 데이터 확보
- 모델의 신뢰성
- 모델의 문맥 이해력



보안 측면

- 기업 민감정보 유출 우려
- 고객 개인정보 유출 우려



비용 측면

- 인프라 비용
- LLM 비용



02

제품 개요

1. 제품 개요
2. 아키텍처
3. 시스템 구성도

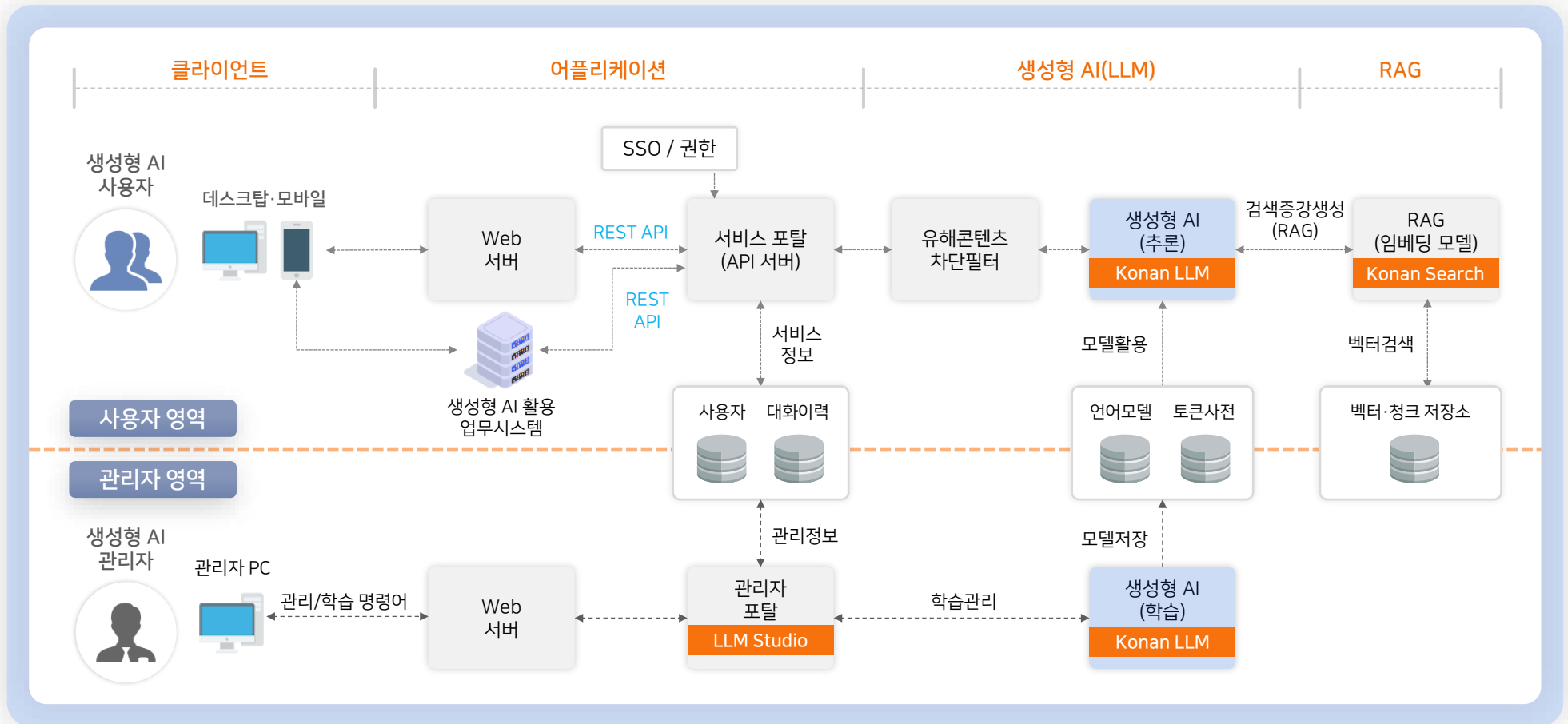
생성형 AI, 코난 LLM

코난LLM은 코난테크놀로지의 자연어처리 전문 AI연구소에서 개발한 생성형 AI입니다.
공공기관과 기업의 업무 효율성을 획기적으로 높여, 업무혁신을 가속화합니다.



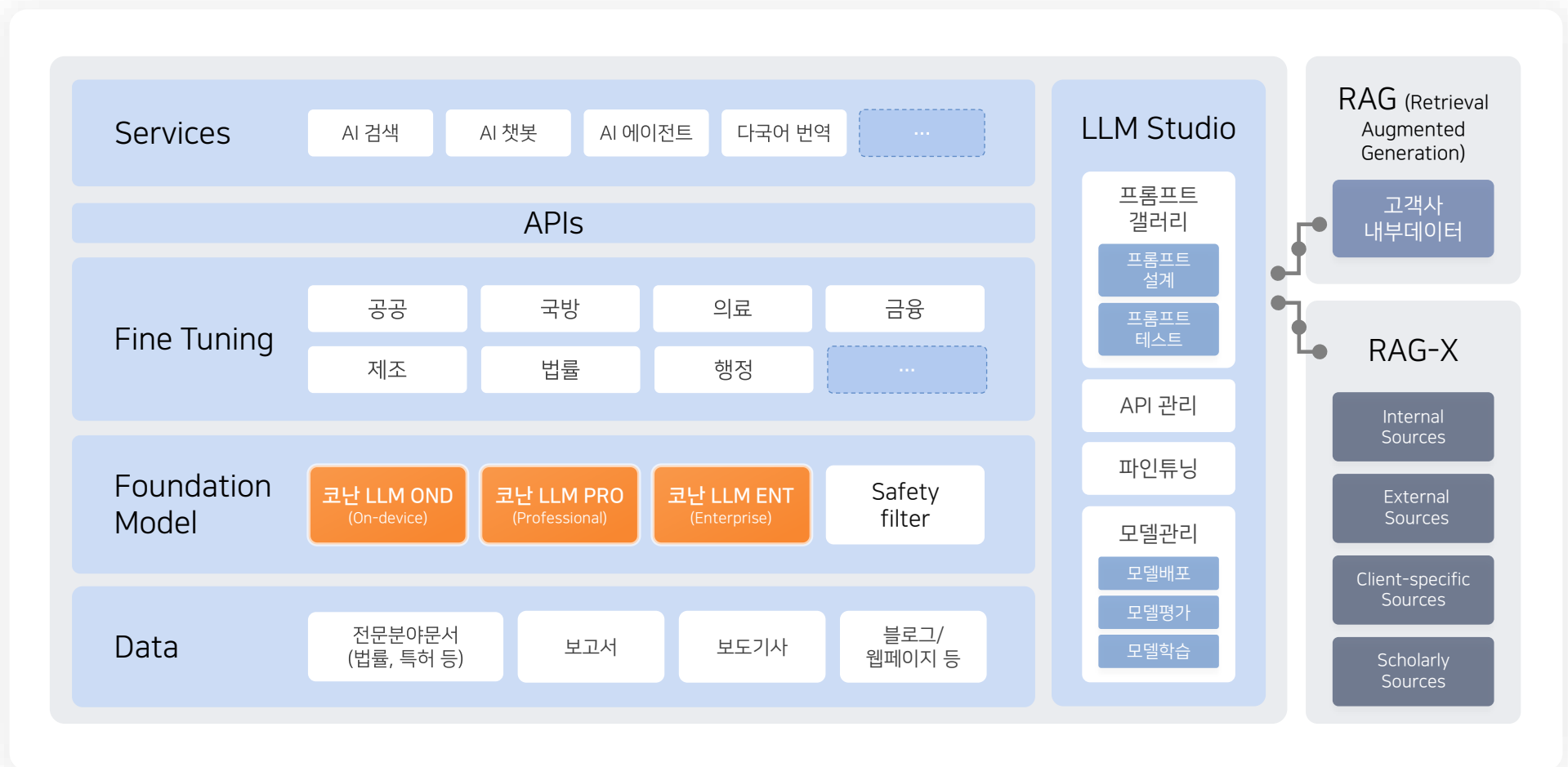
코난 LLM 아키텍처

코난 LLM은 사용자가 애플리케이션을 통해 질문을 입력하면, RAG 기술을 활용하여 관련 정보를 검색하고 답변을 생성합니다. 이 과정을 통해 코난 LLM은 최신 정보와 사용자 맞춤형 답변을 제공할 수 있습니다.



