

Agentic AI 시대,

Agentic AI로 가는 길: AI Ready 데이터와 산업형 AI 에이전트 전략

CrowdWorks

양 수 열 CTO



일반적인 기업 AI 도입 절차

성공적인 AI 도입을 위한 End-to-End 프로세스

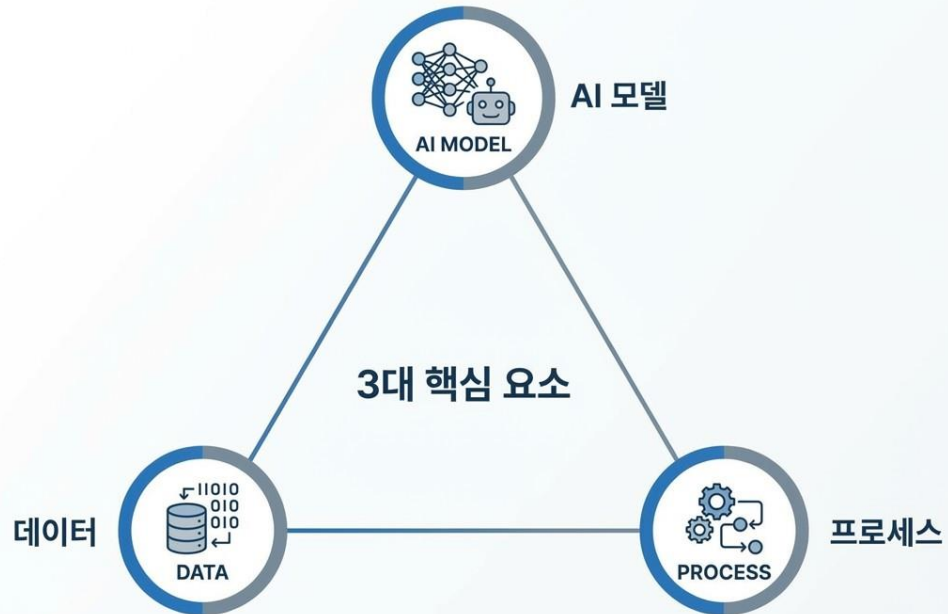
Process Overview



AI의 3대 핵심 요소

성공적인 AI 도입을 위한 DATA, PROCESS, MODEL의 균형

Strategic Framework



"AI 모델은 매년 바뀌지만,
축적된 데이터와 견고한 프로세스는 기업의 영구적 자산이 됩니다"

DATA (기반)

AI의 원료이자 성능의 80%를 결정짓는 핵심 자산입니다.
양질의 데이터 확보(Garbage In, Garbage Out)가 선행되어야 하며,
최신성과 정확성이 보장되어야 합니다.

PROCESS (지속성)

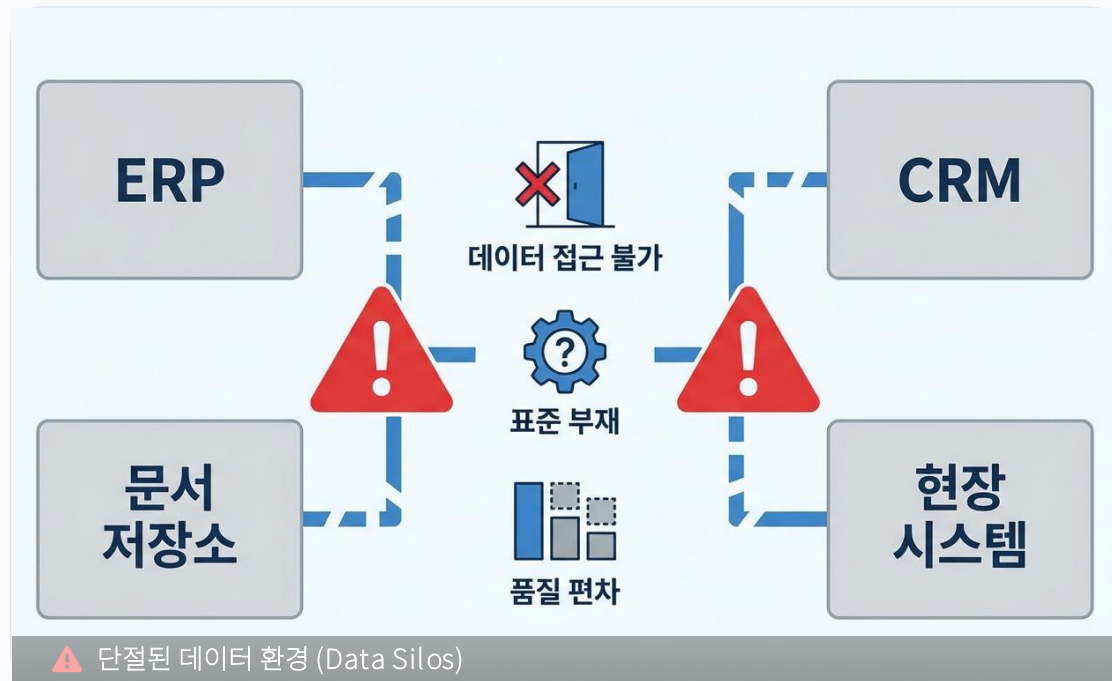
데이터가 수집되고 AI가 학습하는 일관된 흐름입니다.
단발성 프로젝트가 아닌, 지속 가능한 운영 체계(MLOps, DataOps)
를 통해 품질을 유지합니다.

