

[페블러스 데이터클리닉 2.0] Agentic AI 기술을 적용한 AI-Ready 데이터 플랫폼

- 작성일: 2025년 11월 22일
- 기획: (주)페블러스 (Pebblous, Inc.) | 데이터 커뮤니케이션 팀
- 문서 버전: 1.0
- 인터랙티브 콘텐츠: <https://blog.pebblous.ai/>

1. 서론: AI 시대, 데이터 병목 현상 해결을 위한 혁신

인공지능(AI) 기술이 산업 전반의 혁신을 주도하고 있지만, 그 이면에는 성공을 가로막는 치명적인 병목 현상이 존재합니다. AI 개발 프로젝트의 성패가 고품질 데이터에 달려있음에도 불구하고, 실제 개발 과정의 약 80%는 데이터 수집, 정제, 가공 등 지난한 준비 작업에 소요됩니다. 이는 AI 도입의 가장 큰 장벽 이자, 기업의 혁신 속도를 저해하는 막대한 비효율의 근원입니다.

페블러스는 이 문제를 정면으로 해결하기 위해 AI 학습 데이터의 품질을 종합적으로 진단하고 보정하는 솔루션, '데이터 클리닉(Data Clinic)'을 시장에 성공적으로 안착시켰습니다. 이제 우리는 여기서 한 걸음 더 나아가, 데이터 관련 업무 패러다임을 근본적으로 전환할 새로운 비전을 선포하고자 합니다. 바로 '자율 AI 데이터 사이언티스트(Agentic AI Data Scientist, AADS)' 기술을 통해 AI 모델 학습에 즉시 사용 가능한 AI-Ready 데이터를 자율적으로 생산하고 공급하는 지능형 데이터 플랫폼으로의 진화입니다.

본 보고서는 페블러스의 현재 핵심 역량을 시작으로, 데이터 업무의 완전한 자율화를 실현할 AADS 기술의 비전과 구체적인 로드맵을 제시합니다. 이를 통해 페블러스가 어떻게 AI 시대의 데이터 문제를 해결하고, 고객과 투자자에게 지속 가능한 가치를 제공하며, AI-Ready 데이터 공급 시장의 글로벌 리더로 도약할 것인지에 대한 청사진을 공유하겠습니다.

2. 페블러스의 핵심 역량: 지능형 데이터 솔루션 스위트

페블러스의 미래 비전인 AADS 기술은 허공에서 탄생한 것이 아닙니다. 이는 수년간 축적해 온 독자적인 데이터 기술력과 시장에서 검증된 핵심 솔루션들의 유기적인 결합을 통해 실현 가능한, 계산된 목표입니다. 현재 우리가 보유한 지능형 데이터 솔루션 스위트는 AADS라는 거대한 기술적 도약의 단단한 발판이 됩니다.

페블러스의 핵심 솔루션은 다음과 같습니다.

- **데이터 클리닉 (Data Clinic):** AI 학습 데이터의 품질 문제를 종합적으로 진단하고, 데이터의 핵심 가치는 보존하며 불필요한 부분을 제거하는 '데이터 다이어트(Data Diet)'와 부족한 부분을 지능적으로 보강하는 '데이터 벌크업(Bulk-up)' 기능을 통해 품질을 자동 보정하는 페블러스의 핵심 엔진입니다. 데이터의 정확성, 일관성, 완전성 등을 체계적으로 관리하여 AI 모델의 성능을 극대화합니다.
- **페블로스코프 (PebbleScope):** 데이터의 임베딩 구조를 시각적으로 해석하여 데이터셋의 내재적 특성과 잠재적 편향성을 직관적으로 파악하게 해주는 분석 도구입니다. 데이터 전문가가 숨겨진 인사이트를 발견하고 품질 개선 방향을 설정하는 데 기여합니다.
- **합성 데이터 (Synthetic Data):** 현실 세계에서 확보하기 어려운 희귀 데이터나 개인정보보호에 민감한 데이터를 보강하기 위한 고품질 가상 데이터 생성 기술입니다. 데이터 불균형 문제를 해결하고, AI 모델의 강건성을 높이는 데 필수적입니다.

이러한 솔루션들은 이미 시장에서 그 가치를 입증받았으며, 각 기술은 데이터 품질 관리의 특정 단계를 자동화하고 지능화하는 데 기여해왔습니다. 이제 우리는 이 검증된 기술들을 하나의 지능형 자율 시스템으로 통합, 경쟁사가 따라올 수 없는 AADS라는 기술적 해자를 구축할 것입니다.