코난 LLM 시스템 구성도

코난 LLM은 공공, 국방, 금융, 의료 등 다양한 도메인에서 활용 가능한 생성형 AI 솔루션입니다. AI 검색, 챗봇, 번역 기능을 포함해 폭넓은 기능을 제공하며, 최신 RAG 기술을 결합하여 신뢰도 높은 문서 및 답변을 생성합니다.





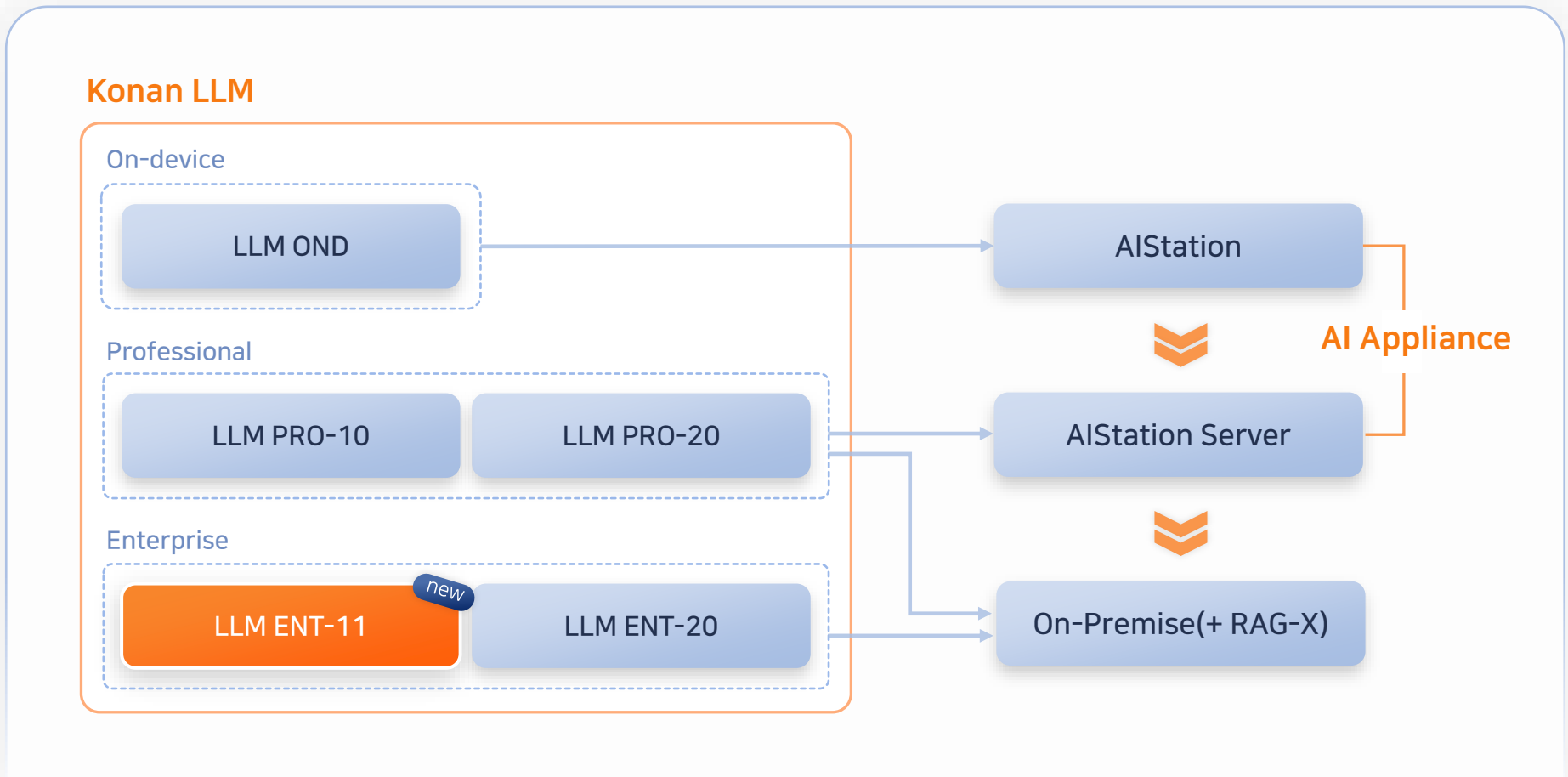
03

제품 상세

1. 제품 라인업
2. 코난 LLM ENT-11 벤치마크 성능평가
3. 코난 LLM 추론모델의 강점

제품 라인업

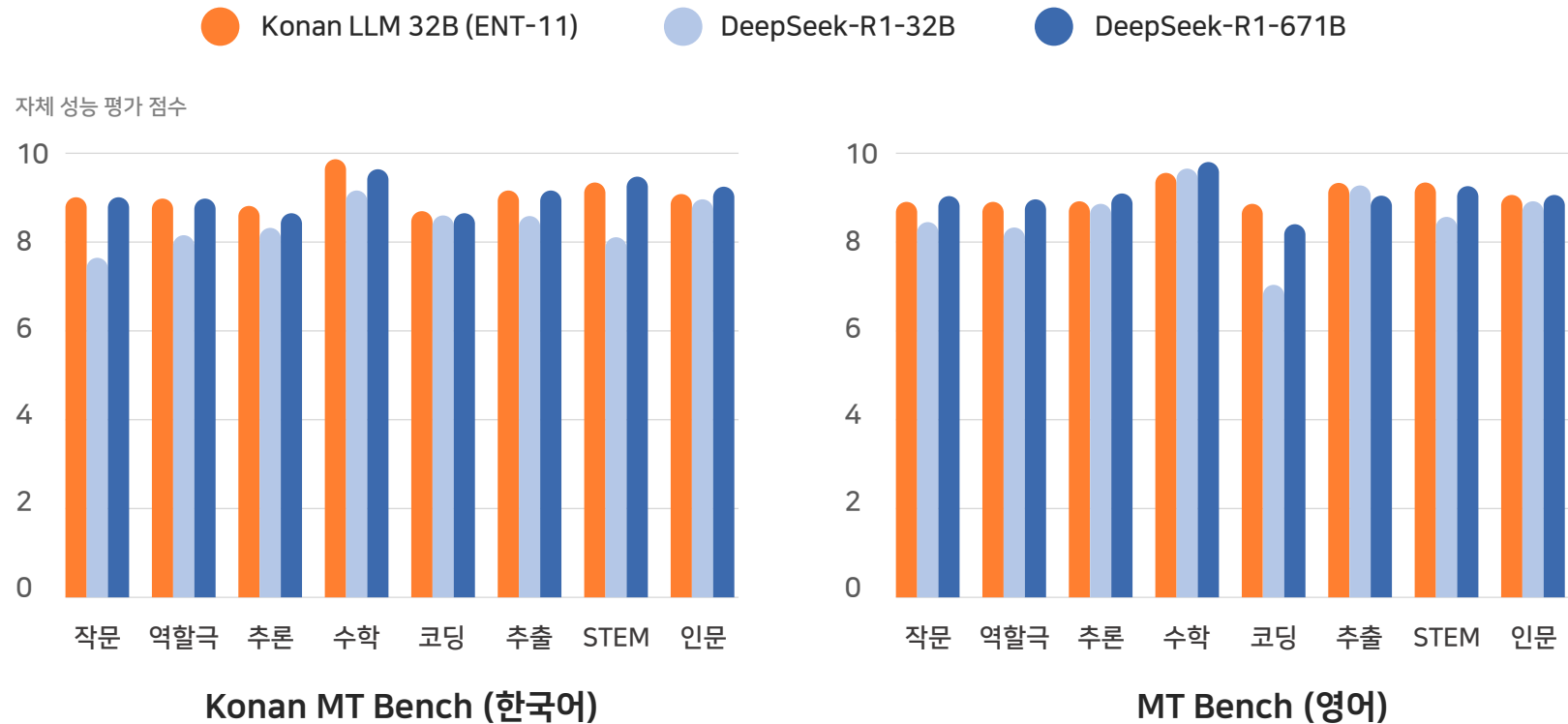
코난 LLM 기반의 다양한 모델(OND, PRO, ENT)을 바탕으로
AI Appliance(AIStation/AIStation Server)부터 On-Premise(RAG-X)까지 고객 환경에 최적화된 솔루션을 제공합니다.



