E EDUMGT AI Agent Studio

Company Profile · 2026

A X

기업의 AI 전환(AX)을 실행 가능한 구조로 만듭니다

LLM · RAG · AI Agent 구축부터 멀티 에이전트 기반 완전 자동화 AI Ops까지,
내부망 · 클라우드 · 하이브리드 어떤 환경에서도 운영 가능한 AI를 설계합니다.

(주)에듀엠지티 · EDUMGT
Fintech · 멀티 LLM 최적화 · 코딩 자동화 · DX · AX
www.edumgt.co.kr

25Y+	113+	4
EXPERIENCE	PROJECTS	AI DEPLOYMENT MODELS

기업의 DX · AX를 실행 가능한 구조로 만드는
AI Agent Engineering Partner

EDUMGT는 교육·공공·엔터프라이즈 분야에서 쌓은 25년 노하우를 기반으로, 단순한 데모가 아닌 **실제 운영 가능한 AI 서비스**를 설계합니다. 전략 수립부터 구축, 배포, 운영 최적화까지 일관된 실행 체계를 제공합니다.

EXPERIENCE
25Y+
AI · Cloud · LMS 중심 실무 구축 경험 (2001 - 2026)

DELIVERY
113+
프로젝트 이력과 운영 중심 산출물 납품

FOCUS
End-to-End
기획 · 개발 · 배포 · 운영 최적화까지 연결

AX · AI Transformation
LLM · RAG · AI Agent를 업무 프로세스에 직접 연결해 사람이 하던 판단과 실행을 AI가 수행하도록 바꾸는 전환. 승인·감사·가드레일을 포함한 운영 모델까지 함께 설계합니다.

DX · Digital Transformation
레거시 분석 → 현대화 → 모듈화 → 클라우드 운영까지 단계별 로드맵. AX가 안착할 수 있는 데이터·API·인프라 기반을 만듭니다.

왜 EDUMGT인가

도메인 이해 + 빠른 실행력
교육·공공·기업 시스템에 대한 깊은 도메인 이해를 바탕으로 PoC → MVP → 운영 안착까지 빠르게 연결합니다.

업무 특성에 맞는 AI 모델·LLM 최적화
Claude · GPT · Gemini · 오픈소스 LLM 가운데 업무에 맞는 모델을 선택·전환하고, Claude Code · Cursor 등 AI 코딩 도구와 MCP 서버 개발로 개발 전 주기를 가속합니다.

보안 경계를 존중하는 아키텍처
내부망(Air-Gap) · 클라우드 · 하이브리드, 데이터 위치와 규제에 맞는 구축 방식을 제시합니다.

자율 운영까지 설계하는 AI Ops
에이전트 하네스와 멀티 에이전트 크루가 계획·구현·검증·배포를 스스로 수행하는 완전 자동화 모델을 보유합니다.

공개된 구현으로 검증
핀테크 AI(투자 분석 · 금융 RAG · 회계·공시 자동화)와 Harbor 폐쇄망 실습을 GitHub에 공개하고 있습니다.

AI Agent Studio · RAG · Vector DB · Workflow · AWS · Azure · Cloud Native · LMS · LCMS · SCORM · xAPI · DX / AX Consulting

02 · ENTERPRISE AI DEPLOYMENT MODELS

내부망 · 클라우드 · 하이브리드 · AI Ops 자동화
LLM · RAG · AI Agent 구축 모델

데이터 위치와 보안 경계에 맞는 구축 방식을 선택합니다. 네 가지 참조 아키텍처 모두 모델 추론 → 지식 검색(RAG) → 업무 도구를 쓰는 Agent → 권한·승인·감사 흐름을 하나의 운영 구조로 연결합니다.