AI MODEL (가속화)

비즈니스 가치를 창출하는 도구이자 엔진입니다.
기술 발전 속도에 맞춰 계속 변화하고 진화하므로, 유연하게 교체 가능
한 구조가 필요합니다.

사내 데이터를 AI에서 쓰려 할 때의 문제점

성공적인 AI 도입을 가로막는 현실적인 데이터 장벽들



"기업 내 데이터는 존재하지만, AI가 학습 가능한 형태(AI-Ready)로 준비되어 있지 않습니다."

물리적으로 분산된 저장소와 일관성 없는 데이터 표준은 AI 모델 개발 시간의 80%를 데이터 전처리에 쏟게 만듭니다."



데이터 사일로 (Data Silos)

Critical

ERP, CRM, 레거시 시스템, 개별 PC 등에 데이터가 파편화되어 있어 통합적인 AI 학습 데이터셋 구성이 불가능함.



표준 및 메타데이터 부재

Critical

같은 의미의 데이터도 시스템마다 컬럼명이나 코드 체계가 달라, 이를 사람이 일일이 매핑해야 하는 비효율 발생.



데이터 품질 편차 (Quality)

Major

결측치(Null), 중복 데이터, 오입력 등이 정제되지 않은 상태로 방치되어 AI 모델의 성능 저하를 유발.



정답지(Label) 부족

Major

AI 학습에 필수적인 라벨링 데이터(예: 불량 이미지 분류값)가 없거나, 전문가의 암묵지에 의존해 기록되지 않음.



접근 권한 및 보안 이슈

Moderate

복잡한 결재 프로세스와 개인정보/보안 규제로 인해 실제 데이터를 확보하여 샌드박스 환경으로 옮기는 데 수주가 소요됨.

현황 1: ERP 중심 프로세스와 데이터 자산 맵

기업의 핵심 비즈니스 프로세스와 여기서 파생되는 정형/비정형 데이터 현황

As-Is Analysis

🏭 Core Business Process (ERP)



1. 수주 (Order)

고객 주문 접수, 계약 체결, 주문 생성



2. 생산 (Production)

자재 투입, 공정 관리, 품질 검사(QA)



3. 출하 (Shipment)

재고 불출, 배송, 송장 발행



4. 정산 (Settlement)

세금계산서 발행, 수금 확인, 회계 처리

🗃️ 정형 데이터 (Structured Data)



ERP 원장



CRM 고객정보



SCM 재고현황



MES 공정실적



재무/회계 데이터

📌 특징: 스키마가 고정되어 있고, RDBMS에 저장됨. AI 학습 용이성 높음.

📁 비정형 데이터 (Unstructured Data)



계약서/시방서 (PDF)



이메일/메신저



IoT 센서 로그 (JSON)



제품 불량 이미지



VOC 음성/텍스트

📌 특징: 형태가 다양하고 용량이 큼. 별도의 전처리(OCR, STT 등) 필요.



핵심 인사이트: AI 모델은 기술 트렌드에 따라 계속 바뀌지만, 기업의 데이터는 지속적으로 축적되는 자산입니다.

따라서 데이터 표준화 및 연계 구조 확립이 우선되어야 합니다.