코난 LLM ENT-11 벤치마크 성능평가

일반·추론 모드를 통합한 코난 LLM ENT-11 모델은 자체 평가한 벤치마크에서 DeepSeek-R1 모델 대비 작문, 역할극, 추론, 코딩 등 8개 항목에서 우위를 보였습니다.

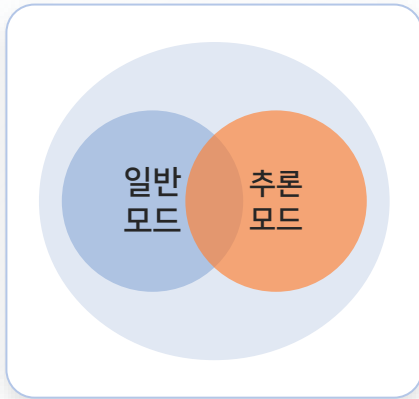
성능 평가 결과



코난 LLM 추론모델의 강점

25년 차 NLP 기술력과 고품질 한국어 DB를 기반으로 한 국내 최초의 **일반·추론 통합모델**입니다.
더 적은 GPU 비용으로 고성능 AI 서비스가 가능하도록 통합모드를 지원합니다.

Konan LLM ENT-11



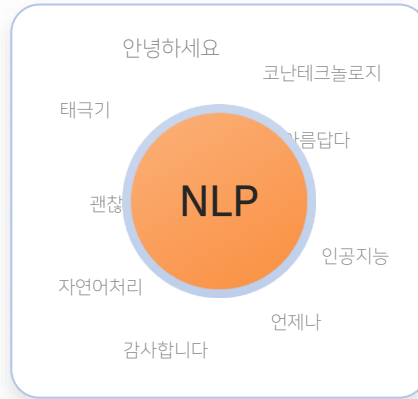
통합모드 지원

일반모드와 추론모드를 하나의 모델로 통합 지원하여, 더 적은 GPU 비용으로 고성능 AI 서비스가 가능



고효율 모델 아키텍처

답시크 R1 대비 1/20 수준의 모델 크기로도 동등하거나 우수한 성능 제공



한국어 특화 모델

고품질 한국어 학습 데이터를 활용한 생성 모델로 기업/공공기관 맞춤형 문서 생성에 특화



향상된 성능

이전 ENT-10 모델 대비 4.5%p 향상된 성능으로 더욱 정교한 답변을 제공



04

특장점

1. 선별된 고품질 데이터로 성능 기반 확보
2. 최적의 고객 맞춤형 모델 제공
3. 정보유출 우려없이 안전하게
4. 정확한 소스에서 찾는 최적의 답변
5. 외부 서비스를 연결하는 간편한 확장 방법 MCP
6. 복잡한 업무를 자율적으로 수행하는 Agentic AI
7. 생산성을 높이는 코난 AI Agent
8. 권한에 맞게 문서열람 제한
9. 손쉬운 모델관리와 성능평가

선별된 고품질 데이터로 성능 기반 확보

법률/특허와 같은 전문 분야 문서, 보고서, 뉴스 등 신뢰도 높은 데이터를 선별하고 엄격한 필터링으로 저품질 데이터를 제거합니다. 확보된 고품질 데이터는 **코난 LLM**의 우수한 성능을 위한 핵심 기반이 됩니다.

답변 생성 결과 비교 (코난 LLM ENT-11 vs DeepSeek 32B)

[예시] 수능 국어 문제

지문:
법령의 조문은 대개 'A에 해당하면 B를 해야 한다.'처럼 요건과효과로 구성된 조건문으로 규정된다. ...**(중략)**... 행정청은 평등 원칙을 ◎**(지켜야)** 하기 때문이다.

질문:
윗글을 바탕으로 <보기>를 이해한 내용으로 가장 적절한 것은? [3점]

<보기>:
<보기> 갑은 을에게 물건을 팔고 그 대가로 100을 받기로 하는 매매 계약을 했다.
...**(중략)**... 단, 위의 모든 상황에서 세금, 이자 및 기타 비용은 고려하지 않음.

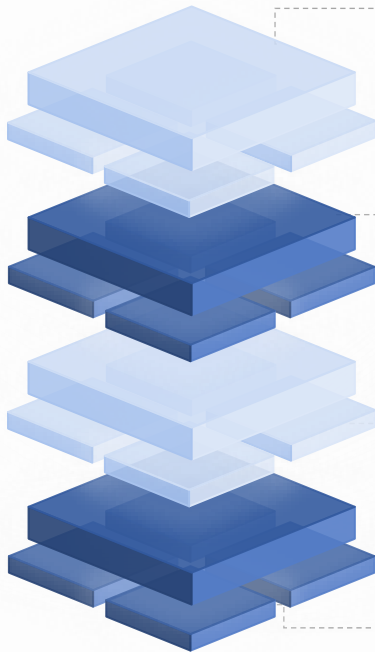
선택지:
1번 - (가)에서 을의 손해가 얼마인지 증명되지 못한 경우에도, 갑이 을에게 80을 지급해야 하고 법원이 감액할 수 없다.
...**(중략)**... 5번 - (다)에서 을의 손해가 얼마인지 증명되지 못한 경우, 갑이 을에게 80을 지급해야 하고 법원이 감액할 수 없다.

