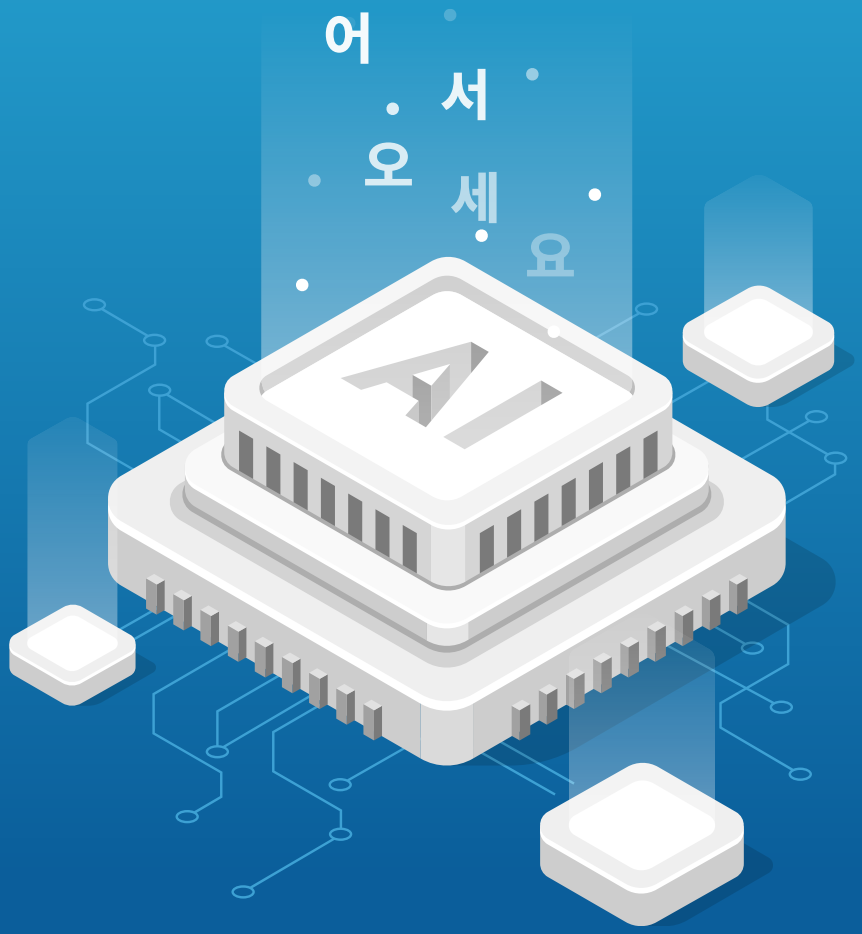
KONAN
TECHNOLOGY

Leap, Leap ahead

Konan LLM

대규모 언어모델



개요

코난 LLM은 코난테크놀로지의 자연어처리 전문 AI연구소에서 개발한 문서생성AI(대규모 언어모델)로 공공기관 및 기업의 업무혁신을 가속화합니다.

