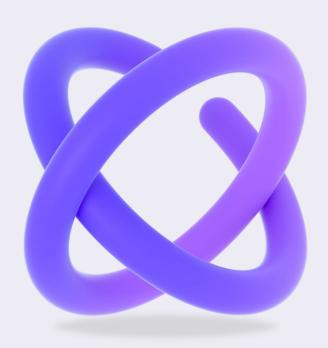
커스터마이징이 가능한 미래형 A.I. 플랫폼 서비스 Clevi

CleviCS, Company Introduction

2025



CONTENTS

Chapter I. Clevi 플랫폼 개요

Chapter II. Clevi 플랫폼 및 모델 소개

Chapter Ⅲ. Clevi 플랫폼 특장점

Chapter IV. 도입 시 고려사항

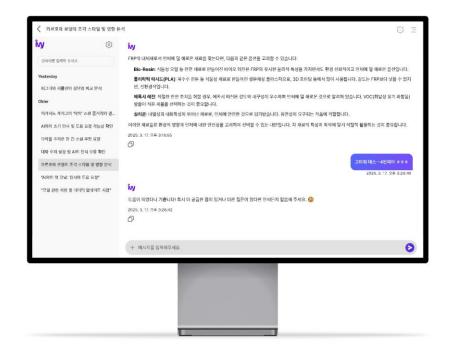
Chapter V. LLM 적용사례





"사용자가 역할을 지정하면 해당 역할을 전문적으로 수행하는 커스터마이징이 가능한 미래형 A.I."

Clevi는 다양한 정보를 실시간으로 학습하고 개선하여 사용자에게 맞춤형 도움을 제공하는 고수준 지능형 어시스턴트입니다.



Clevi는 행동을 정하거나 스케쥴을 지정하면 그에 따라 행동합니다.

반복되는 업무 지시 가능

EX) "매주 월요일마다 스포츠 소식을 정리해서 알려줘"

이벤트 알림 및 행동 지시 가능

EX) "손흥민 선수가 골을 넣으면 알려주고 골 장면을 보여줘"

주요 프로젝트 일정 및 업무 관리

EX) "일요일이 프로젝트 마감이니, 금요일에 비용 내역을 모두 뽑아줘"



새로운 형태의 LLM - 자체 기술인 생각의 사슬 (Chain of Thought, CoT)을 기반으로 추론합니다.

인간과 기계의 대화중 복잡한 문제를 단계별로 추론 과정을 설명하며 푸는 알고리즘으로 LLM의 주요 추론 능력 향상을 위한 기술입니다.

사용자 요청

AI가 사용자에게 응답할 때 주의해야할 점은?

생각 시작

7초동안 고민했습니다. ^

Alright, I need to address the key points Al should be cautious about when responding. First off, accuracy is crucial. I should make sure the information I provide is correct and up-to-date, especially when it comes to facts, figures, and industry specifics.

Transparency is another big one. I should always be clear about what I can and cannot do, and if I'm generating information based on patterns in data rather than concrete facts. I need to make that clear to the user.

Ethical considerations are important too. I must avoid anything that could be harmful or support unethical behavior. This includes being mindful of biases and ensuring my responses are fair and inclusive.

Respecting privacy is essential. I shouldn't ask for or share any personal information without proper consent, and I need to handle any sensitive data with care.

Encouraging responsible use is also key. If my response could have significant real-world implications, I should advise the user to seek professional help or do more research.

Clarity and simplicity in my explanations are important to ensure the user understands what I'm saying. I should avoid jargon unless it's necessary and always check if the user needs further clarification.

Being user-centered means I should tailor my responses to the user's needs and preferences. Whether they prefer formal or casual language, I should adapt accordingly.

Staying professional and respectful is a must, no matter the situation. I need to maintain a polite and considerate tone in all interactions.

000

응답







온프레미스 솔루션

다양한 용량의 자체 모델을 통해 고객사 요구에 맞춘 다양한 AI 모델을 제공합니다.

