deepseek AI 서버

ZDS-671B

Video Memory	768GB
Computing Power	FP8 10560 TFLOPS
Memory	1TB
Throughput (Total tokens/s)	1400
Concurrent Users (Simultaneous Access)	100

ZDS-70B

Video Memory	192GB
Computing Power	BF16 1320 TFLOPS
Memory	512GB
Throughput (Total tokens/s)	2100
Concurrent Users (Simultaneous Access)	260

ZDS-32B

Video Memory	96GB
Computing Power	BF16 660 TFLOPS
Memory	256GB
Throughput (Total tokens/s)	2100
Concurrent Users (Simultaneous Access)	260

deepseek On-Premise Server 기능별 활용 제안

고성능 챗봇
GPT-4 수준의 자연어 대화

문서 요약 및 생성
보고서, 이메일, 회의록 자동화

전문 Q&A (RAG)
사내 문서 기반 질의응답 시스템

코드 생성 및 분석
개발자 도우미, 코드 설명

음성 챗봇
Whisper + TTS 연동, AI 스피커 구현

이미지 생성
Stable Diffusion 프롬프트 생성

LoRA 파인튜닝
특정 업무에 특화된 사내 모델 구축

완전 오프라인 지원
인터넷 없이 자체 AI 운용 가능

deepseek On-Premise Server 업종 별 활용 예시

금융/보험
오프라인 챗봇, 문서 요약, 내부 규정 Q&A

공공/국방
내부망 AI 비서, 민원 자동화, 음성 보조

의료/제약
EMR 요약, 의료 Q&A, 환자 상담 자동화

제조/산업
매뉴얼 요약, 고장 진단, 음성 지시 시스템

통역/번역
실시간 다국어 통역 / 번역

ZETACUBE™
The Solution Leader of DePIN & AI

T +82-2-569-2227 F +82-2-2088-4655 E allen@zetacube.net
OFFICE 14786 경기도 부천시 소사구 양지로 237 광양프런티어밸리 5차 지식산업센터 1078,1079호
SHOWROOM 06720 서울시 서초구 효령로 304 국제전자센터 30호
DATACENTER 고양1센터 / 광주1센터 / 부천1센터 / 부천2센터 / 서울1센터

zetacube.net nanodc.info filsogood.io