기술력

AI 전문 연구개발인력이 전체의 67%

최고 수준의 한국어 심층 자연어 처리 기술 적용
SP 2등급 받은 표준 개발 프로세스를 적용

Data

205억건 이상 문서 보유

2007년부터 '펄스K' 서비스를 위해 문서 수집
이 중 양질의 문서 약 20억 건을 코난 LLM 학습에 사용

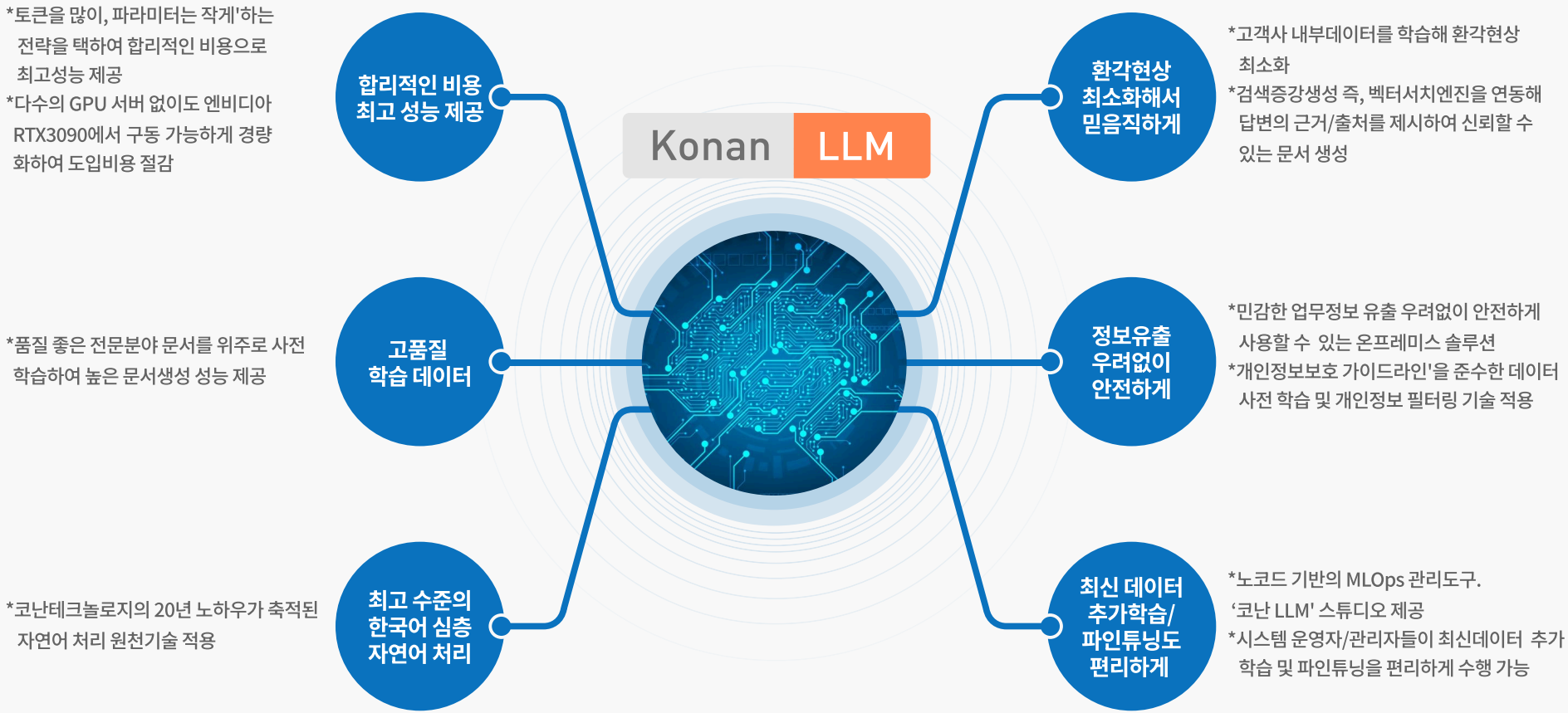
GPU서버

고성능 GPU서버 운영 노하우

대규모 언어모델 학습에 필요한
고성능 GPU 서버 다수 보유 & 운영 경험

특징

코난 LLM은 합리적인 비용으로 최고 성능을 제공하는 온프레미스 솔루션입니다.
RAG를 연동해 고객사 내부데이터와 최신데이터를 기반으로 신뢰도 높은 답변을 생성합니다.



활용분야

코난 LLM은 대표적인 능력인 문서생성 외에도 질문-답변, 요약, 고쳐쓰기, 분류 등 다양한 자연어 처리 능력을 갖고 있습니다.