3. AADS(Agentic AI Data Scientist): 데이터 업무의 완전한 자율화를 향한 도약

AADS는 페블러스 기술 전략의 정점이자, 데이터 관련 업무의 패러다임을 '수동'에서 '자율'로 전환하는 혁신적인 개념입니다. 이는 단순히 기존 작업을 자동화하는 것을 넘어, AI 에이전트가 스스로 데이터 품질 문제를 진단하고, 해결책을 계획하며, 실행하고, 그 결과를 보고하는 완전한 자율 운영 체계를 지향합니다.

3.1. 개발 동기: "AI 개발의 80%는 데이터 업무"

AI 개발의 성공이 데이터에 달려있다는 것은 자명한 사실이지만, 현실은 데이터 과학자와 엔지니어들이 모델 개발보다 데이터 준비에 대부분의 시간을 쏟고 있는 실정입니다. 데이터의 수집, 품질 진단, 정제, 개선, 그리고 결과 보고에 이르는 이 지난한 과정은 전체 AI 개발 파이프라인의 80%를 차지하며 심각한 병목을 유발합니다. AADS는 바로 이 비효율을 근본적으로 해결하고자 하는 문제의식에서 출발했습니다. 우리는 반복적이고 노동 집약적인 데이터 업무를 지능형 AI 에이전트에게 위임함으로써, 인간 전문가들이 더 창의적이고 전략적인 고부가가치 업무에 집중할 수 있는 환경을 만들고자 합니다.

3.2. 기술 목표: 자율성과 신뢰성을 갖춘 AI 데이터 전문가

AADS의 핵심 기술 목표는 명확합니다. 인간 데이터 과학자와 같이 자율적으로 데이터 품질을 진단하고 개선하며, 그 모든 수행 과정을 투명하게 기록하여 신뢰성을 보장하고, 이를 통해 데이터 관리 업무의 효율을 극대화하는 것입니다.

이를 실현하기 위해 페블러스는 한국과학기술정보연구원(KISTI)과 긴밀히 협력하고 있습니다. KISTI가 보유한 과학기술 분야 특화 거대언어모델(LLM)인 'KONI LLM'을 기반으로, 데이터 품질 진단 및 개선

에 특화된 **'AADS-LLM'**을 공동 개발할 것입니다. 이 특화 LLM은 AADS의 두뇌 역할을 하며, 복잡한 데이터 문제에 대한 깊이 있는 이해와 추론 능력을 제공할 것입니다.

3.3. 전략적 의의: 국제 표준 기반의 데이터 거버넌스 확립

AADS의 전략적 가치는 단순한 생산성 향상을 넘어섭니다. AADS는 데이터 처리의 모든 과정을 ISO/IEC 5259(데이터 품질 모델) 및 ISO/IEC 42001(AI 경영시스템)과 같은 국제 표준에 기반하여 수행하고 기록합니다. 이는 문제가 발생한 뒤에 대응하는 '수비적(defensive)' 데이터 관리를 넘어, 설계 단계부터 고품질 데이터를 보장하는 '공격적(offensive)' 패러다임으로의 전환을 의미합니다.

AADS는 '설계 기반 데이터 품질(Data Quality by Design)' 원칙을 시스템에 내재화합니다. 이는 저 품질 데이터가 생성될 가능성 자체를 사전에 차단하는 접근법으로, 사후에 데이터를 정제하는 막대한 비용을 근본적으로 절감시킵니다. 결과적으로 AADS는 고객의 데이터 거버넌스 체계를 국제 표준 수준으로 격상시키고, 모든 AI-Ready 데이터에 감사 가능한(auditable) 투명성과 책임성을 부여하는 핵심 동력이 될 것입니다.

4. 비전의 구체화: AADS 1차년도 6대 정량 목표

비전의 실행은 환상이 아닙니다. 페블러스는 AADS의 원대한 포부를 현실에 단단히 고정시키기 위해, 1차년도에 달성을 6가지의 타협 없는 정량적 이정표를 정의했습니다. 이것은 단순한 기술 목표가 아니라, 우리의 이해관계자들에게 규율 있는 실행과 검증 가능한 진척을 약속하는 계약입니다.

AADS 1차년도 6대 정량 목표는 다음과 같습니다.