현황 2: 정형/비정형 파이프라인과 거버넌스

데이터 유형별 최적화된 처리 파이프라인과 통합 품질/보안 관리 체계

Pipeline Architecture

정형 데이터 파이프라인



1. 수집 (CDC / Batch)

ERP, CRM의 변경 데이터 실시간 캡처 및 배치 수집



2. 정제 및 정합성 검사

결측치 처리, 타입 변환, 비즈니스 규칙 검증



3. 데이터 카탈로그 등록

메타데이터 추출, 용어 사전 매핑, 계보(Lineage) 관리



4. Feature Store 적재

모델 학습용 피쳐 생성 및 서빙 (Online/Offline)

비정형 데이터 파이프라인



1. 수집 및 전처리

문서(PDF), 이미지, 로그 수집 및 노이즈 제거



2. 분류 및 정보 추출

OCR, STT, 개체명 인식(NER)을 통한 텍스트화



3. 임베딩 (Vectorization)

텍스트/이미지를 고차원 벡터로 변환 (Embedding Model)



4. Vector DB / RAG 인덱스

의미 검색을 위한 벡터 인덱싱 및 지식 베이스 구축

통합 거버넌스

Unified Governance Layer



보안/프라이버시 (PII 마스킹)



데이터 계약 (Data Contract)



품질 규칙 (Quality Rule)



접근 제어 (RBAC)

현황 3: 모델이 바뀌어도 흔들리지 않는 대응 전략

단기적 모델 성능 개선보다 장기적 데이터 자산화에 집중하는 단계별 로드맵

Strategic Roadmap



데이터 표준화 & 품질 확보

- 전사 데이터 표준 용어 정의
- 결측치/이상치 정제 규칙 수립
- 정형 데이터 스키마 확정



메타데이터 관리 & 카탈로그 구축

- 데이터 계보(Lineage) 시각화
- 비정형 데이터 태깅/분류
- 데이터 자산 검색 포털 오픈



피처 스토어 & 벡터 DB 도입

- 모델용 Feature 재사용성 증대
- RAG를 위한 벡터 인덱스 생성
- Online/Offline 정합성 보장



평가 체계 및 API 추상화

- 모델 교체를 위한 API 추상화
- Golden Set 기반 성능 평가
- 지속적 모니터링 대시보드

핵심 전략 Key Actions



Model-Data 분리
모델 의존성 제거 및 독립적 관리



Schema Versioning
데이터 구조 변경 이력 추적



Benchmark Sets
객관적 성능 평가 기준 데이터 확보

핵심: AI-Ready Data 관리 역량

모델 성능을 결정짓는 고품질 데이터 파이프라인(Data Pipeline) 구축 전략

Core Competency

AI-Ready Data Pipeline



성공적인 운영을 위한 5가지 데이터 관리 원칙



데이터 계약 (Data Contract)

데이터 생산자와 소비자 간의 스키마 및 품질 기준을 명문화하여 변경에 따른 장애 방지



메타데이터 & 계보 (Lineage)

데이터의 생성 시점부터 가공, 소비 되는 전 과정의 이력을 추적하여 신뢰성 확보



품질 SLO (Service Level)

결측치, 타입 오류, 최신성 등 핵심 품질 지표에 대한 목표 수준을 설정하고 모니터링



접근 제어 & 감사 (Audit)

민감 정보(PII) 자동 마스킹 및 역할 기반(RBAC) 접근 제어로 보안 리스크 최소화



비용 가시성 (FinOps)

스토리지 및 컴퓨팅 비용을 파이프라인 단계별로 시각화하여 운영 효율성 제고

모델 변경 시 대응 방법

지속적인 AI 서비스 운영을 위한 MLOps 순환 체계 (Cycle)

Continuous Adaptation



MLOps Feedback Loop

1 변경 감지 (Detection)

데이터 분포 변화(Drift)나 스키마 변경을 실시간 모니터링하여 알림 발송

2 영향 분석 (Analysis)

변경된 데이터가 기존 파이프라인과 모델 성능에 미칠 영향을 자동 분석

3 호환 계층 (Adapter)

즉시 수정이 어려울 경우, 어댑터 패턴을 통해 데이터 입출력 규격을 맞춤

4 재학습 및 배포 (Retrain & Deploy)

최신 데이터로 모델을 재학습하고, 챔피언-챌린저 테스트 후 점진적 배포



불변 레이어 (Invariant Layer)

모델과 프로세스가 바뀌어도 흔들리지 않고 유지되어야 하는 핵심 자산입니다.