자연어 인터페이스

물류, 제조, ERP 등
기간계 시스템에 자연어로
데이터 조회/분석/답변 가능

업무문서/보고서 초안 생성

문서 작성 요청 프롬프트에
입력한 내용에 따라
자동으로 초안 생성

고객 맞춤 상담

AICC, 챗봇 등
고객 상담 시스템에서
고객 개인 정보에 기반한
답변 초안 생성

계약서 검토

계약서 내 불공정 또는
함정 조항 탐지

고객 상담 후처리 자동화

상담 내용 자동요약/분류/예약 등
레거시 시스템과 자동 연동

타겟 마케팅

상담 내용 기반 고객 관심사
맞춤형 문구 초안 생성

지속적 정보 수집과 실시간 대응 방안 초안 생성

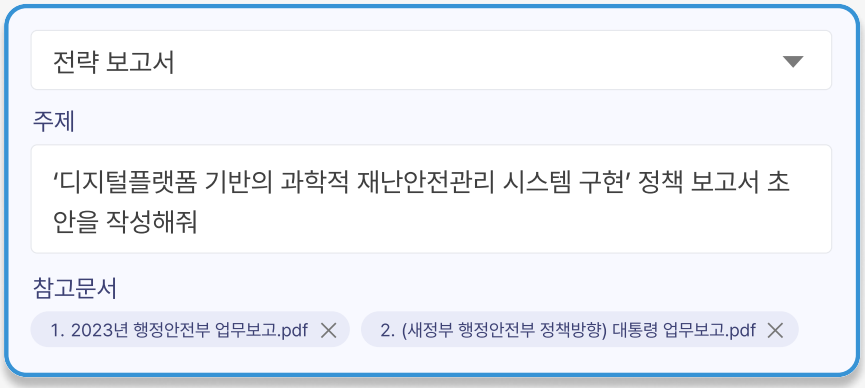
정보나 국방 분야에서 지속적으로
정보수집 시 효과적인 요약 제공/
긴급 시 대처방안 초안 생성

생성사례

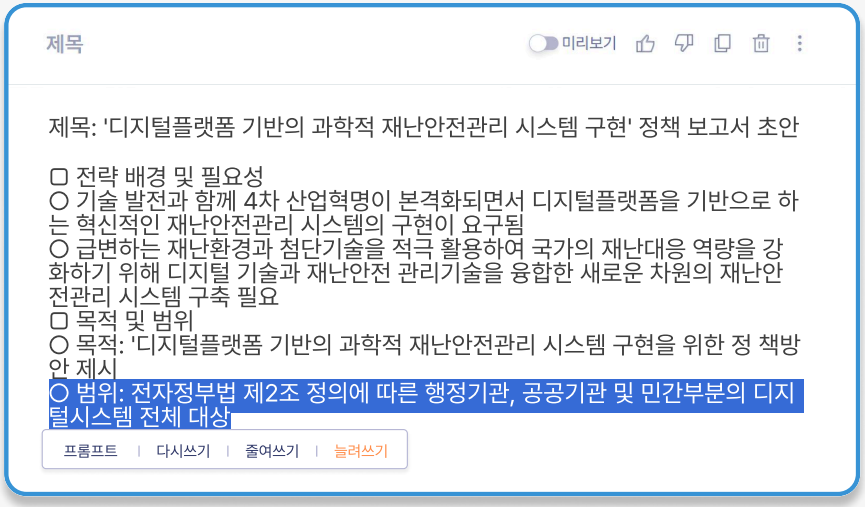
코난 LLM이 생성한 '행정기관의 정책 보고서 초안' 사례입니다.