Al Models



Cobalt

효율성 중시 다중 사용자 제공 다목적 AI. 솔루션



Onyx * Clevi App 제공 모델

넓은 지식 베이스를 가진 효율화 모델



Obsidian

확장된 추론 능력을 제공하는 고성능 Al. 솔루션



Diamond_vNext
* State of the Art

최고 수준의 성능과 정밀도를 보장하는 초대형 AI. 솔루션

Advantages



온프레미스 설치

- 고객사 서버에 직접 설치 가능
- 데이터 보안 강화
- 맞춤형 솔루션 및 유지보수 지원



고객 혜택

- 안정적인 데이터 보호
- 맞춤형 AI 솔루션 제공
- 고도화된 분석 및 인사이트 도출



Cobalt



AS 접수 등 다양한 업무 시나리오에 유연하게 대응

고객 문의나 AS(After Service) 접수와 같은 업무 절차도 자연어로 쉽게 처리할 수 있어, 기업 내 고객지원 부서나 서비스 센터 운영 효율을 크게 높여줍니다.

- ❖ 빠른 응답 속도를 자랑하는 비교적 작은 LLM 모델
 Cobalt는 규모가 작음에도 불구하고, 고도화된 아키텍처를 통해 빠른 출력이 가능
 이를 통해 다양한 환경에서도 즉각적인 정보 제공과 사용자 응대 실현
- ❖ 여러 사용자를 동시 지원 동시 접속자가 많은 상황에서도 안정적인 처리 성능을 유지 여러 사용자에게 끊김 없는 고객 지원 및 정보 전달 제공
- ❖ 손쉬운 유지 · 확장으로 비즈니스 효율 극대화 상대적으로 작은 모델 사이즈 덕분에 인프라 부담이 적고, 필요한 경우 확장이나 연동이 용이 다양한 규모의 비즈니스 요구사항에 유연하게 대응 가능





어시스턴트와 문서 검색에 최적화

자연어 기반의 문맥 분석 능력을 갖춘 대화형 어시스턴트로서 풍부한 답변을 생성하고 대규모 문서를 빠르게 검색 · 분석합니 다. ❖ 더 큰 규모의 파라미터로 향상된 성능
기존 모델 대비 파라미터 수가 많아 복잡한 연산이나 복수의 태스크를 처리할 때 우수한 성능 발휘

❖ 코드 해석 및 작성 능력
코드의 문맥을 이해하고, 오류를 찾거나 코드 작성을 보조하는 등 개발자를 위한 다양한 기능을 제공

❖ 효율성과 확장성을 극대화한 아키텍처
 높은 처리량과 유연한 구조로, 조직의 다양한 요구사항에 맞춰 손쉽게 확장할 수 있어
 대규모 환경에서도 안정적으로 운영이 가능



Obsidian



혁신적인 아이디어 발굴

문제 해결에 있어 대규모 데이터 분석과 문맥 추론을 통해 창의적이고 다양한 관점을 제시 및 높은 수준의 협업 환경에 최적화되어 있습니다.

❖ 복잡한 연산과 예측을 안정적으로 처리 방대한 양의 데이터에서도 높은 정확도 및 세밀한 해석 능력 보유 다중 작업 및 고급 코드 해석 지원

❖ 여러 태스크를 동시에 처리할 수 있어 팀 생산성 향상 깊이 있는 코드 이해 및 작성 기능을 통해 개발자의 업무 효율을 극대화 새로운 인사이트 창출

❖ 다양한 툴과 연동되며, 복잡한 비즈니스 프로세스에도 유연하게 대응 프로젝트 단위로 워크플로우를 자동화하고, 전문가 수준의 지원으로 협업 성과 극대화



Diamond_vNext



다양한 학문 분야나 업무 환경과 연동해 적용 범위 확장

내부 프로세스와의 연계 및 맞춤형 튜닝을 통해 복잡한 프로젝트에도 원활히 대응합니다.