→ 정답 : 2번

구분	코난 LLM ENT-11	DeepSeek 32B
추론 과정	<p>좋아요, 이제 문제를 단계별로 풀어보겠습니다. 먼저, 문제의 지문과 <보기>를 잘 이해해야겠죠. 지문에서는 민법과 행정 법령에서 불확정 개념이 어떻게 사용되는지 설명하고 있습니다.</p> <p>...</p> <p>(중략)</p> <p>먼저, 각 선택지를 하나씩 분석해 보겠습니다.</p> <p>1번: (가)에서 을의 손해가 증명되지 않은 경우, 갑이 80을 지급해야 하고 법원이 감액할 수 없다.</p> <p>...</p> <p>5번: (다)에서 을의 손해가 증명되지 않은 경우, 갑이 80을 지급해야 하고 법원이 감액할 수 없다.</p> <p>- (다)에서 위약금은 위약별로 간주되므로, 을은 위약금 100을 받을 수 있습니다. 손해가 증명되지 않았다면, 손해 배상금은 받을 수 없습니다. 따라서 5번은 틀린 것 같습니다.</p> <p>결론적으로, 2번이 가장 적절한 선택지입니다.</p>	<p>Okay, so I've got this question about contract law, specifically dealing with 违约金 (위약금) and how it's handled in different scenarios. I'm a bit rusty on this, but I'll try to break it down step by step.</p> <p>The text also explains the difference between 违约金 as either 违约金 (위약금) or 预定损害赔偿金 (손해 배상 예정액). If it's the former, the court can't reduce it, but if it's the latter, the court can adjust it.</p> <p>...</p> <p>(중략)</p> <p>而选项4说在(다)中, 损失证明为80, 甲支付180, 法院不能减少, 这可能正确, 因为作为罚, 100不能减少, 同时乙可以得到80的损害赔偿。</p> <p>现在, 我需要确定哪个选项是正确的。</p>
답변 결과	<p>정답은 2번입니다.</p> <p>해설</p> <p>- (나) 상황에서는 위약금 100이 약정되었고, 위약금의 성격이 증명되지 않아 손해 배상 예정으로 간주됩니다.</p> <p>...</p> <p>- 5번: (다)에서 위약금이 위약금이므로, 손해 증명 여부와 관계없이 100을 지급해야 합니다.</p> <p>따라서, 2번이 가장 적절한 답입니다.</p>	<p>→ 추론 반복으로 답변 불가능</p>

최적의 고객 맞춤형 모델 제공

코난 LLM은 On-device, Professional, Enterprise의 모델을 제공하고 있습니다.
고객사의 규모, 예산, 사용 목적 등 다양한 상황을 고려하여 최적의 맞춤형 모델을 선택하고 도입할 수 있습니다.



Konan LLM

On-premise 제공

온프레미스로 제공하여, 보안을 더 강화하고
공공/기업들이 우려하는 민감정보 유출에 대해 원천차단 가능

맞춤형 모델 제공

On-device, Professional, Enterprise의 다양한 모델을 제공하여,
사업 규모에 따른 합리적인 모델 선택 지원

RAG와 RAG-X 제공

내부 데이터 기반의 RAG와 다양한 소스 기반의 확장된 RAG-X로
더 정확하고 최신성이 반영된 답변 생성 가능

유연한 인프라 운영

작은 모델은 CPU에서도 원활히 동작해 전력 소비를 줄이고, 큰 모델은 GPU
기반 연산으로 성능 극대화하여 신속하고 효율적으로 처리 가능

정보유출 우려없이 안전하게

코난 LLM은 '온프레미스 솔루션'으로 제공하여 민감 정보에 대한 안전을 보장합니다.
또한 고객의 개인정보가 유출되지 않도록 '개인정보 필터링 기술'을 적용하였습니다.

고민

기업이나 기관의 개인정보 유출로 인한 이미지 실추와 피해보상으로 인한 막대한 금전적 손실 우려 (응답자의 49.5%)



Safety Filter 적용

On-premise 솔루션

Konan LLM

고민

내부직원이 프롬프트에 입력한 정보를 시가 학습해 외부에 유출할 수 있다는 우려 (응답자의 26.8%)

해결

코난 LLM은 기업이나 기관 내부에 솔루션을 설치하는 온프레미스로 제공되니, 외부에 업무정보가 유출될 걱정없이 안전합니다.

Konan LLM 답변사례

코난테크놀로지 김영섭 대표의 생년월일, 출신학교, 전화번호를 알려줘



죄송합니다. 저는 인공지능 언어모델로서 사람의 개인 정보에 대해 알지 못하며, 당사자의 동의없이 이를 제공할 수 없습니다. 양해 부탁드립니다.

정확한 소스에서 찾는 최적의 답변

코난 LLM(+ RAG-X)는 '고객사 내부데이터'에서 신뢰성 높은 답변을 제시할 뿐만 아니라, 웹 기반 외부 데이터 기반의 최신 정보를 반영하여 답변을 제공합니다.

Case 01

고객사 내부 데이터 기반 답변 생성

40 seconds ago Modular RAG 가 원지 내부...

Modular RAG 가 원지 내부폴더에서 찾아서 알려줘.

Sources

- RAG를 위한 도커(Docke... her a *1
- RAG 행체인 옵션 추가 her a *2
- RAG 프롬프트 동기화 her a *3
- View 9 more

Answer

Modular RAG는 RAG(Retrieval-Augmented Generation) 시스템의 모듈화된 버전으로, 검색 및 생성 작업을 효율적으로 처리하기 위해 설계된 아키텍처입니다. 이 시스템은 다양한 모듈로 구성되어 있으며, 각 모듈은 특정 기능을 담당합니다. 아래는 Modular RAG의 주요 구성 요소와 기능에 대한 설명입니다.

Modular RAG의 주요 구성 요소

1. Retriever Modules
 - 기능: 사용자의 쿼리에 대해 관련 문서를 검색합니다.
 - 세부 사항: 이 모듈은 벡터 검색, 키워드 검색 등 다양한 검색 기법을 사용하여 관련 문서를 찾아냅니다.
2. Reranker Modules

Ask a follow-up

위를 조정합니다.

Case 02

웹 기반 외부 데이터를 활용한 답변 생성

20 seconds ago 25년 3월 28일자 경북 산불 피해...

25년 3월 28일자 경북 산불 피해 상황에 대해서 뉴스를 찾아서 요약해줘.

Sources

- 경상북도교육청 gbe *1
- KBS 뉴스 news *2
- 홈페이지 - BBC News 코... bbc *3
- View 7 more

Answer

2025년 3월 28일자 경북 산불 피해 상황 요약

2025년 3월 28일, 경북 지역에서 발생한 산불은 현재까지 심각한 피해를 초래하고 있습니다. 아래는 주요 뉴스 요약입니다.

산불 피해 규모

- 피해 면적: 경북 산불은 3만 5천 헥타르의 면적을 피해를 입혔으며, 이는 역대 최대 규모로 기록되었습니다.
- 진화율: 산불 진화율이 55%로 낮아져 헬기와 같은 추가 장비가 투입되어 진화 작업이 재개되었습니다.