문서초안 생성사례

프롬프트 입력



문서 생성

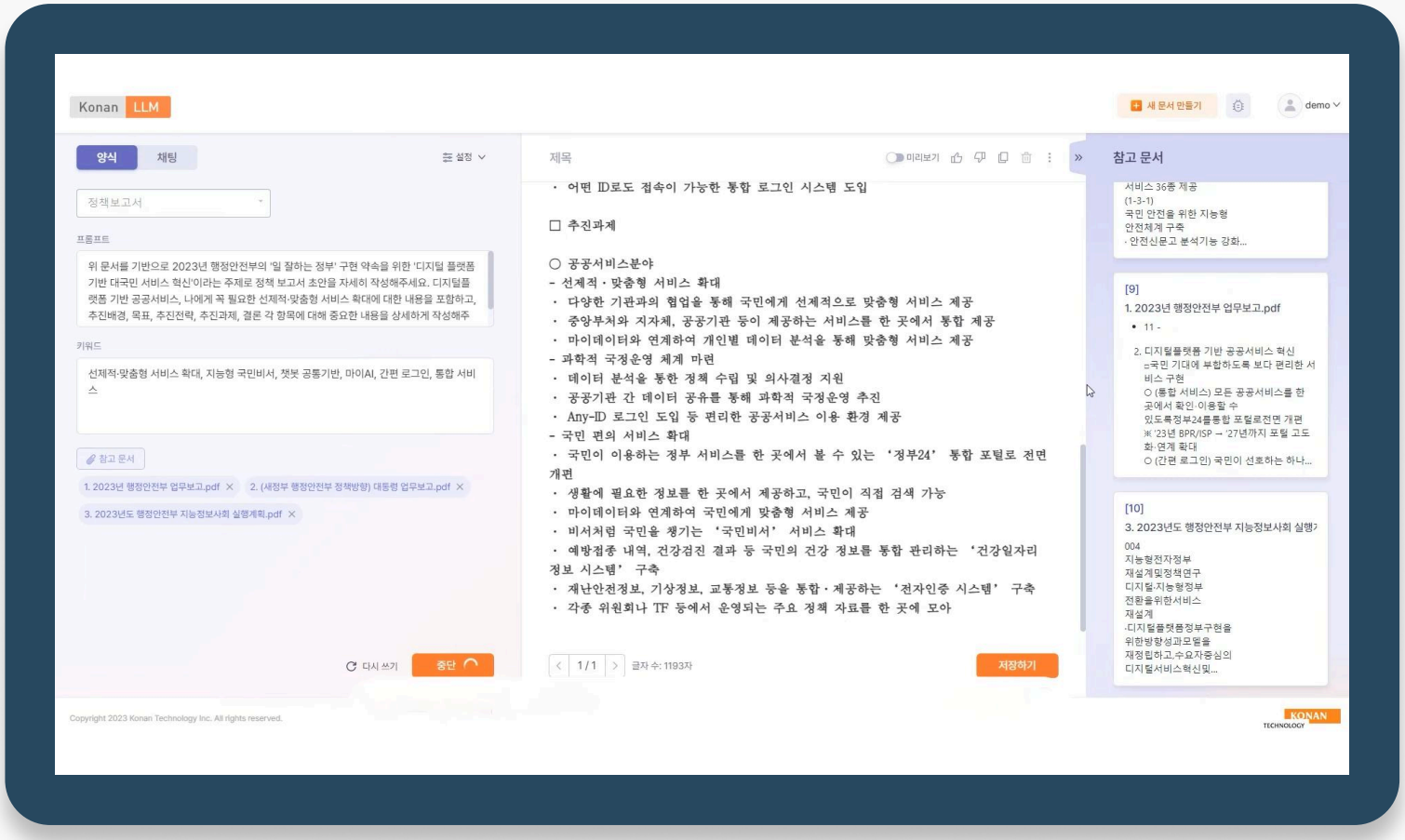


보고서 초안 작성 방법

1. **프롬프트 입력**
   원하는 문서에 대한 질문 작성
2. **초안 생성**
   공공기관 보고서 형식에 맞춰 제목, 목차, 내용 생성
3. **문서 보완**
   선택한 문장에 '늘려쓰기', '줄여쓰기', '다시쓰기' 기능을 활용하여 문서 보안
4. **보고서 완성**
   보고용 문서로 완성하기 위해 미리보기, 저장하기