- ❖ 모호한 질문에도 단계별 추론 과정을 제시하여, 결론 도출의 근거를 명확히 설명 방대한 연구 및 검증을 목적으로 활용 시, 학습 및 개선 포인트를 쉽게 파악 가능 과학·수학 분야에서의 탁월한 성능
- ❖ 복잡한 공식이나 이론을 이해·적용해 정확한 계산 및 해석 제공 데이터 분석, 수치 모델링 등 고난도 작업에서도 우수한 결과 보장 높은 수준의 신뢰도와 재현성
- ❖ 완성된 답변 뒤에 숨겨진 논리적 설계를 그대로 확인할 수 있어. 결과물의 신뢰도 상승 협업 및 공동 연구 시, 동일한 과정을 재현해 검증 가능 연구·개발 분야에 적합한 확장성

⟨ Chapter III ⟩ Clevi 플랫폼의 특장점



데이터 품질 및 관리

- ❖ 적대적 데이터 증강 (Adversarial Data Augmentation):
 - 가상 샘플 생성(텍스트 무작위 변환, 이미지 회전·반전 등) 및 적대적 및 공격적 데이터 생성 및 분류로 고품질 데이터 세트 구현
- ❖ 정밀 리뷰 (Expert Review):
 - 전문가가 샘플 단위로 리뷰하여 최종 정확도 보장재귀적 피드백이 가져다 주는 이점
- ❖ 재학습 (Continual Learning) 기반 교정 모델이 지속적으로 자신의 예측 결과를 반영해 데이터 품질을 점차 개선

SVCE

- Secure Container Environment: 네트워크·파일·메모리 등 다층 격리로 외부 침입을 최소화
- * Al-Driven Orchestration: LLM이 컨테이너 자원 할당과 스케줄링 결정을 자동으로 내림 → 최적화된 부하 분산
- ❖ Zero-Trust Architecture: 내부·외부 트래픽을 균일하게 모니터링하여, 최소 권한 원칙으로 불필요한 접근을 차단
- ❖ Stateful Container 관리: 세션 정보, 사용자 맞춤 데이터 등을 컨테이너 내부에서 안전하게 지속 및 백업



ARI

❖ Role-Based Prompting:

'데이터 수집 담당', '품질 검수 담당' 등 역할별 명령을 할당하여 AI 간 협업을 체계화

❖ 명시적 Input/Output 정의:

각 단계의 입력과 출력을 분명하게 설정함으로써 에러를 최소화, 협업 시 인과관계가 명확

❖ 체크포인트 (Checkpoint) 기반 검증:

단계별 산출물을 저장・검증해 안전하고 일관된 프로세스 구축

❖ 멀티 모달 연계:

텍스트, 음성, 이미지, 센서 데이터 등을 연동하여 더욱 다양한 워크플로우 개발 가능

vNext

❖ Chain-of-Thought (생각의 사슬):

모델이 답을 도출하는 모든 중간 단계를 투명하게 보여주어 신뢰도 상승

Self-Critique & Correction:

모델 스스로 오류 가능성을 검토하고, 필요하면 즉시 수정 제안

Explainable AI (XAI):

의사결정 로직을 시각적으로 표현하여, 인간이 결과를 검증하고 학습 과정에 참여 가능

Multi-Task Mastery:

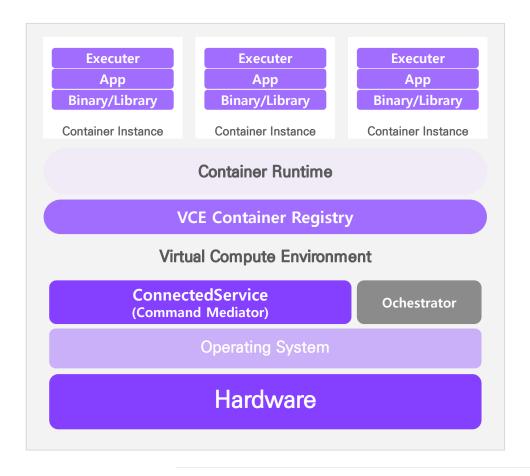
단순 질의응답을 넘어 코딩 보조, 문서 요약, 데이터 분석 등 다양한 업무를 동시에 수행