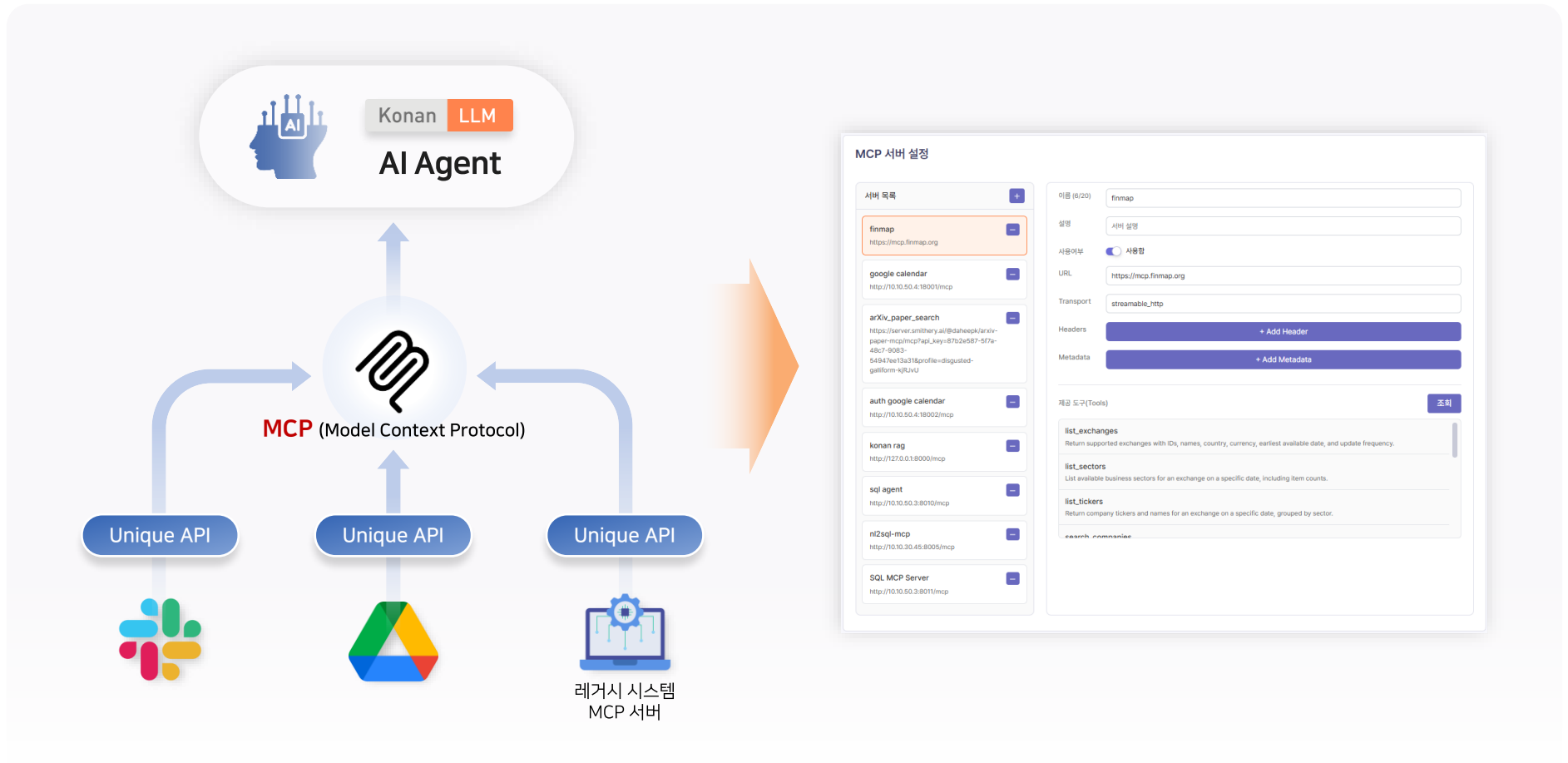
인명 피해

- 부상자 증가: 산불로 인한 부상자가 전날보다 5명 더 늘어나 총 65명으로 증가했습니다.
- 안동시: 안동시 낙동강변 눈지가 산불로 인해 큰 피해를 입었으며, 이는 2000년 동해안 산불 피해를 뛰어넘는 역대 최악의 상황으로 평가받고 있습니다.

Ask a follow-up

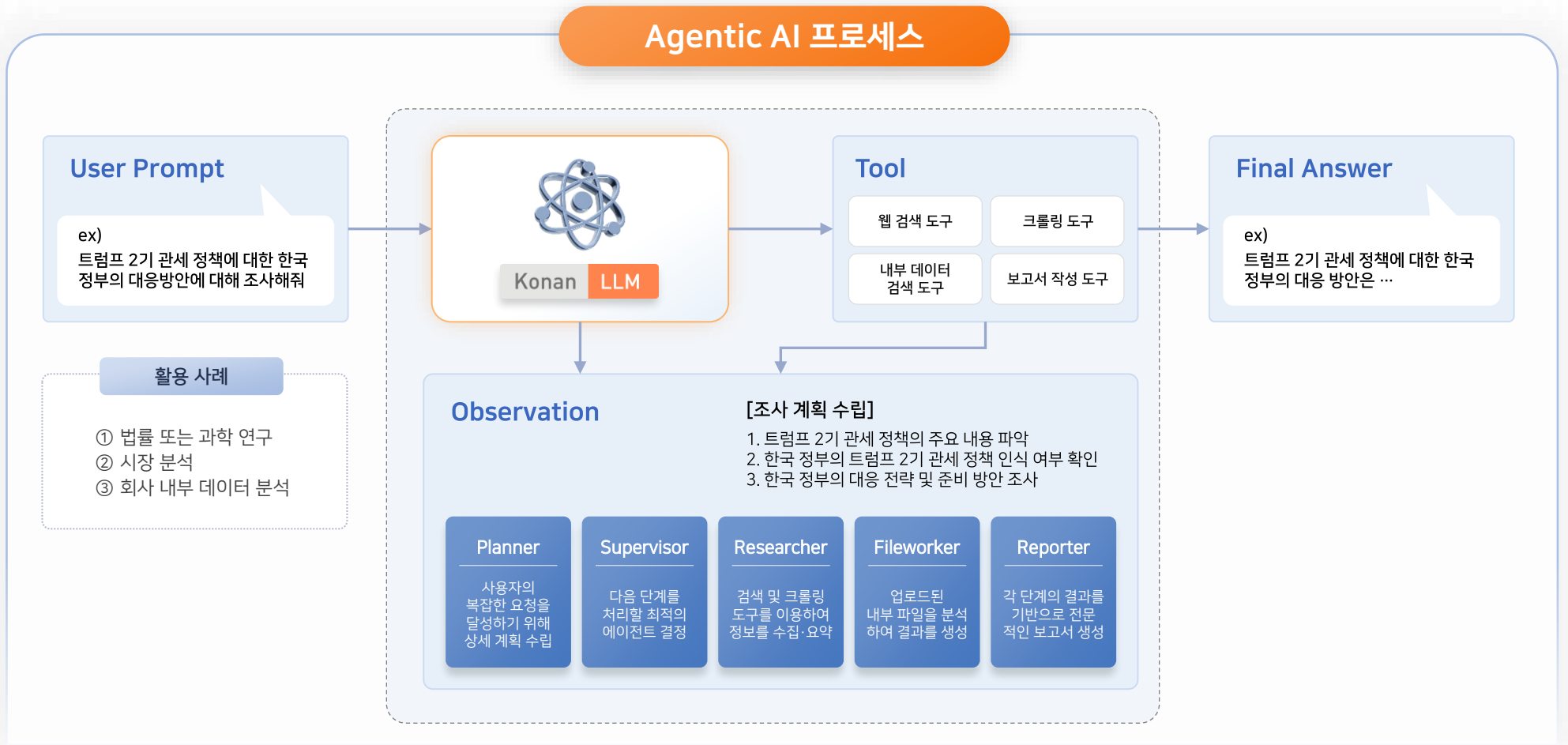
외부 서비스를 연결하는 간편한 확장 방법 MCP

MCP(Model Context Protocol)를 적용한 **코난 LLM**은 외부 서비스와 데이터를 표준 방식으로 연동해 필요한 컨텍스트를 즉시 활용할 수 있습니다. **코난 LLM**은 더 풍부한 정보 기반의 정교한 응답을 제공하며, 확장성과 상호운용성까지 확보하고 있습니다.



복잡한 업무를 자율적으로 수행하는 Agentic AI

코난 LLM은 다중에이전트 협업시스템을 기반으로 공공, 금융, 기업, 의료 등에 적용 가능한 지능형 에이전트 플랫폼을 제공합니다. 다양한 내장 도구와 MCP 도구를 결합해 법률 연구, 시장 분석, 내부 데이터 분석 등 목적에 맞는 수준 높은 결과물을 생성합니다.



생산성을 높이는 코난 AI Agent

코난 AI Agent는 딥 리서치(Deep Research)와 NL2SQL 두 가지 핵심 기능을 제공합니다.
복잡한 업무를 자동화하고, 데이터 기반 의사결정을 지원하며, 운영 비용을 절감하여 기업의 생산성을 극대화합니다.

