KT NPU LLM Station

공공기관 폐쇄망 안에서 바로 쓰는
온프레미스 구독형 생성형 AI

kt Enterprise enterprise.kt.com ax.kt.com 기업고객센터 1588-0114

AX Platform Company kt

서비스 소개

폐쇄망·보안 규정 때문에
외부 클라우드 AI는
검토 단계에서 막혀요.

데이터는 기관망 안에 그대로

질문·문서·답변의 모든 처리가 기관 네트워크 안에서 끝납니다.
행정 데이터와 AI 연산이 밖으로 나가지 않습니다.

자체 GPU 인프라는
수급 기간도 길고
비용이 부담돼요.

서버 · 모델 · 플랫폼 일체형

국산 NPU 서버와 KT AI 모델, 운영 플랫폼을
한 패키지로 제공해 별도 구축 없이 설치 당일 바로 사용합니다.

쌓이는 내부 문서를
찾고 활용하는 일은
여전히 직접 해요.

내부문서 기반 답변 제공 (RAG)

규정·지침·업무 매뉴얼을 올려두면 직원 누구나 묻고 출처와 함께 답을 받습니다.

국산 NPU 서버에 KT AI 모델을 탑재해
기관 안에 설치하는 온프레미스 구독형 AI 서버

이런 기관에 적합해요

01

보안 규정 때문에 외부 AI를 못 쓰는 기관

폐쇄망, 망분리 환경에서도 생성형 AI 도입

02

GPU 구축이 부담되는 기관

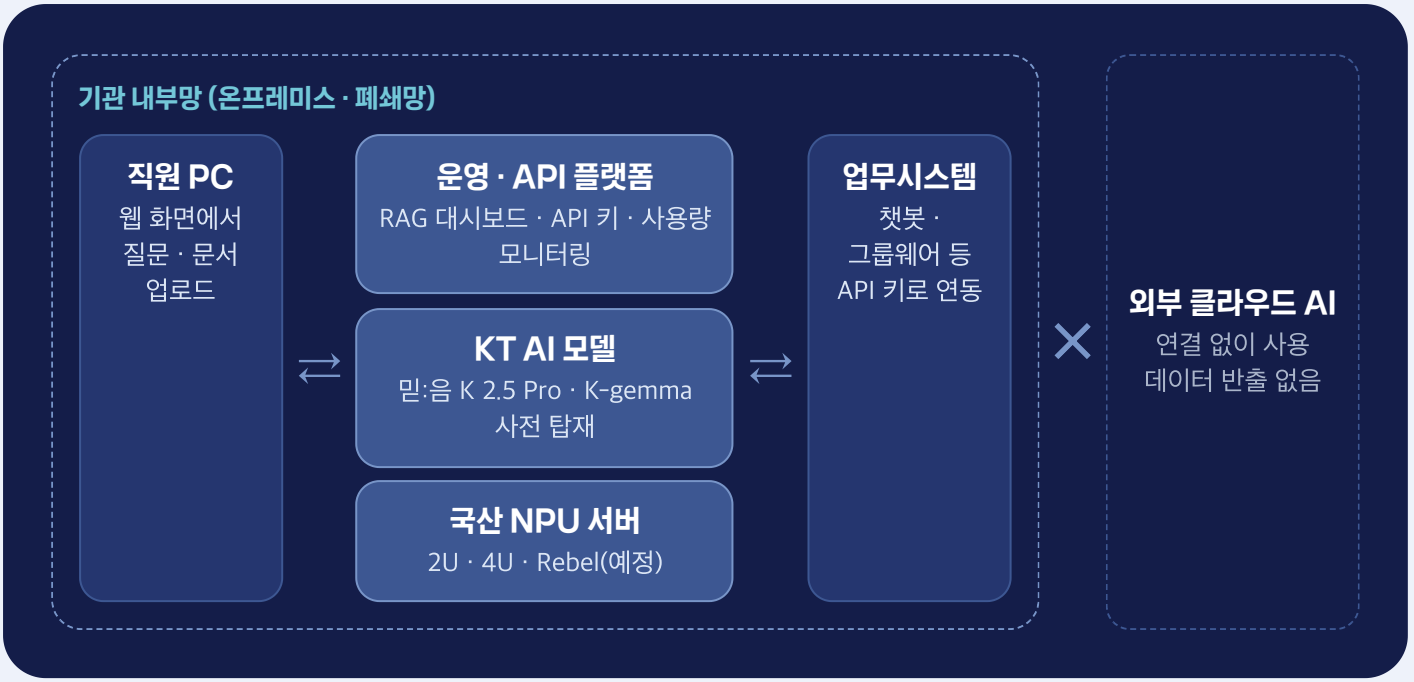
GPU 수급 대기, 초기 구축비 없이 월납으로 시작

03

국산 AI를 우선으로 검토하는 기관

국산 NPU와 KT AI모델을 묶은 소버린 AI 패키지

서비스 구성도



· RAG(Retrieval-Augmented Generation) : 기관 문서를 근거로 답변을 만드는 방식
· 외부 연결 없이 구성 가능하며 적용 범위는 도입 구성에 따라 협의