〈 Chapter IV 〉 도입 시 고려사항



Clevi-X MDLM 모델을 위한 안전한 격리 컴퓨터 환경

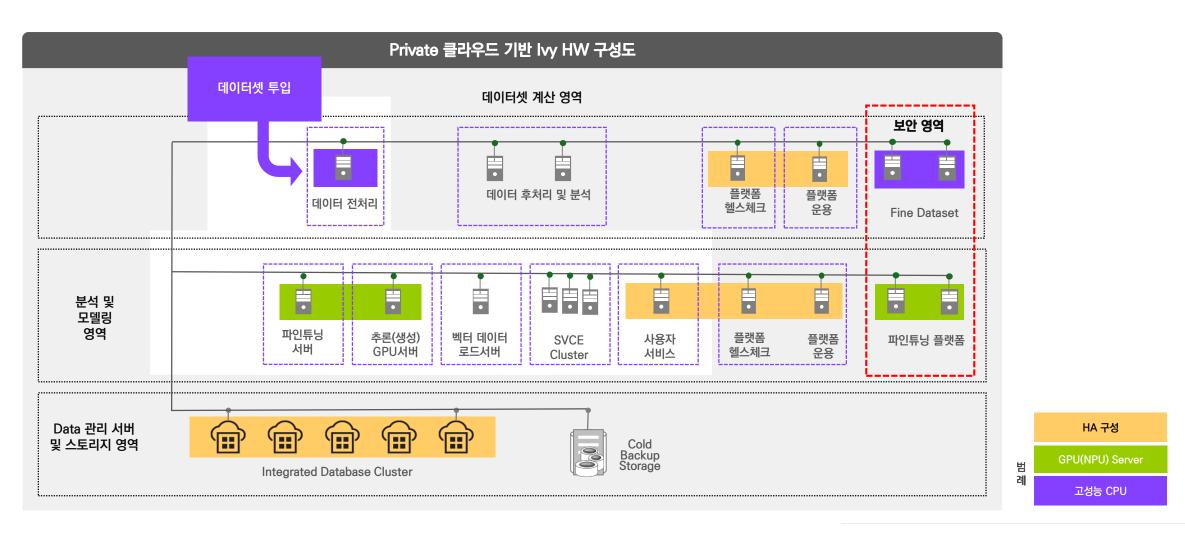
- Clevi-X MDLM (M Double LM Massive Multimodal Language Model)은 인간의 컴퓨팅 사용 환경 및 방식, 구조, 절차를 학습하여 컴퓨팅을 지원할 수 있도록 설계됨
- CLK (Clevi Linux Kernel) 기반 경량 컨테이너 환경은 바이러스의 침입 및 악의적 접근은 물 론 권한을 넘어서는 모든 작업이 불가능하도록 설계되어 있음
- 여러 익명 유저가 다양한 목적을 가지고 접속하는 클라우드 컴퓨팅 환경은 **악의적 사용자**도 물론 접속할 수 있음
- 이런 유저에게서 안전을 보장하기 위해 Clevi는 SVCE로 불리는 안전한 가상 격리 컴퓨팅 환경을 만들어 모델에게 사용할 수 있도록 제공
- 각종 인터넷 접속이나, 계산, 코드 실행, 결과 검증, 파일입/출력을 위한 안전한 격리 공간을 제공함으로 모델에게 출력 데이터의 신뢰성과 작업 실행 능력을 부여하는 한편 악의적 의도를 가진 해킹, 데이터 유출, 모델을 이용한 공격, 모델을 향한 공격 등에서 안전한 환경을 제공