핵심 목표	성공 기준 (1차년도)
1. 자율적 데이터 품질 개선	국제 표준(ISO/IEC 5259) 기반 품질 지수 88점 달성.
2. 에이전트 기반 기능 수행	데이터 품질 관련 기능(진단, 개선 등)을 포함한 100개 이상의 Task 를 에이전트가 성공적으로 수행 (성공률 95%).
3. 투명한 수행 과정 보장	국제 표준(ISO 42001 Annex A) 스키마를 따르는 로그 생성 충족률 80% .
4. 품질 관리 효율 향상	기존 수작업 워크플로우 대비 업무 소요시간 30% 단축 .
5. 고품질 학습 데이터 구축	AADS 특화 LLM 학습을 위한 Q&A 데이터셋 학습 10만 건 , 테스트 1만 건 생성.
6. 우수한 챗봇 성능	데이터 품질/규제 관련 질의응답에서 ChatGPT-4o 대비 90% 수준의 정확도 달성.

이 6가지 목표는 AADS의 핵심 가치인 자율성, 기능성, 신뢰성, 효율성, 데이터 기반, 성능을 각각 정량적으로 입증하도록 설계되었습니다. 이 목표들의 성공적인 달성을 페블러스의 기술력을 증명하는 동시에, 장기적인 기술 고도화 로드맵의 성공적인 첫 단추가 될 것입니다.

5. 3개년 기술 고도화 로드맵: AADS 플랫폼의 완성

완전한 자율 데이터 플랫폼으로의 여정은 단 한 번의 도약이 아닌, 치밀하게 계산된 3단계의 캠페인입니다. 각 단계는 이전 단계의 성공 위에 구축되며, 기술적 경로의 리스크를 체계적으로 제거하는 동시에 시장 점유율을 공격적으로 확장하여 최종적으로 글로벌 리더십을 확보하도록 설계되었습니다.

5.1. 1단계 (기반 구축): 핵심 기술 연동 및 시장성 검증

- **목표:** KISTI의 기반 기술과 페블러스의 데이터 클리닉 엔진을 연동하여 텍스트 데이터 품질 관리의 핵심 기술을 확보하고, 초기 시장성을 검증합니다.
- **핵심 개발 내용:**
 - KISTI의 KONI LLM을 기반으로 데이터 품질 진단에 특화된 **AADS-LLM v0.9**를 공동 개발 합니다.
 - 텍스트 데이터 품질 진단 및 리포팅 API를 포함한 **데이터 클리닉 v2.0 MVP(최소기능제 품)**를 출시합니다.
- **예상 성과:**
 - 핵심 타겟 도메인인 로봇 1곳, 제조 2곳의 고객사를 대상으로 PoC(기술검증)를 성공적으로 수행하여 기술의 효용성을 입증합니다.

5.2. 2단계 (기술 고도화): 멀티모달 확장 및 상용화

- **목표:** 텍스트를 넘어 차트(Chart)와 같은 멀티모달 데이터까지 처리 범위를 확장하고, 완성도 높은 상용 서비스 패키지를 출시하여 본격적인 시장 진입을 시작합니다.
- **핵심 개발 내용:**
 - 차트의 구조와 내용을 이해하고 추론할 수 있는 특화 **VLM(Vision-Language Model)** 을 개발합니다.
 - 멀티모달 데이터 개선 기능과 에이전트 실행 플랫폼을 통합한 **데이터 클리닉 v3.0 상용 패키지**를 개발합니다.
- **예상 성과:**
 - PoC 고객의 유료 전환 및 신규 고객 확보를 통해 5곳 이상의 유료 상용 고객을 확보하고, 시장에서의 입지를 다집니다.

5.3. 3단계 (플랫폼 완성): 완전 자율화 및 글로벌 사업화

- **목표:** 완전한 자율 운영(Autonomous)이 가능한 AADS 플랫폼을 완성하고, 글로벌 클라우드 마켓플레이스 등록을 통해 본격적인 글로벌 사업을 전개합니다.

- **핵심 개발 내용:**
 - 다수의 AI 에이전트를 조율하고 제어하는 에이전트 오케스트레이션 및 도구 연동 시스템 기술을 적용하여 AADS v1.0 플랫폼을 정식 출시합니다.
 - ISO 42001 규제 준수를 위한 거버넌스 대시보드와 원클릭 감사 리포트 기능을 구현합니다.
- **예상 성과:**
 - AWS 등 글로벌 클라우드 마켓플레이스에 AADS 플랫폼을 정식 등록하고, 해외 기술 파트너십을 통해 초기 해외 매출을 발생시킵니다.