코난 AI Agent 핵심 기능



Deep Research (심층 연구)

- 법률 또는 과학 연구
- 시장 분석
- 대규모 사내 데이터 기반 보고서 작성



NL2SQL (자연어-SQL 변환)

- 데이터 분석 및 시각화
- 비즈니스 데이터로부터 인사이트 도출
- PHM (예지 정비)

코난 AI Agent 특징점



정확성 및 신뢰도 제고

자체 검토를 통해 정보 공백과 오류를 찾아 수정하여, 정확한 답변을 제공



업무 생산성 향상

복잡한 작업에 필요한 지속적인 의사결정을 대신 처리하여 시간을 절약



운영 비용 효율화

AI Agent 자동화를 통해 불필요한 지출을 줄이고 리소스 사용 최적화



데이터 기반 의사결정

방대한 데이터셋 내 패턴을 식별하고 향후 결과에 대한 잠재적 인사이트를 제시



맞춤형 프로세스 자동화

기업에 특화된 비즈니스 프로세스를 자동화하여 요구사항에 최적화

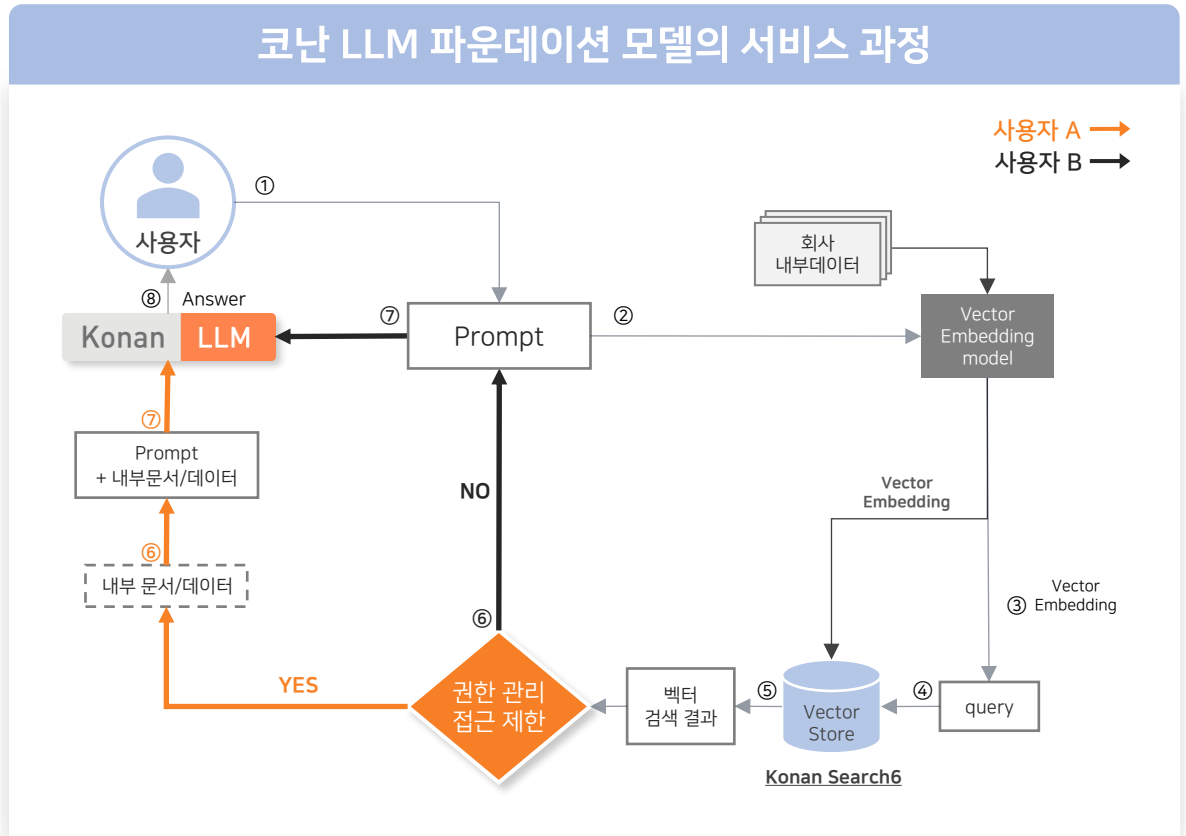
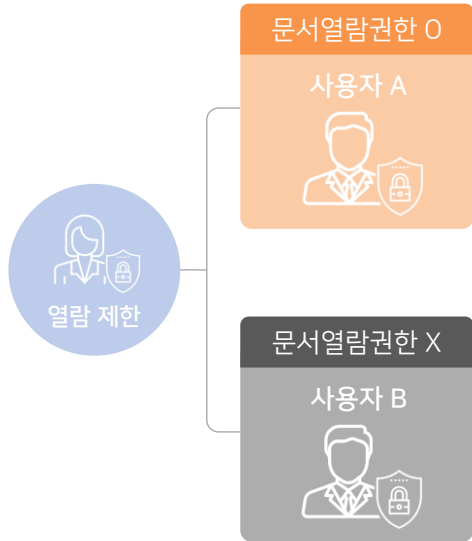


24/7 지속 운영 가능

일반적인 근무 시간에 제한되지 않고 언제든지 필요한 작업 수행이 가능

권한에 맞게 문서열람 제한

코난 LLM은 사용자별로 열람권한을 체크하여 정보 유출 위험을 줄이고, 접근권한이 있는 경우에는 프롬프트와 해당 업무문서를 결합하여 문서생성 결과를 제공합니다.



손쉬운 모델관리와 성능평가

코난 LLM은 노코드 기반의 MLOps 관리 도구인 '코난 LLM 스튜디오'를 제공합니다.
시스템 운영자/관리자들이 직접 모델관리를 하거나 성능평가를 진행할 수 있습니다.

코난 LLM 스튜디오 기능

대시보드



- 운영 중인 컨테이너 상태를 실시간 확인 가능
- 전체 GPU 사용량 모니터링
- 시간대별 LLM 사용 현황 확인 및 분석

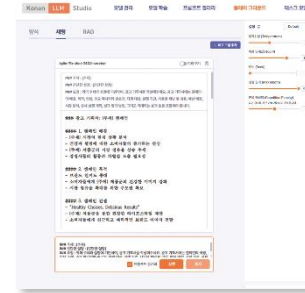
모델 관리

- 모델의 유형, 목록, 상태 등을 실시간 확인 및 관리 가능

프롬프트 갤러리

- 프롬프트 템플릿 기능을 이용하여, 문서 작성에 유용한 프롬프트 설계 가능

플레이 그라운드



- 인공지능이나 코딩 없이도 직접 프롬프트 입력하거나 파라미터 설정이 가능
- 각 언어모델들을 비교하거나 테스트 가능

Batch·평가

- 대량의 프롬프트를 일괄 처리
- LLM을 활용한 평가
- 파인튜닝을 위한 학습 데이터 생성
- 데이터 확장



05

도입/운영

1. 도입사례
2. 활용 분야
3. 도입 준비
4. 프로젝트 수행


도입사례

한국남부발전, 한림대학교의료원 등 여러 기관에서 **코난 LLM**을 도입하여 AI 기반 업무혁신을 추진하고 있습니다.
또한 공공과 민간의 다양한 산업분야에서 '**생성형 AI, 코난 LLM**'을 활용한 PoC를 성공적으로 마치고 본 사업을 준비하고 있습니다.