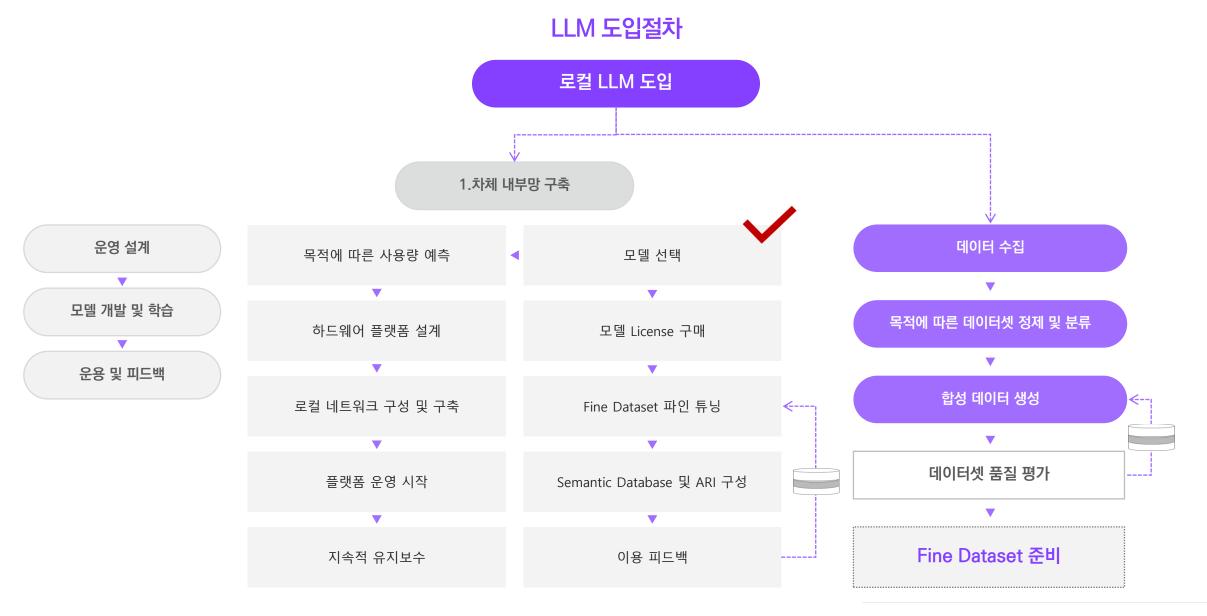


HW 요구사항

폐쇠망 기반 Full Service 구성 시 구조





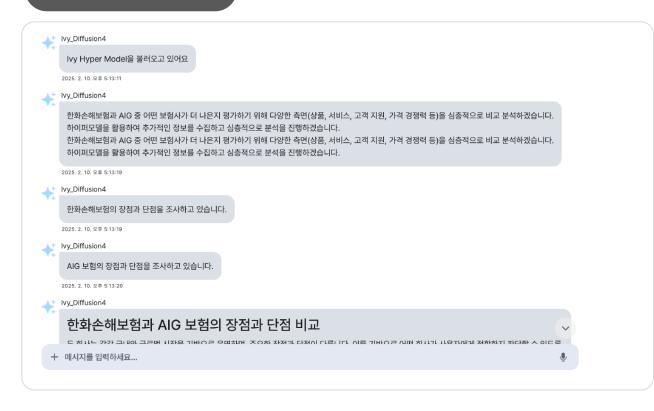


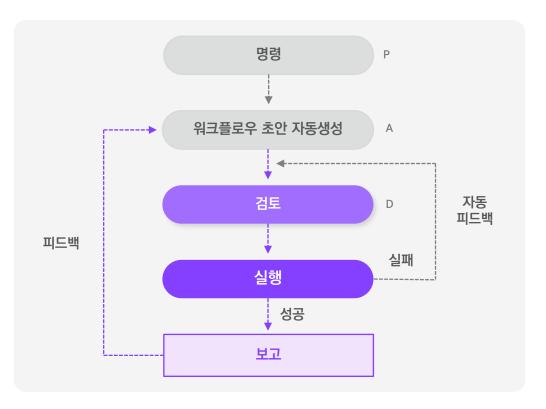


LLM 적용 - ARI Sample

- ✓ 자연어 명령을 통해 AI가 자율적으로 노드를 생성 · 해석하여 기본적인 실행 워크플로우를 만듦
- ✓ AI 기반 능동 워크플로우 시스템은 AI에게 가이드라인을 제공하며 필요한 경우 AI가 생성한 워크플로우를 수정 가능
- ✓ 작업 수행시 발생하는 예외상황은 AI에 의한 동적 의사결정을 통해 자율적으로 행동하여 사용자의 요청을 이행