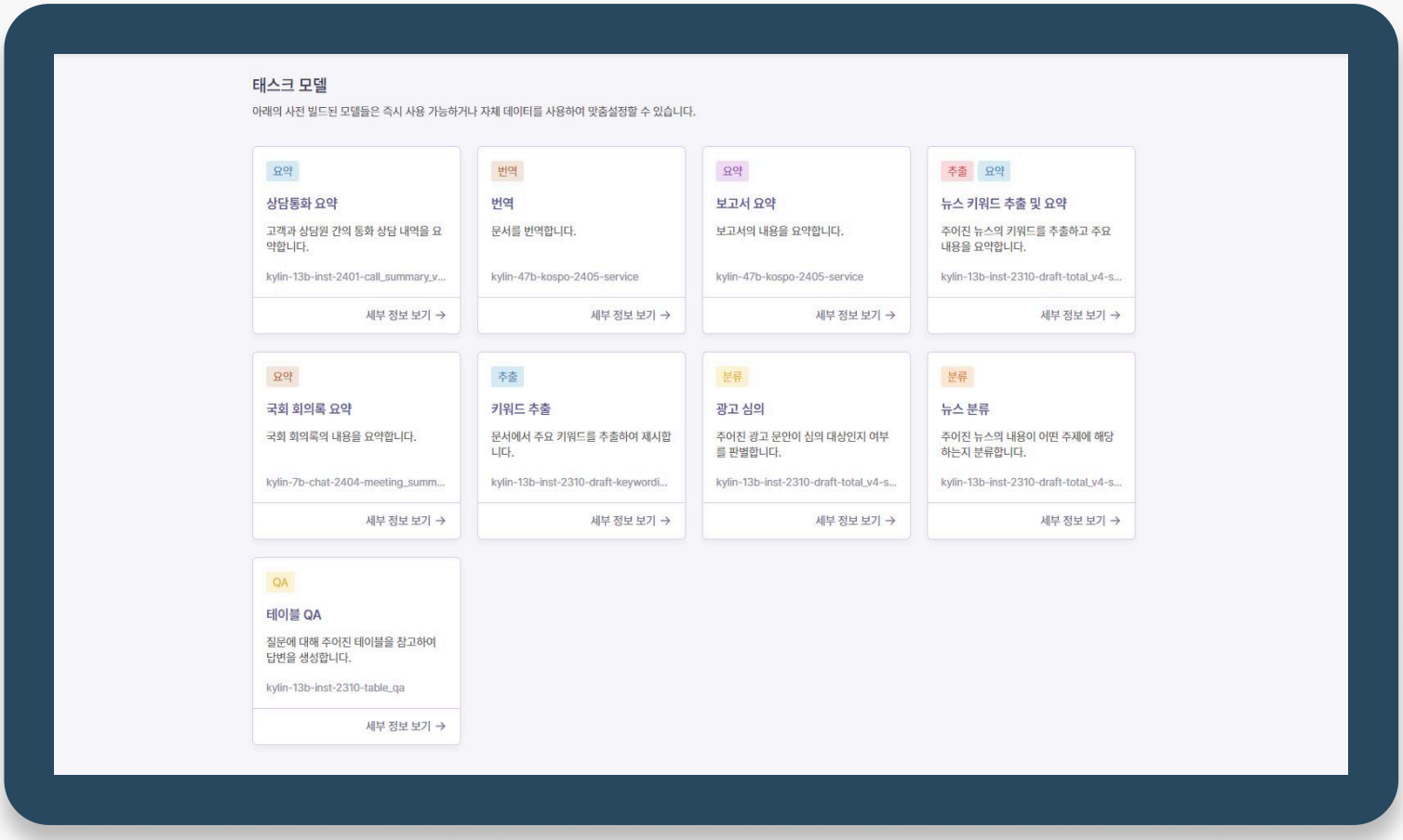
질문답변 사례

관련 법령 등 참고 문서에 근거하여 적절한 답변을 제시합니다.