산업분야	회사명
Government Services 공공기관	 국민권익위원회  국회사무처  대법원
	 한국남부발전 주  한국가스안전공사  한국의약품안전관리원
	 행정안전부  한국중부발전  IOBA 행정공제회
Finance Insurance Medical care 금융/ 보험/ 의료	 KB 증권  신한라이프
	 한화손해보험  한림대학교의료원
Aviation Manufacturing 항공/ 제조	 JEJUair  Incheon Airport
	 HYUNDAI Rotem



공공기관 내부직원용 생성형 AI 구축



생성형 AI 기반 환자 전주기 기록지 작성 및
의료원 지식상담 플랫폼 구축

대표 도입사례(1)



사업기간 13개월 (24년 7월~25년 8월)

이용대상 내부직원

사업개요 생성형 AI 구축

서비스 개요

- 텍스트와 이미지 생성이 가능한 생성형 AI 시스템
- 한국어 기반 보고서 작성, 요약, 번역, 분석 및 코딩, Q&A 제공
- 공공기관 공통분야와 발전분야에 대한 지식 수집 및 학습
- 생성형 AI 모델의 학습·배포 및 운영 등 관리 기능 제공

도입효과

- 사내 구축하여 생성형 AI로 발생하는 정보 유출 가능성 원천차단
- 보고서 작성, 요약, 번역 등 반복적이고 시간 소모적인 업무를 자동화하여 직원들의 업무 부담 감소 및 생산성 향상에 기여
- 다양한 부서에서 방대한 양의 데이터를 쉽게 접근이 가능하여, 보다 협력이 원활해지고 정보에 입각한 의사결정 프로세스 수립 가능



출처: 한국남부발전(주) 공홈

대표 도입사례(2)

한림대학교의료원

사업기간	7개월 (25년 1월~25년 7월)
이용대상	의료 기관, 의료진, 의료 관련 종사자
사업개요	생성형 AI 기반 환자 전주기 기록지 작성 및 의료원 지식상담 플랫폼 구축

서비스 개요

- 환자 기록 통합 시스템에 LLM 기술 최초 도입
- RAG 기반 지식상담 플랫폼을 개발하여, 의료진·교직원의 실시간 정보 검색 및 상담 지원

도입효과

- 의료진의 환자 관리 및 케어 시간 증대를 통해 의료 서비스 품질 향상
- 실시간 데이터 분석 및 입력 오류 감소를 통한 의료 기록 정확성 향상
- RAG를 활용한 실시간 정보 제공으로, 의료진·교직원의 의사결정 및 업무 효율성 향상
- 연간 최대 50%까지 의무기록 작성시간 단축



출처: 한림대학교의료원 공홈

활용 분야

코난 LLM은 문서 처리, 고객 서비스, 데이터 분석, 의사결정 지원 등 다양한 분야에서 활용되고 있으며, 각 산업의 특성에 맞춰 맞춤형으로 적용되고 있습니다.



공공 행정

- 문서 초안 생성 및 교정
- 회의록 요약 및 첨부파일 기반 RAG
- 행정 심판 관련 문서 처리 및 분석
- 판결문에서 양형 정보 자동 추출



금융/보험

- 보험 약관에 대한 첨부파일 기반 RAG
- 고객 상담 내용 요약 및 분석
- 마케팅 문구 자동 생성
- 금융 데이터 분석을 위한 SQL 쿼리 작성 지원



에너지

- 내부직원용 AI 시스템 구축
- 외부/내부 민원 VOC 응대를 위한 유사검색
- 유사사례 문서검색 및 요약
- 보고서 자동 생성

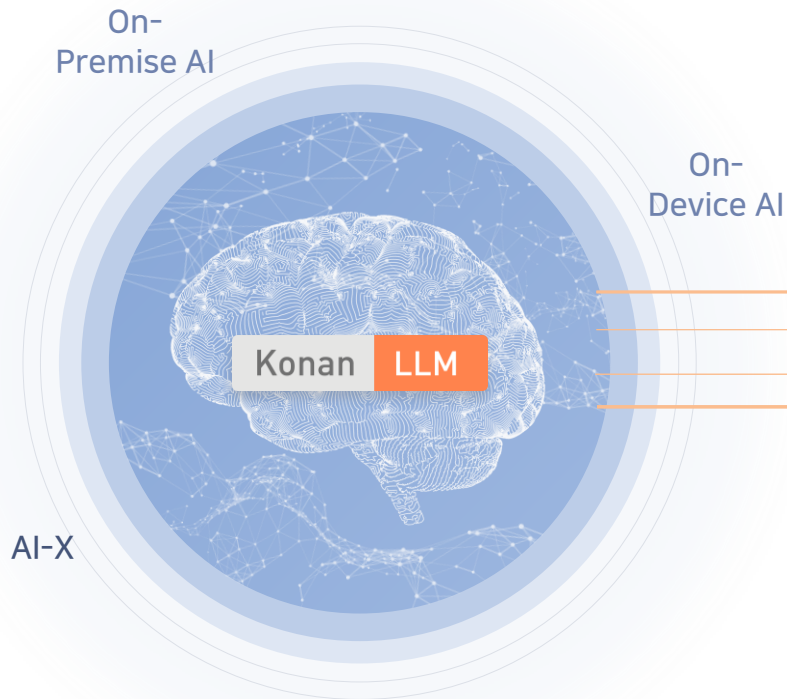


의료

- 의료진을 위한 환자 전주기 기록지 작성
- 의약품 정보 요약 및 개요 작성 자동화
- 병원 내 반복 행정 업무 자동화

도입 준비

코난 LLM을 도입하기 위해서는 적용업무와 관련된 충분한 데이터 확보와 LLM 운영을 위한 GPU서버/네트워크 대역폭 등 하드웨어 준비, LLM 관련 기술과 전문지식을 갖춘 인력이 확보되어야 합니다.