실행 예시









글로벌 제조공정(MES) 최적화와 예측적 유지보수

도메인

제조(MES) + 공급망 관리 + 예측 분석

사용자

• 글로벌 생산관리 총괄 '오준영'

ARI가 분석한 업무 요구사항

- 전 세계 3개 공장의 센서 데이터를 실시간 수집, 품질 검사, 생산 실적 분석
- 불량률이 2% 이상 상승하는 시점에 자동으로 생산 파라미터 재조정
- 원자재 재고 모니터링 후 부족 시 대체 공급업체와의 긴급 계약 자동 수행
- 일정 주기(월말)마다 과거 한 달치의 생산/품질 데이터 기반으로 LLM 노드가 머신러닝 코드 자동 생성, 향후 한 분기의 수요 예측 및 생산 일정 재배치
- 사내 MES 표준, GMP (우수제조관리기준) 문서, 국제 품질규정 KB 참조, 권한 관리 (RBAC) 통해 특정 고급 노드 변경은 관리자만 가능
- 인증/보안 정책 변경 시 자동 반영 (예: 새로운 API 인증 토큰 처리 노드 삽입)

실행 과정

1. 자연어 명령

"매시간 전 세계 공장의 센서 데이터 수집 후 품질 검사, 불량률 2% 초과 시 설비 파라미터 수정, 원자재 부족 시 대체 공급업체 계약. 매월 말 과거 데이터 기반으로 생산계획 예측 모델 업데이트"

2. LLM 해석 & 워크플로우 생성

- LLM(MES 전문가 페르소나) 노드가 센서 데이터 수집 노드(각 공장 API), 품질검사 코드 노드. 파라미터 재조정 노드, 재고 확인 및 주문 노드, 머신러닝 예측 노드 생성
- MES용 노드 템플릿과 GMP/품질 규정 KB 적극 활용
- 정책 변경(인증 절차 강화) → LLM 노드가 인증 노드 삽입

3. GUI 편집

- 오준영은 GUI에서 노드 배치 및 병렬 처리 그룹 정의, 재고 조회 노드와 공급업체 계약 노드에 조건 분기 연결
- LLM 노드 클릭 → "MES 전문가" 페르소나 + 국제 품질 기준 문서 경로 설정

4. 동적 의사결정

- 센서 포맷 변경 시 LLM 노드가 에러 로그를 분석하고 파싱 코드 수정 → 재실행
- 불량률 급상승 시 대체 알고리즘(다른 파라미터 튜닝 방식) 제안

5. 성능 모니터링/로깅

- 모든 활동 JSON 로그로 기록, 감사 추적 가능
- 대규모 병렬 처리(3개 공장, 수백 센서)에도 안정적 실행, Kubernetes로 스케일 아웃

6. 장기 분석

- 한 달 데이터 축적 후 LLM 노드가 머신러닝 코드 자동 생성, 예측 모델 개선
- 예측 결과 기반으로 생산일정 재배치 노드 추가 제안 → GUI에서 승인 후 반영



고급 DevOps 파이프라인 개선과 지속적 코드 품질 관리

도메인

• DevOps, CI/CD, 마이크로서비스

사용자

DevOps 리드 '이영희'

ARI가 분석한 업무 요구사항

- 수십 개의 마이크로서비스 코드 저장소에 대해 커밋 발생 시 자동 빌드, 유닛/통합 테스트 실행
- 컨테이너 이미지 보안 취약점 스캔 후 품질 미달 시 LLM이 코드 개선안 제안
- 테스트 커버리지·성능 측정 → 기준치 이하일 경우 Canary 배포 대신 Slack 알림 노드로 라우팅
- 사내 DevOps 베스트 프랙티스 문서, 클라우드 배포 정책 KB 참조
- RBAC로 특정 고급 노드(배포 관련 노드) 수정은 DevOps 관리자만 가능