데이터 계약

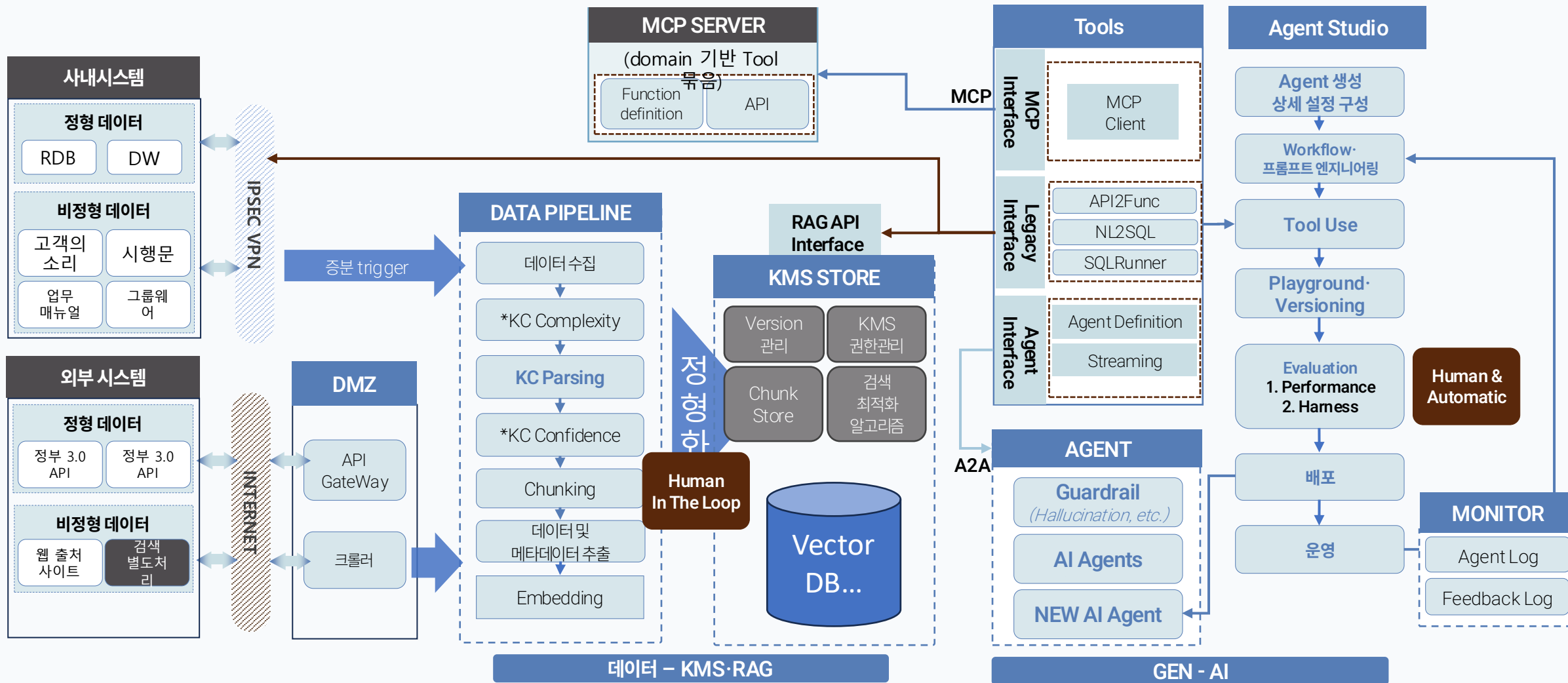
Feature Store

평가 Benchmark

Agentic AI System Architecture

내부 시스템과 외부 시스템의 다양한 데이터를 AI Ready Data로 전환 하여, **최적화된 AI 시스템 지원**

Agentic AI Adaptation



Q&A

발표 내용과 관련하여 궁금한 점을 자유롭게 질문해 주세요.

Interactive Session

Q&A

무엇이든 물어보세요

기업 데이터 전략, 파이프라인 구축, 또는 MLOps 운영 경험에 대해 깊이 있는 논의를 환영합니다.

Join at
slido.com
#3021 671



감사합니다.

기업의 데이터를 자산으로 만드는 여정,
성공적인 AI 도입의 파트너가 되겠습니다.



Presentator

양수열 CrowdWorks CTO



Email

sales@crowdworks.ai

Next Steps

지금 바로 시작할 수 있는 다음 단계

01

현황 진단 워크숍

현재 데이터 인프라와 보유 데이터 자산을 분석하고 AI 준비도를 진단합니다.

02

로드맵 수립

비즈니스 목표에 맞춘 단계별 데이터 구축 및 거버넌스 로드맵을 제안합니다.

03

PoC 설계 및 검증

핵심 과제에 대한 파일럿 프로젝트를 수행하여 가능성을 검증합니다.