서비스 유형

구분	2U 서버	4U 서버	Rebel 서버 (예정)
구성	NPU 카드 4장	NPU 카드 8장	차세대 Rebel 카드 탑재
동시 사용	약 44명	약 88명	500B급 대형 모델 대응
타깃고객	부서 · PoC 단위 도입	기관 전체 도입 · 여러 모델 동시 운영	대형 모델 수요 기관

· 조직 규모에 맞춰 2U → 4U → Rebel 순으로 확장 (스케일업 · 스케일아웃)

차별화 요소

데이터 주권 확보
질문 · 문서 · 답변의 모든 처리가 기관망 안에서 끝남

예측 가능한 구독 요금
토큰 사용량과 관계없이 요금 동일
설치부터 유지보수까지 월납

저전력 국산 NPU
동급 GPU 서버 대비
전력 25% 절감 (KT 산정)

기대효과

외부로 나가지 않는 AI로 보안 규정이 엄격한 기관도 생성형 AI 도입

보안성

직원 PC
업무망에서
질문 · 문서 업로드

›

기관 내부 · KT NPU LLM Station
문서 저장 · AI 연산 · 답변 생성
모두 기관 네트워크 안에서 처리

×

외부 클라우드
데이터가
밖으로 나가지 않음

서버 · 모델 · 플랫폼 일체형으로 설치 당일 바로 사용

편의성

1 **문서 등록**
규정 · 지침

›

2 **질문**
모델 선택 후

›

3 **답변 확인**
출처 함께
표시

›

4 **시스템 연동**
API 키 발급

›

5 **운영 관리**
사용량 확인

일반 GPU 서버 구축 약 3개월 → 설치만으로 바로 사용 (Plug & Play)

사용량이 늘어도 요금이 같은 예측 가능한 비용

경제성

요금 방식
월납 구독
설치부터 3년 유지보수까지
하나의 월납계약

전력
-25%
동급 GPU 서버 대비
(14B 모델 기준)

수치는 KT 산정 기준

보안 · 구축 · 비용, 생성형 AI 도입의 세 가지 장벽을 한 번에 제거
폐쇄망 기관도 즉시 AI 활용

활용 예시

1 문서 질의 응답

규정·지침·민원 기준 즉시 확인

예 : 「관내 출장 여비 지급 기준은?」

· 내부 규정·매뉴얼 근거로 답변
· 참고 조문 출처 함께 표시

2 업무시스템 연동

기존 챗봇에 AI 연결

API키로 기존 시스템에 AI 기능 연결

· 민원 챗봇, 그룹웨어, 내부포털에
· AI 기능 연동 가능

3 현장 음성

무전 통신 결합 AI

현장 음성 질의에 폐쇄망 AI로 응답

· 현장 음성 질의에 폐쇄망 AI로 응답
· 데이터는 내부망 밖으로 나가지 않음

4 영상 분석

CCTV 이상 상황 탐지

화재, 쓰러짐 등 이상 상황 탐지

· 이상 상황 탐지와 상황 브리핑 생성
· 기존 CCTV 에 연결

자주 묻는 질문

Q. 외부 클라우드 연결 없이 사용할 수 있나요?

네. 서버·AI 모델·운영 플랫폼이 모두 내부망에 설치되어 외부 클라우드 연결 없이 폐쇄망에서 사용합니다.
질문·문서·답변의 모든 처리가 내부 네트워크 안에서 끝나 데이터가 외부로 나가지 않습니다.

Q. 어떤 AI 모델이 탑재되나요?

KT AI 모델 믿:음 K 2.5 Pro 또는 K-gemma가 사전 탑재되어,
별도 설치 없이 폐쇄망 안에서 바로 사용할 수 있습니다.

요금 안내

월납 구독형 : 설치부터 3년 유지보수까지 하나의 월납계약
서버 규모(2U · 4U)와 도입 구성에 따라 견적 협의