실행 과정

1. 자연어 명령

"각 서비스에 커밋 발생 시 빌드, 테스트, 취약점 스캔 후 품질 기준 만족하면 Canary 배포, 아니면 Slack 알림. 2주마다 LLM이 코드 품질 패턴 분석해 파이프라인 구조 개선 제안"

2. LLM 해석 & 워크플로우 생성

- LLM(DevOps 엔지니어 페르소나) 노드가 빌드 노드, 테스트 노드, 취약점 스캔 노드, 성능 분석 노드, Canary 배포 노드, Slack 알림 노드 구성
- DevOps 가이드 KB 참조, 인증 토큰 관리 노드 삽입

3. GUI 편집

- 이영희는 GUI에서 병렬 빌드 노드(다수 마이크로서비스 동시 처리) 구성
- Canary 배포 노드 클릭 → "DevOps 엔지니어" 페르소나, 특정 API 사양 문서 경로 지정

4. 동적 의사결정

- 테스트 실패 시 LLM 노드가 로그 분석, 코드 수정안(예: 테스트 커버리지 향상용 코드) 생성 → 재실행
- 성능 저하 발생 시 대체 설정(리소스 할당 증가) 자동 제안

5. 장기 개선

- 2주 후 LLM 노드가 반복적 오류 패턴 감지, 파이프라인 구조 최적화(불필요한 테스트 단계 제거, 병렬 처리 최적화) 제안
- GUI에서 승인 후 즉시 반영, JSON 로그로 감사 추적



연구(R&D) 대규모 데이터 분석과 지능형 파이프라인 리팩토링

도메인

• 연구 데이터 분석, 머신러닝, 고성능 계산

사용자

• 연구소 데이터 사이언티스트 '이창호'

ARI가 분석한 업무 요구사항

- 매일 수백 GB 규모의 실험데이터 HPC(고성능 컴퓨팅) 클러스터에서 수집
- Spark 기반 전처리, 머신러닝 이상치 검출, 결과 그래프, PDF 보고서 자동 생성
- 한 달 주기로 LLM 노드가 결과 분석. 효율적 알고리즘 코드를 새로 생성해 파이프라인 자동 리팩토링
- 연구 윤리, 데이터 프라이버시 규정 KB 참조, 민감정보 마스킹 자동 적용

실행 과정

1. 자연어 명령

"매일 HPC 클러스터 데이터 전처리. 이상치 감지. 그래프+PDF 보고서 생성. 매달 성능 분석 후 LLM이 더 효율적인 알고리즘 코드로 파이프라인 자동 개선"

2. LLM 해석 & 워크플로우 생성

- LLM(데이터 사이언티스트 페르소나)가 데이터 수집 노드, Spark 전처리 코드 노드, 머신러닝 모델 노드, 그래프 생성 노드, PDF 리포트 노드, 성능 모니터 노드 구성
- 연구 윤리 KB 참조. 민감 데이터 마스킹 노드 삽입

3. GUI 편집

- 이창호는 GUI에서 반복 실행(매일), 월말 성능 분석 노드(조건 분기) 추가
- LLM 노드 프롬프트 패널에서 "고성능 알고리즘 우선" 옵션 선택

4. 동적 의사결정

- 특정 날 데이터 포맷 변경 시 LLM 노드가 오류 감지→코드 개선 후 재실행
- 머신러닝 모델 성능 저하 감지 시 LLM 노드가 다른 알고리즘 (예: 랜덤포레스트→XGBoost) 코드 생성 제안

5. 장기 개선

- 한 달 후 LLM 노드가 반복적 패턴 분석, 파이프라인 노드 재배치(전처리 단계를 병렬화), 더 효율적 API 호출 구조 제안
- GUI에서 승인 후 JSON 업데이트, 로그 기록



HR 글로벌 인력 관리와 자동 교육 커리큘럼 추천

도메인

• 인사관리(HR), 교육 추천

사용자

HR 총괄 '정미희'

