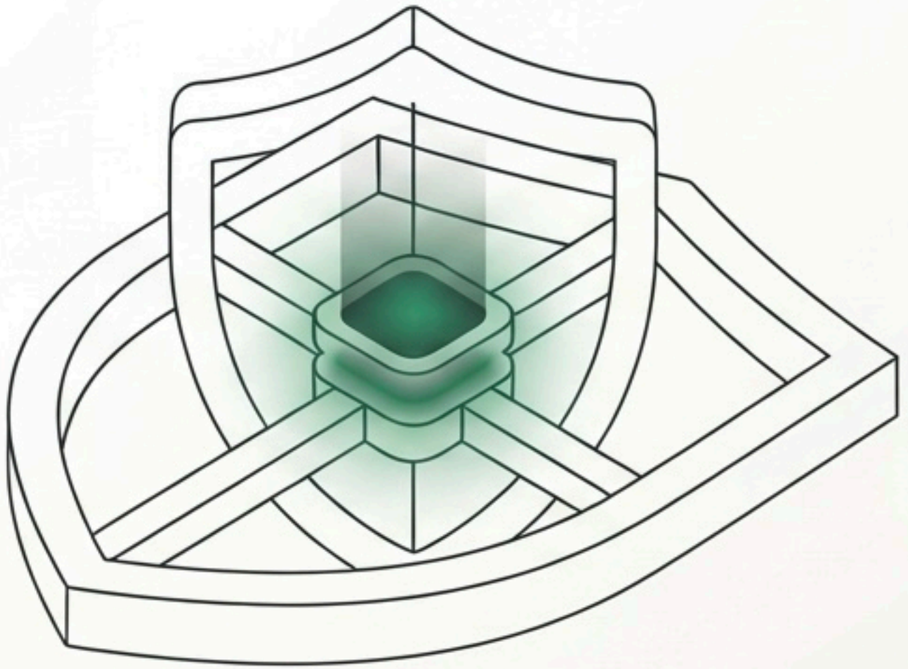
GovOn



온프레미스 AI 기반 민원 처리 인프라, GovOn

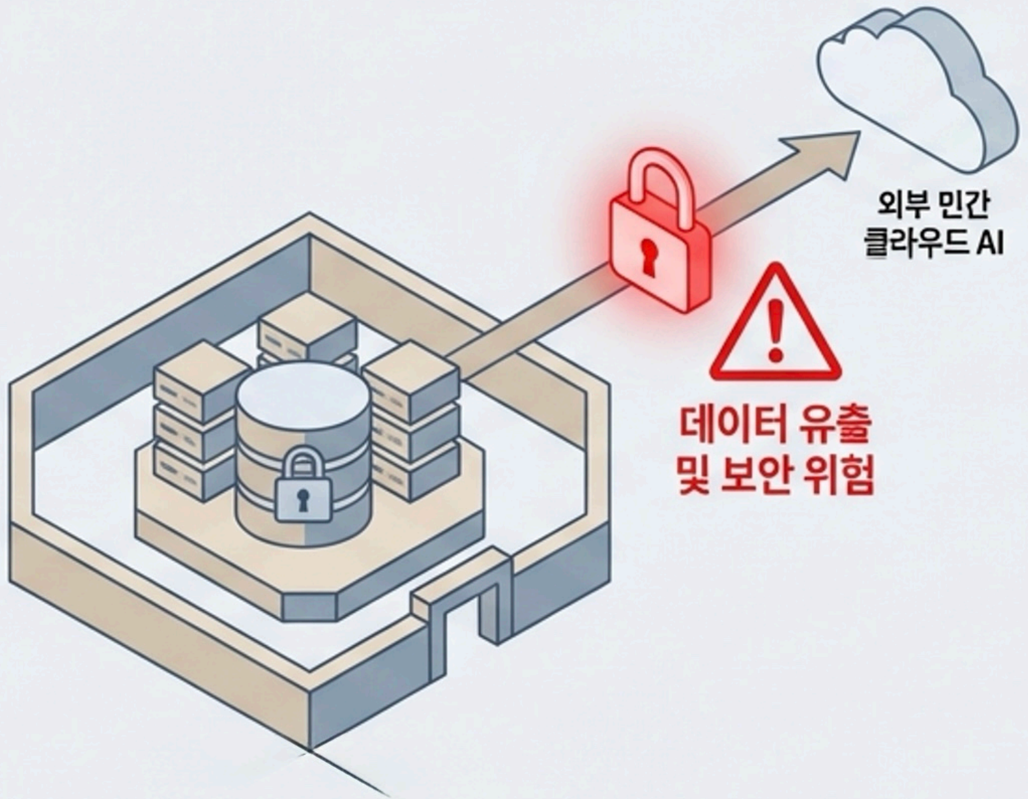
일선 공무원의 업무 부담 최소화 및 국가 정보 보안의 완벽한 보장

동아대학교 DONG-A UNIVERSITY 동아대학교 SW 현장미러형 연계 프로젝트 SW

보안과 효율의 딜레마, 온프레미스 AI로 돌파하다

AS-IS: 공공 행정망의 한계

민감 정보 유출 우려로 외부 민간 클라우드 AI 도입 제약



TO-BE: GovOn Architecture

온프레미스 LLM 구축 (On-Premise Core)

망 분리된 내부 서버에서 독립적으로 구동되는 완벽한 데이터 통제 환경 (EXAONE-Deep-7.8B 활용).

파인튜닝 (Fine-Tuning)

정부 행정 규정과 민원 데이터를 집중 학습하여 실무 공무원의 업무 전문성을 극대화.

양자화 (Quantization)

한정된 내부 서버 자원에서도 빠르고 가볍게 연산되도록 모델 경량화 및 최적화 달성.

검색 증강 생성 (RAG)

폐쇄망 환경의 정보 단절을 극복하고, 최신 법령 및 유사 민원 사례를 실시간으로 참조(FAISS 기반).

부처 간 데이터 사일로

부처별 분절된 시스템으로 인한
정책 판단 지연 및
종합적 대응 한계.

국가인공지능전략위원회,
「대한민국 인공지능 행동계획(2026~2028)」 p.89

공무원의 위험 회피 문화

규제와 정보 유출 리스크로 인해
실무진이 새로운 AI 기술 도입을
선제적으로 주저하는 구조적 환경.

국가인공지능전략위원회,
「대한민국 인공지능 행동계획(2026~2028)」 p.94

행정망 해킹 사고 현실화

보안인증 기업과 국가 행정망
대상의 고도화된 타겟 해킹 등
심각한 사이버 위협 증가.