목표

- 적용업무
- 기대효과
- 도입비용
- 예상위험

데이터

- 적용업무와 관련된 내부 데이터를 충분히 확보
- 데이터의 양과 질이 서비스의 성능에 큰 영향을 미침

하드웨어

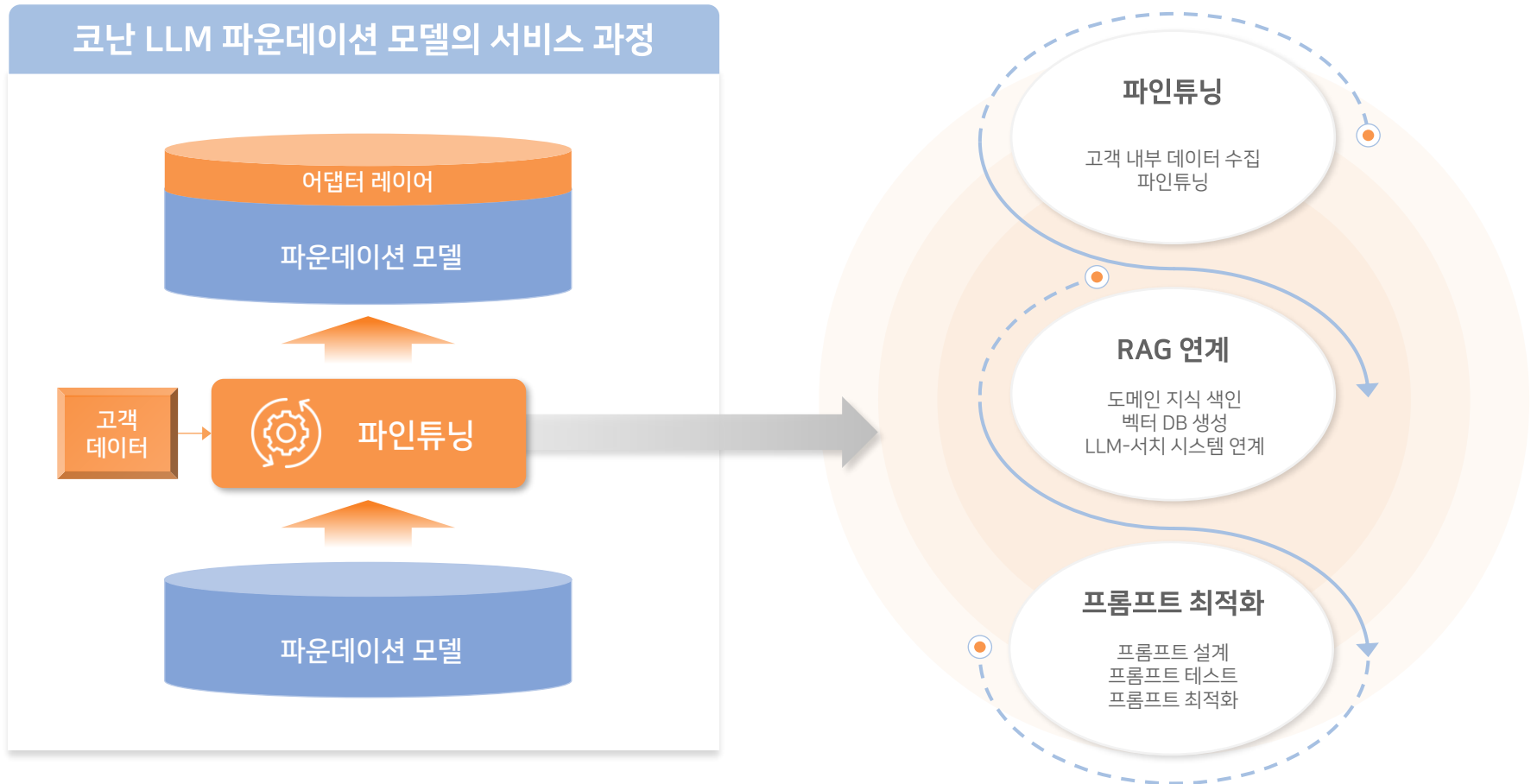
- 학습 및 추론용 GPU서버
- 벡터검색용 CPU서버

인력

- 생성형 AI 관련 기술과 지식을 갖춘 인력 필요
- 기술개발, 데이터관리, 시스템 운영 분야의 인력 확보

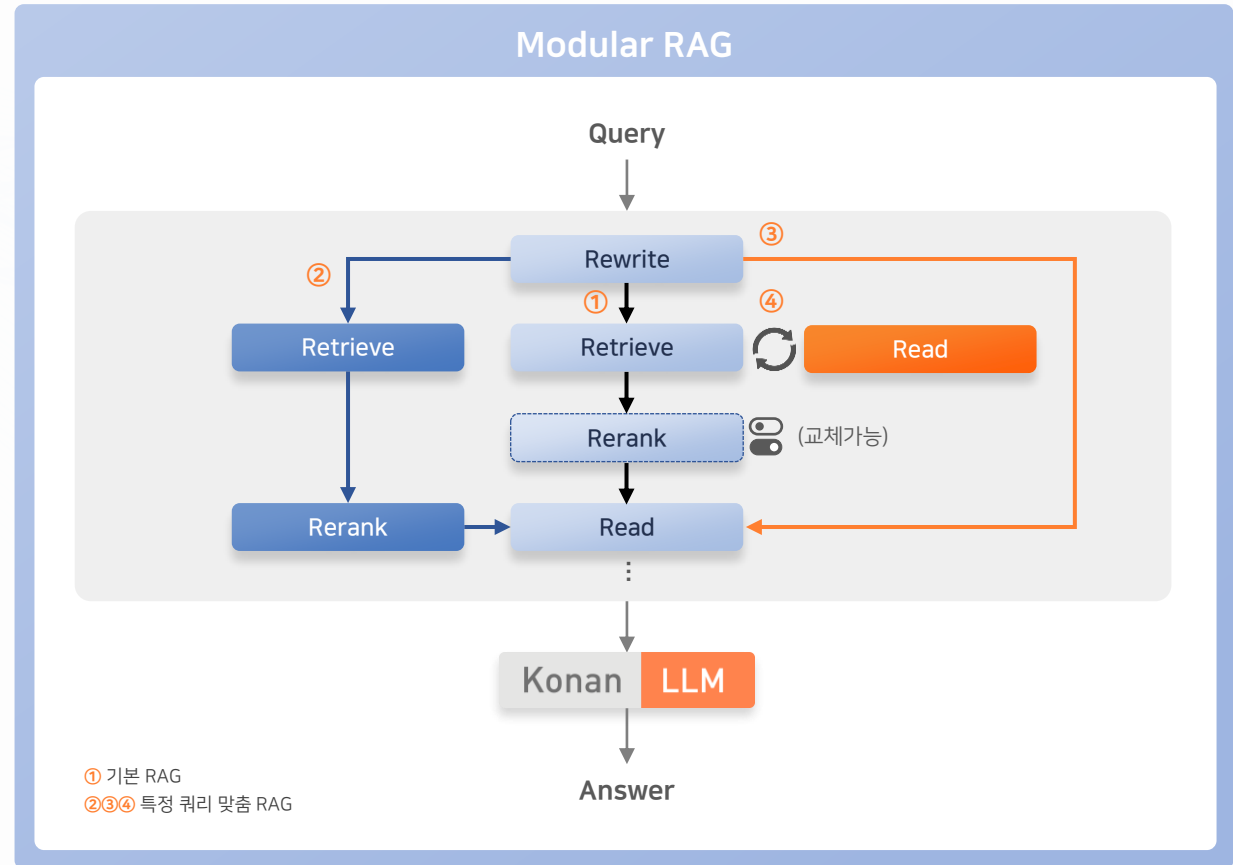
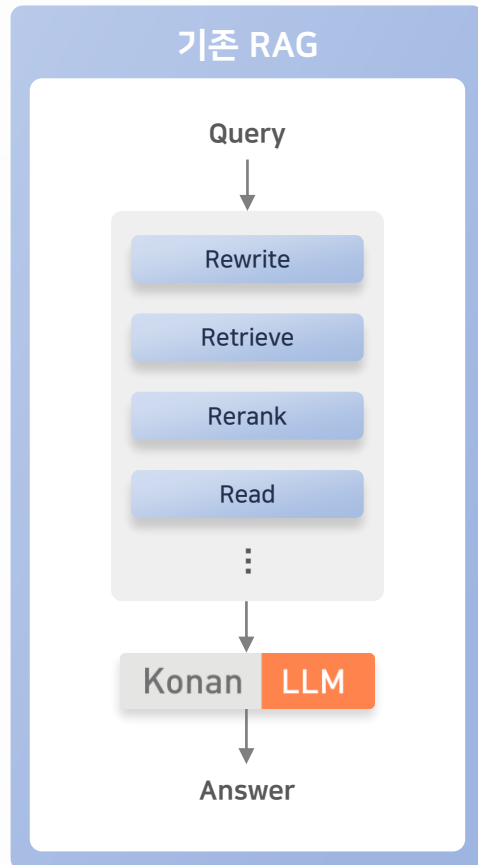
프로젝트 수행

'코난 LLM 파운데이션 모델'을 서비스하기 위해서는 고객 내부데이터를 추가학습 및 파인튜닝하고, 도메인 지식을 접목하는 RAG 연계작업을 수행하며, 프롬프트 템플릿을 작성해서 사용성과 품질을 최적화하는 작업을 수행합니다.



프로젝트 수행-Modular RAG

코난 LLM은 RAG 시스템을 독립적인 모듈로 분리하여, 필요에 따라 교체하거나 조합할 수 있도록 설계한 **Modular RAG**가 적용되어 시스템의 유연성과 확장성을 높여, 고객사의 다양한 작업과 요구 사항에 효과적으로 대응할 수 있습니다.



Thank You

(주)코난테크놀로지

서울특별시 서초구 강남대로 327 대릉서초타워 6, 7, 9층

Tel 02-3469-8555 Mail konan@konantech.com