ARI가 분석한 업무 요구사항

- 매월 말 각 국가 지사의 신규 입사자 목록 수집, 이력서 스킬 분석, 부족 스킬별 맞춤 교육 추천
- 지역별 HR 규정, 개인정보 보호 정책 준수, 민감 데이터 마스킹
- 6개월 주기로 성과 데이터 수집 후 LLM 노드가 교육 커리큘럼 코드 재생성. 국제 HR 정책 변화 시 자동 반영(예: 추가 인증 노드 삽입)

실행 과정

1. 자연어 명령

"매월 말 각 지사 신규 입사자 정보 가져와 이력서 분석. 부족 스킬 교육 배정. 민감정보 보호. 6개월 후 성과 검토 후 커리큘럼 재설계"

2. LLM 해석 & 워크플로우 생성

- LLM(HR 전문가 페르소나): HR API 노드, 이력서 파싱 코드 노드, 교육 추천 노드, 개인정보 마스 킹 노드, 성과 평가 노드, 커리큘럼 재설계 노드 구성
- 국제 HR 정책 KB 참고, 보안 정책 문서 분석해 인증 노드 삽입

3. GUI 편집

- 정미희는 GUI로 조건 분기(특정 스킬 부족 시 특정 교육과정 할당) 구성
- LLM 노드 클릭 → "HR 전문가" 페르소나 설정, 프롬프트 패널에서 "문서 KB경로" 지정

4 동적 의사결정

- 특정 지역 이력서 포맷 변경 → LLM 노드가 파싱코드 수정→재실행
- 6개월 후 성과데이터 분석: LLM 노드가 교육 커리큘럼 구조적 변경 (예: 코드로 표현된 교육 과정 리스트) 제안

5. 정책 반영

개인정보 보호 규정 업데이트 시 LLM 노드가 마스킹 노드 파라미터 변경, 승인 후 반영



교육 분야 맞춤형 학습 경로 제안과 자동 난이도 조정

도메인

• 교육, 이러닝

사용자

• 교육기관 이러닝 매니저 '김민수'

ARI가 분석한 업무 요구사항

- 매주 학생들의 강의 수강 기록, 퀴즈 성적 분석
- 부족 개념에 맞춘 추가 학습자료 추천, 난이도별 퀴즈 자동 재생성
- 한 학기 후 LLM 노드가 전체 성과 분석해 커리큘럼 재배치 코드 생성
- 교육 정책 문서(KB) 참조. 접근성 가이드 준수

실행 과정

1. 자연어 명령

"매주 학생별 퀴즈 성적, 시청 이력 분석 후 부족한 개념 자료 추천, 난이도 조절 퀴즈 재생성, 한 학기 후 성과 보고서 바탕으로 커리큘럼 재배치"

2. LLM 해석 & 워크플로우 생성

- LLM(교육 전문가 페르소나): LMS API 호출 노드, 퀴즈 생성 코드 노드, 자료 추천 노드, 난이도 조절 파라미터 노드, 커리큘럼 재배치 노드 구성
- 교육 정책 KB(학습 윤리, 접근성) 참조, 필요 시 인증 노드 삽입

3. GUI 편집

- 김민수는 GUI에서 반복 실행(매주), 조건 분기(성적 기준치 이하 학생에게 추가 자료) 노드 연결
- LLM 노드 프롬프트 패널에서 "초보자친화 난이도" 옵션 선택

4. 동적 의사결정

- LMS 응답 포맷 변경 시 LLM 노드가 파싱 코드 개선→재실행
- 특정 학생군 반복 학습 어려움 감지 시 LLM 노드가 추가 지원자료 코드 노드 생성 제안

5. 장기 개선

• 한 학기 후 축적 데이터로 LLM 노드가 커리큘럼 재배치 코드 제안, GUI 승인 후 업데이트



COPYRIGHT © 2025, CleviCS Inc. ALL RIGHTS RESERVED. CONFIDENTIAL & PROPRIETARY.

POWERED BY CleviCS Inc.