비디오

코난 LLM 스튜디오는 생성형 언어 모델 코난 LLM을 빠르게 프로토타이핑하고 테스트 해볼 수 있는 도구입니다.



FAQ

코난 LLM 도입 전 궁금증은 FAQ를 통해 확인하세요.

자주 묻는 질문

기본 제공 탬플릿 외 맞춤보고서 생성도 가능한가요?

네. 파인튜닝을 통해 맞춤 형식의 보고서 출력이 가능합니다.

표나 그래프도 작성 가능한가요?

표는 마크다운과 html 형식으로 작성 가능합니다. 향후 그래프 작성도 가능해지면, 제품 업데이트를 통해 알려드리겠습니다.

한국어 외 어떤 외국어를 지원하나요?

현재 영어를 지원하고 있으며, 다양한 언어로 지원 확대할 예정입니다.

프롬프트에 입력 가능한 데이터는 어떤 것들이 있나요?

현재는 텍스트만 입력이 가능합니다.

프롬프트에 참고자료로 추가되는 첨부파일의 개수 제한이 있나요?

아니요. 제한 없습니다.

프롬프트에 입력할 수 있는 글자 수나 단어 수 제한이 있나요?

4k 토큰까지 입력 가능합니다. 4k는 A4 용지 6페이지 분량입니다.

문서저장 시 파일포맷은 어떤 것들을 지원하나요?

docx, txt, hwpx 등 다양한 형식으로 저장 가능 합니다.

개인정보 보호조치는 어떻게 하고 있나요?

고객 개인정보 유출 피해가 발생하지 않도록 안전하게 보호하고 있습니다. 데이터 사전학습 과정에서부터 개인정보 보호 가이드라인을 준수하여 '개인정보 필터링'을 지원합니다.

파인튜닝을 쉽게 할 수 있는 방법을 제공하나요?

네. '코난 LLM 스튜디오'를 제공하여 최신데이터 추가학습 및 파인튜닝을 쉽고 편리하게 수행할 수 있습니다.

기반모델은 무엇인가요? 자체개발했나요?

Transformer 모델의 디코더에 기반하여 자체개발한 모델입니다.

클라우드로도 제공하나요?

아니요. 클라우드 상에서 SaaS형태로 제공하지 않지만, 고객사의 프라이빗 클라우드 위에서는 제공 가능합니다.

상품라인업은 어떻게 되나요?

코난 LLM은 고객의 사용목적에 따라 3종의 모델을 제공하고 있습니다.

구분	모델
On-Device용	코난 LLM OND
단위업무 규모	코난 LLM PRO
전사업무 규모	코난 LLM ENT

GPU 하나로 돌아가나요?

네. 코난 LLM PRO 모델의 추론서버는 GPU 하나로 운영 가능합니다. 다만, 학습 데이터양에 따라 필요 GPU개수는 더 추가될 수 있습니다. 자세한 내용은 아래 표를 참조하세요.

모델	NVIDIA H100		NVIDIA A6000		비고
	추론[1]	학습[3]	추론[2]	학습[3]	
OND 10	-	-	-	-	1) 입력 3K, 출력 1K, 동시 사용자 500명 기준 2) 동시 사용자 50명 기준 3) 4K 토큰 길이 10,000건 학습 데이터 12시간 이내 학습 기준 4) 10,000건 학습 데이터 17시간 학습 기준
OND 20	-	-	-	-	
PRO 10	1장	2장	1장	2장	
PRO 20	1장	2장	1장	2장	
ENT 10	2장	2장	2장	4장	
ENT 20	2장[2]	2장	4장	8장[4]	

*참고로 추론 외에도 서버운영, 토크나이저, 검증 등의 작업을 위해서는 별도의 CPU장비가 필요합니다.
이때 CPU 최소스펙은 "Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz" 입니다.

필요한 하드웨어는 무엇 무엇이 있나요?

학습 및 추론용 GPU서버가 필요합니다. RAG 적용 시 추가적으로 벡터검색용 GPU 및 CPU 서버가 필요합니다.