01 · ON-PREMISE

내부망 (폐쇄망) 구축

기업 내부망에서 모델, 사내 지식, 업무 도구를 연결하는 참조 아키텍처

폐쇄망 공급망

외부 검증·반입 승인된 이미지·모델 파일 반입

Harbor + 사내 CA 내부 레지스트리 · TLS 신뢰 체인

Robot Account K8s 최소 권한 Pull

내부 AI 플랫폼

Kubernetes 사내 클러스터 배포

GPU + vLLM LLM 추론 API

NFS / 모델 저장소

기업 업무 서비스

RAG 파이프라인 수집·청킹·임베딩·벡터 검색

AI Agent 사내 도구 호출 · 승인 흐름

02 · CLOUD

클라우드 구축

관리형 인프라 위에 LLM, RAG, Agent와 운영 계층을 연결

접속·보안

사용자 포털 업무별 접속 화면

ID·권한 접근 제어와 인증

API Gateway 라우팅 · 사용량 제어

AI 실행

클라우드 런타임 컨테이너·서버리스

LLM 서비스 관리형 또는 자체 호스팅

RAG·Vector DB 문서 검색 · 근거 연결

업무·운영

AI Agent 도구 호출 · 승인 흐름

업무 시스템 API·데이터 연계

관측·감사 품질 · 비용 · 실행 이력

03 · HYBRID

하이브리드 구축

민감 데이터는 내부망에 두고, 승인된 요청만 클라우드 AI 자원과 연결

내부망 데이터

사내 문서·DB 원본과 접근 권한 유지

내부 검색·RAG 권한별 문서 검색

내부 LLM 민감 요청의 로컬 추론

보안 연결

정책 라우터 요청·데이터 분류

비공개 연결 VPN · 전용 연결 경로

승인·감사 반출 승인과 호출 기록

클라우드 확장

클라우드 LLM 허용된 작업의 추론 확장

AI Agent 내·외부 도구 오케스트레이션

04 · AI OPS AUTOMATION

완전 자동화 AI Ops

에이전트 하네스가 이벤트를 받고, 역할 에이전트 크루가 계획·구현·검증·배포까지 수행

에이전트 하네스

이벤트·트리거 Git push · 알람 · 스케줄 · 채팅

Agent Harness 도구 루프 · 스킬 · 장기 메모리

정책·가드레일 비밀정보 분리 · 실행 한도

멀티 에이전트 크루

Planner · Builder · Reviewer · Deploy 역할 분담과 되돌림 루프

AWS 배포·운영

S3 → CloudFront 동기화 · 캐시 무효화

CloudWatch 환류 자가 복구

AI Sec Ops · Air-Gap

외부 반입 승인, 이미지 검증, 내부 인증서·접근 권한 관리 등 폐쇄망 운영 절차를 모델링합니다. Harbor 폐쇄망 반입·vLLM 배포 템플릿은 공개 실습 저장소(harbor-lab)로 제공되며, RAG·Agent 연결은 고객 환경에 맞춰 설계합니다.

03 · FULLY AUTOMATED AI OPS

사람이 지시하면, 에이전트 크루가 배포까지 끝냅니다

Hermes 방식의 상시 실행 에이전트 하네스가 Git push·모니터링 알람·스케줄 이벤트를 받고, CrewAI 방식의 역할 에이전트 크루가 계획·구현·검증·배포를 수행합니다. www.edumgt.co.kr 정적 사이트의 S3·CloudFront 배포를 기준으로 실제 운영 중인 참조 구성입니다.

01 에이전트 하네스 Hermes 방식

- **이벤트 · 트리거**
  Git push · 알람 · 스케줄 · 채팅 지시
- **Agent Harness**
  도구 루프 · 스킬 · 장기 메모리 · 세션 관리
- **정책 · 가드레일**
  비밀정보 분리 · 실행 한도 · 승인 규칙

→

02 멀티 에이전트 크루 CrewAI 방식

- **Planner Agent**
  작업 분해와 완료 기준 정의
- **Builder Agent**
  코드·콘텐츠 수정과 사이트 빌드
- **Reviewer · QA Agent**
  테스트·보안 점검 후 되돌림 루프

→

03 AWS 배포 · 운영 www.edumgt.co.kr

- **GitHub → Deploy Agent**
  커밋·PR 기록 후 승인된 산출물 배포
- **Amazon S3 → CloudFront**
  정적 원본 동기화와 캐시 무효화
- **모니터링 · 피드백**
  CloudWatch·로그를 하네스로 환류해 자가 복구

이벤트 기반 자율 실행
사람의 개입 없이 트리거만으로 작업이 시작되고, 완료 기준을 만족할 때까지 반복합니다.

역할 분리로 품질 확보
Planner · Builder · Reviewer가 서로 검증하며, QA 실패 시 자동 되돌림 루프가 동작합니다.

정책 안의 자율성
비밀정보 분리, 실행 한도, 승인 규칙을 하네스에 내장해 통제 가능한 자동화를 구현합니다.

관측 → 자가 복구
CloudWatch·로그 지표가 하네스로 돌아와 장애를 감지하고 스스로 복구를 수행합니다.

