

랩 위드의 AI 기반 솔루션 >

"외부 API 없이, 자체 AI로 컨설팅한다 — 랩 위드 R&D AI 서버 구축기"



기술사업화 랩 위드
2026.07.08. 15:06 조회 26

댓글 0 URL 복사

랩 위드 R&D AI 서버 상태

정상 가동

갱신 2026-07-08 14:19:00 · 30초 자동

시스템 작동

정상

컨테이너 18개

지식DB 보유

3,104건

연구계획서 3,086 · 제안서 18

임베딩 모델

bge-m3

한국어 특화 벡터검색

검색(RAG)

정상

응답 173ms · 결과 8

일별 질의수(24h)

1

/rmd-chat 기준

인프라 · 보안 (실시간 점검)

nginx

정상

설정·기동

홈페이지

정상

www.labwith.kr

포탈 R&D

정상

로그인 보호

SSL 인증서

정상

만료 2026-07-08

공인 IP

정상

172.18.0.1

크래시방지

정상

RND 토큰 3/3

재임베딩 / 색인 진행

100% · 완료 3,104 / 처리중 0 / 오류 0 / 전체 3,104

컨테이너 · 서비스

nginx

plugin_daemon

sandbox

worker

worker_beat

api_websocket

api

web

unruffled_dijkstra

authentik-server

authentik-worker

authentik-postgresql

authentik-redis

certbot

redis

ssrf_proxy

db_postgres

weaviate

GPU (RTX 5060 Ti)

사용률 0% · VRAM 8,087/16,311MB

상주 모델: exaone3.5:7.8b, bge-m3:latest

자료 · 보안

임베딩: bge-m3

색인 오류: 0건 (0%)

로그인: Authentik 로그인 · 병렬4

분석 · 진단

컨테이너·nginx·사이트·인증서·IP 모두 정상

최근 접속 로그

시간	IP	경로
08/Jul/2026:05:19:32 +0000	172.18.0.1	/md/
08/Jul/2026:05:19:32 +0000	172.18.0.1	/portal/
08/Jul/2026:05:19:32 +0000	172.18.0.1	
08/Jul/2026:05:19:32 +0000	172.18.0.1	
08/Jul/2026:05:19:32 +0000	172.18.0.1	
08/Jul/2026:05:19:31 +0000	172.18.0.1	
08/Jul/2026:05:19:31 +0000	172.18.0.1	
08/Jul/2026:05:19:30 +0000	172.18.0.1	

기술사업화 컨설팅 업계에서 AI를 도입하는 방식은 크게 두 가지입니다.
하나는 ChatGPT·Claude 같은 외부 서비스에 문서를 올려 답을 얻는 방식, 또 하나는 자체 서버에 AI를 직접 구축해 운영하는 방식입니다.

랩 위드(Lab WiTH)는 두 번째 방식을 선택했습니다.

① 왜 자체 AI 서버인가? — 필요성

컨설팅 현장에서 가장 민감한 것은 **고객사 데이터의 보안**입니다.
기업의 사업계획서, R&D 전략, 기술 정보는 외부 서버에 올리기 어렵습니다.

또한 국가 R&D 과제 데이터를 기반으로 한 전문 답변은, 일반 AI가 학습한 공개 데이터만으로는 한계가 있습니다.

이에 랩 위드는 다음 3가지 목표를 세웠습니다.

- ① **데이터 외부 유출 Zero** — 고객 정보가 서버 밖으로 나가지 않는 폐쇄형 구조
- ② **도메인 특화 AI** — 국가 R&D 사업계획서·제안서 전용 지식베이스 구축
- ③ **24시간 안정 운영** — 컨설턴트가 언제든지 즉시 AI 검색 가능

② 무엇을 만들었나? — 구축 내용

2026년 7월 현재 운영 중인 랩 위드 R&D AI 서버 현황입니다.

항목	내용
운영 컨테이너	18개 (Docker 기반 마이크로서비스)
AI 모델	EXAONE 3.5 (7.8B) — LG AI Research 한국어 특화 LLM
임베딩 모델	bge-m3 — 한국어 최적화 벡터 검색
GPU	RTX 5060 Ti (VRAM 32GB)
지식DB 규모	3,104건
색인 완료율	100% (오류 0건)
RAG 검색 응답속도	173ms
보안	Authentik SSO 로그인 보호, SSL 인증서 유지
인프라	nginx + Weaviate 벡터DB + PostgreSQL + Redis

모든 서비스는 **정상 작동 중**이며, 30초 단위 자동 상태 점검이 이루어집니다.

핵심은 **RAG(Retrieval-Augmented Generation)** 구조입니다.

질문을 입력하면, 3,104건의 R&D 문서 중 가장 관련성 높은 내용을 벡터 검색으로 추출한 뒤, 한국어 특화 LLM이 전략적 답변을 생성합니다.

외부 인터넷 없이, 완전히 사내 서버 안에서 처리됩니다.

③ 어떤 성과가 나오나? — 활용 효과

이 시스템이 실제 컨설팅에서 만들어내는 차이는 명확합니다.

Before (기존 방식):

과제 선정 분석에 1~2일 소요

담당자 경험과 직관에 의존

사업계획서 유사 사례 검색 수작업

After (AI 지원 방식):

관련 R&D 과제 검색 → **173ms 이내**

3,104건 사업계획서 기반 전략 방향 즉시 도출

기술동향·경쟁 과제·성공 패턴 자동 분석

기술이전, R&D 사전기획, BM 도출, 사업화전략 수립 전 과정에서 **AI가 컨설턴트의 분석 속도와 깊이를 높여주는 파트너**로 작동합니다.

마치며

랩 위드의 AI 서버는 상품이 아닙니다.

고객에게 더 빠르고, 더 정확하고, 더 안전한 기술사업화 컨설팅을 제공하기 위해 내부에서 쌓아온 역량입니다.

앞으로도 지식DB를 지속 확장하고, 기업진단·IR·R&D전략 전 서비스에 AI를 깊이 연동해 나갈 계획입니다.

기술과 사업화의 간극을 AI로 연결합니다.

☎ 010-6408-8557 | ✉ antipoto@gmail.com | 🌐 www.labwith.kr



기술사업화 랩 위드

댓글을 남겨보세요