국가인공지능전략위원회,
「대한민국 인공지능 행동계획(2026~2028)」 p.12

GovOn: 온프레미스 AI 기반 행정 혁신 프로젝트 진행 현황

프로젝트 마일스톤 (Project Milestones)

M1 & M2: 기획 및 핵심 MVP (100% 완료)

시스템 아키텍처 설계와 EXAONE-Deep 모델 파인튜닝 및 양자화 배포를 완료했습니다.

M3: 시스템 고도화 (46% 진행 중)

FAISS 기반 RAG 파이프라인과 멀티턴 에이전트 시스템 최적화를 진행하고 있습니다.

M4: 테스트 및 최종 발표 (예정)

통합 테스트 및 사용자 수용 테스트(UAT)를 통한 최종 검증을 앞두고 있습니다.

핵심 기술 스택 및 개발 지표 (Core Tech Stack & Dev Metrics)

AI Model Stack
EXAONE-Deep, QLoRA, AWQ, vLLM

RAG Pipeline
FAISS, BM25, MultiIndexManager

Agent System
Session, Multi-turn, Context, Streaming

Automated DevOps (GitHub Actions)
테스트부터 빌드, Docker 배포까지 전 과정을 자동화하여 개발 효율성을 극대화했습니다.

Frontend UI
Next.js, Auth, Figma MCP

Issue Stats
총 95개 중 45개 완료
안정적인 개발 속도 유지

주요 기술 스택 EXAONE-Deep | FastAPI | vLLM | FAISS | GitHub Actions | Next.js | Docker

GovOn

차세대 행정 인프라의 새로운 표준

행정망의 사일로 현상과 클라우드 AI의 보안 취약성을
동시에 극복한 최적화된 온프레미스 AI.

강력한 보안 + 뛰어난 연산 효율 = 실무 공무원 업무 생산성 혁신