사내 RAG 생성형 CrewAI, 서버 한 대에 가상 직원 여러 명

PM·PD의 지시를 받은 역할별 AI 에이전트가 서로 의사소통하며 업무를 분배하고, 사내 문서 기반 RAG와 외부 리서치 데이터를 근거로 결과물을 생성합니다. 사무실 PC(Mac mini) 한 대에 구성하는 표준 구축 패키지입니다.

A · 클라우드 (AWS)

외부 데이터 수급 크롤러 · 리서치 서버

사내 IP는 크롤링 중 차단될 수 있어 클라우드 분리 운영. 귀사 명의 AWS 계정, 사용량 비례 과금.

크롤러 · 수집
웹 · 공시 · 뉴스 · API 등 허용된 범위 수집

정제 · 구조화
중복 제거 · 분류 · 요약 후 사내 RAG로 전달

수집 범위 정책
robots · 로그인 · API 제한 준수

→

B · 사내 OnPrem 서버 (Mac mini 1기 = CrewAI 환경 1세트)

사내 RAG 생성형 CrewAI 멀티 에이전트 환경

PM · PD 지시
업무 목표 · 완료 기준 · 우선순위를 채팅으로 전달

생성 결과물
보고서 · 기획안 · 콘텐츠 · 코드 · 데이터 분석 결과

↓ ↑

CrewAI 에이전트 크루 · 가상 직원
역할 분배 · 에이전트 간 의사소통 · 되돌림 루프

- **Planner** — 작업 분해 · 담당 배정
- **Researcher** — 사내 RAG · 외부 데이터 조사
- **Writer · Builder** — 초안 · 산출물 생성
- **Reviewer · QA** — 검증 · 수정 요청

↕

사내 문서 · DB
규정 · 보고서 · 회의록 · 업무 데이터

임베딩 · Vector DB
청킹 · 임베딩 · 권한별 검색

근거 기반 응답
출처가 포함된 검색 결과를 에이전트에 제공

↔

C · LLM 연동 계층

기존 LLM 그대로 연결 OpenRouter 다중 모델 라우팅

귀사에서 이미 쓰는 계정·API를 연동하고, 에이전트별로 모델을 선택·전환합니다. API 사용료는 귀사 계정 과금.

- ChatGPT (OpenAI)
- Claude (Anthropic)
- Gemini (Google)
- OpenRouter 다중 라우팅

모델 선택 · 전환
역할·비용·품질에 따라 에이전트마다 다른 모델 사용

키 분리 · 사용량 통제
비밀정보 분리 · 호출 한도 · 로그

표준 구축 일정 · 3주 + 하자보수 2개월

1일차 · 설치	1주차 · 경과보고	2주차 · 실사용	3주차 · 인수
Mac mini에 CrewAI SW 설치, 크롤러·정제 SW 클라우드 셋업	에이전트 역할·업무 흐름 정의, 경과보고 기반 파인튜닝	실사용 시작, 프롬프트·분배 규칙 조정	안정화·검수 후 2개월 하자보수. 매주 1회 방문 회의

표준 구축 패키지 비용

₩3,000,000 부가세 별도 · 3주 구축 용역

CrewAI SW 설치·구성
2,000,000원 · 서버 1기당

크롤러·정제 SW 셋업
1,000,000원 · 클라우드

고객 직접 구매
Mac mini 1기 약 400~500만원

고객 사용량 과금
AWS 월 20~30만원 · LLM API

업무 특성에 맞는 AI 모델·LLM 최적화로 코딩 자동화와 AI 업무 응용, 조직의 생산성을 근본적으로 바꿉니다

MULTI-LLM · AI CODING

업무 맞춤 LLM 기반 AI 코딩 서비스

Claude Code · Cursor · Copilot 등 AI 코딩 도구와 MCP 서버 개발로 코드 생성부터 리뷰·배포까지 개발 전 주기를 AI로 가속합니다. 고객의 바이브코딩 MVP를 코딩 완결성 확보 · 서비스 고도화 · 클라우드 배포까지 완성합니다.

- 업무 특성에 맞는 LLM (Claude · GPT · Gemini · 오픈소스) 연동 코드 생성·자동완성 파이프라인
- 코드 리뷰 자동화 · 버그 탐지 · 테스트 코드 생성
- Claude Code · Cursor · MCP 서버 개발환경 구축
- 레거시 코드 분석·문서화·리팩터링 자동화

MULTI-LLM · BIZ CONSULTING

AI 업무 응용 컨설팅 (AX)

조직의 핵심 업무 프로세스를 진단하고 업무 특성에 맞는 AI 모델·LLM을 선택·최적화해 실질적 생산성 향상 전략을 수립·실행합니다.

