

기업 데이터 기반의 학습, 호환성과 보안까지 갖춘 솔트웨어 Sapie-LLM을 소개합니다.

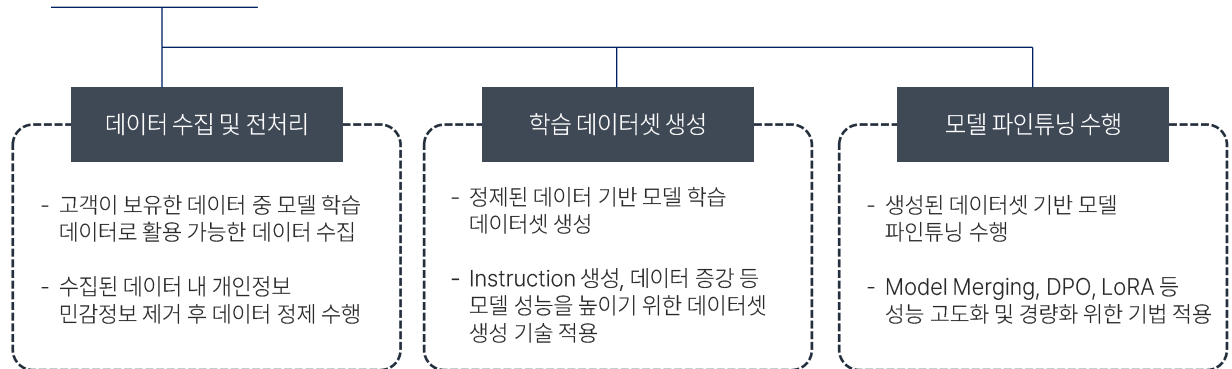
Sapie-LLM은 솔트웨어가 자체 개발한 AI 솔루션으로 뛰어난 한국어 이해력과 향상된 생성 성능을 제공합니다. 외부 연결 없이 기업 내에서 안전하게 AI를 활용할 수 있도록 설계되었으며, 고객의 민감 정보와 데이터를 수집해 제공하고, 학습 데이터셋을 생성해 모델 파인튜닝까지 수행합니다.

01. 어떤 고객에게 필요할까요?

- 폐쇄망 환경에서 보안과 높은 성능을 갖춘 AI 기능을 구축하고 싶은 분
- 금융, 법률, 제조 등 특정 도메인에 최적화된 생성형 AI가 필요한 분
- 자체 데이터를 학습시켜 비즈니스에 특화된 AI를 구현하고 싶은 분

02. 어떤 특징점이 있나요?

- 미국 웨이트앤바이어시스(Weights & Biases)에서 오픈소스 LLM 분야 세계 1위를 기록한 검증된 기술력
- LLaMA, Gemma 등 오픈소스 모델의 한계를 뛰어넘는 한국어 이해력과 도메인 최적화
- 클라우드 및 온프레미스 환경에서의 유연한 연동 지원
- OpenAI API와 완벽하게 호환되는 인터페이스 제공 및 유연한 API 설계와 표준 준수
- 고객이 제공하는 데이터를 기반으로 산업 맞춤형 문서 자동 작성 및 반복 업무 자동화
- 도메인 특화 학습 지원



* Model Merging : 복수 모델의 가중치(가치)를 결합하여 최종 모델 성능을 향상시키는 방법
* DPO : 인간의 선호도를 학습에 직접 반영하는 방법
* LoRA : 원본 모델의 가중치는 고정된 채로 저자원의 행렬(LoRA 모듈)을 추가해 해당 행렬의 파라미터만 수정하는 방법

03. 서비스 도입 절차는?

- vLLM이나 Ollama 등 주요 모델 서빙 프레임워크를 지원해 고객 환경에 따라 유연한 구축 지원

Step 1. 실행 환경 구성

- 서비스 사용을 위한 인프라 준비
- 운영체제 : Ubuntu 22.04 이상
- 필수구성 : GPU Server(NVIDIA 지원), Docker 기반의 운영 환경

Step 2. 서버 실행 및 배포

Docker 기반의 LLM 추론 서버(vLLM) 구축 및 로컬 또는 Hugging Face의 Sapie-LLM을 이용한 서비스 제공

Step 3. 서비스 구축

OpenAI API기반 구축 및 제공되는 서비스의 종류에 따라 추가적인 지원 (챗봇 서비스 요청 시 Sapie-Bot 제안)