- 업무 프로세스 AI 전환 로드맵 수립 (As-Is → To-Be)
- 보고서 작성·데이터 분석·의사결정 지원 자동화
- 업무 맞춤 LLM 기반 사내 지식관리 시스템 설계·구축
- AI 거버넌스·보안 정책·임직원 역량 내재화 교육

FINTECH × AI · GITHUB

핀테크 AI 프로젝트

금융 데이터에서 실행 가능한 AI까지. 공개 GitHub 저장소로 구현 역량을 검증합니다.

- **투자 분석 · 퀀트** KRX 데이터 · QuantConnect LEAN 백테스트 · 포트폴리오 최적화
- **금융 AI Agent · RAG** ReAct 질의응답 · Qdrant 문서 검색 · 대화형 분석
- **회계 · 전자공시 자동화** 증빙 OCR · 계정 분류 · 이상거래 탐지 · OpenDART 연동

investment-analysis · lumina-invest · stock-kms-portal · vdp-agent

Core AI · LLM

Claude · GPT · Gemini · OpenRouter 다중 LLM 라우팅 · Claude Code · MCP · Amazon Bedrock · Azure OpenAI · Ollama · vLLM (온프레미스) · HuggingFace OSS LLM

AI Agent · RAG

LangChain / LangGraph · LlamaIndex · AutoGen · CrewAI · Qdrant · Pinecone · Weaviate · pgvector · ChromaDB · FAISS · Cohere Rerank

MLOps · LLMOps

MLflow · LangSmith · Weights & Biases · SageMaker · Kubeflow · Arize Phoenix · Helicone · Prometheus · Grafana

Cloud · Platform

AWS Lambda · Step Functions · S3 · API GW · ECS/EKS · Glue · Athena · Redshift · Azure DevOps · AKS · Kafka · Kinesis · Airflow · Docker · Kubernetes · Harbor

06 · AI PROJECT DELIVERY FRAMEWORK

설계부터 운영까지 이어지는 4단계 실행 프레임워크

AI Agent, RAG, LLMOps, DevOps, MLOps를 하나의 실행 체계로 묶어 기업의 AX · DX 전환을 추진합니다.

DISCOVER 01

업무·데이터 진단

현행 프로세스, 문서, DB, API, 인프라를 분석해 AI 적용 범위와 우선순위를 정의합니다.

As-Is 분석 · 데이터 맵 · KPI 정렬

DESIGN 02

AI Agent · RAG 설계

프롬프트 전략, 툴 호출, 권한 흐름, 벡터DB, 검색 구조를 설계해 신뢰 가능한 응답 체계를 만듭니다.

Prompt · RAG · Guardrail

DEPLOY 03

클라우드 구현·배포

AWS/Azure 기반 인프라, CI/CD, API, 권한·감사 체계를 연결해 엔터프라이즈 운영에 맞게 배포합니다.

AWS · CI/CD · Security

OPTIMIZE 04

운영 최적화

품질, 비용, 응답속도, 장애 지표를 모니터링하며 DevOps·MLOps·AIOps 관점으로 지속 개선합니다.

Observability · Cost · Feedback Loop

핵심 산출물과 운영 기준

핵심 산출물	AI PoC, RAG 지식베이스, Agent 워크플로우, 운영 대시보드, 클라우드 배포 자동화
적용 영역	교육 플랫폼, 기업 내부 업무자동화, 대시보드, 클라우드 전환, AI 서비스 구축
운영 기준	보안, 추적성, 비용 통제, 확장성, 현장 안착까지 고려한 실무형 Delivery
전환 경로	POC → MVP → 운영 안착 · 성능·보안·품질 점검 · 내부 역량 강화 교육

CONTACT

프로젝트 문의

kimdypm@gmail.com

AI Agent · RAG · 클라우드 구축 · DX/AX 전환 등 어떤 주제든 편하게 메일 주세요. 빠르게 검토 후 연락드리겠습니다.

(주)에듀엠지티 · Since 2010
서울 성동구 연무장5가길 현대테라스타워 E동 1206호
www.edumgt.co.kr · github.com/edumgt