이 로드맵을 통해 탄생할 '데이터 클리닉 2.0'은 단순한 데이터 품질 관리 도구를 넘어, 고객의 AI 개발 생태계 전체를 혁신하는 전략적 플랫폼으로 자리매김할 것입니다.

6. Data Clinic 2.0: AADS가 제공하는 고객 가치와 미래 비전

AADS 기술이 완전히 통합된 '데이터 클리닉 2.0'은 고객에게 단순한 기능 개선 이상의 전략적 가치를 제공합니다. 이는 AI 개발의 가장 큰 걸림돌이었던 데이터 준비 과정을 자동화하고, 데이터 자산의 신뢰도를 극대화하며, 복잡한 규제 환경에 능동적으로 대응할 수 있는 혁신적인 솔루션입니다.

데이터 클리닉 2.0이 제공하는 핵심 고객 가치는 다음과 같습니다.

- **개발자 속도 해방 (Unlocking Developer Velocity):** AI 개발 파이프라인의 80%를 차지하는 데이터 병목을 제거하여, 귀사의 핵심 인재가 데이터 정제가 아닌 가치 창출에 집중하도록 만듭니다.
- **데이터 신뢰성 극대화:** ISO 국제 표준을 준수하는 자동화된 품질 관리 및 리포팅을 통해 모든 데이터 처리 과정을 추적하고 증명할 수 있습니다. 이는 외부 감사에 대응하고 데이터 기반 의사결정의 신뢰도를 확보하는 강력한 기반이 됩니다.
- **자율적 문제 해결:** AI 에이전트가 데이터 품질 문제를 스스로 진단하고 최적의 개선 시나리오를 제시함으로써, 소수의 데이터 전문가에게 의존하던 기존의 업무 방식을 탈피합니다. 예를 들어, 로봇 경로 계획 데이터에서 발생하는 '간헐적인 데이터 누락'이나 제조 공급망 데이터의 '부품 규격 단위 (mm/inch) 오기입'과 같은 치명적인 오류를 자율적으로 진단하고 해결책을 제안하여 운영 리스크를 최소화합니다.
- **AI 도입 리스크 제거 (De-risking AI Adoption):** 복잡한 글로벌 AI 규제(ISO 42001, EU AI Act)를 위한 감사 가능한 증적과 보고서를 원클릭으로 생성하여, 컴플라이언스를 비용이 아닌 경쟁 우위로 전환합니다.

이러한 독보적인 고객 가치를 바탕으로, 페블러스는 데이터 클리닉 2.0을 통해 시장을 선도하는 기업으로 성장할 것입니다. 우리의 장기적인 목표는 단순히 솔루션을 판매하는 것을 넘어, AI-Ready 데이터를 자율적으로 생산하고 공급하는 플랫폼 생태계를 구축하는 것입니다.

7. 페블러스 성장 전략: AI-Ready 데이터 공급의 글로벌 리더

페블러스는 AADS 기술을 중심으로 명확하고 야심찬 장기 성장 전략을 추진하여, AI-Ready 데이터 공

급 분야의 글로벌 리더로 도약하고자 합니다. 우리의 전략은 투자자 관점에서 지속 가능한 성장과 높은 기술적 해자를 구축하는 데 초점을 맞추고 있습니다.

페블러스의 3대 핵심 성장 전략은 다음과 같습니다.

- AI-Ready 데이터 플랫폼 사업자로의 전환:** 우리는 고객의 요청에 수동적으로 대응하는 '솔루션' 제공자를 넘어설 것입니다. 우리의 비전은 고품질의 AI-Ready 데이터를 자율적으로 대량 생산하고, 이를 필요로 하는 모든 기업에 안정적으로 공급하는 지능형 '플랫폼' 사업자로 진화하는 것입니다. 이를 통해 페블러스는 단순 판매 모델에서 벗어나, 데이터 공급 생태계를 주도하는 시장의 핵심 플레이어로 자리매김할 것입니다.
- 자율 데이터 품질 관리 기술의 초격차 확보:** AADS의 핵심은 **'자율적인 데이터 품질 관리 및 개선 기술'**입니다. 우리는 이 핵심 기술을 지속적으로 고도화하고 관련 특허 포트폴리오를 확장하여 강력한 기술적 해자(Technological Moat)를 구축할 것입니다. 경쟁사가 모방하기 어려운 압도적인 기술 격차를 확보함으로써, 시장의 선점 효과를 극대화하고 장기적인 경쟁 우위를 유지할 것입니다.
- 차세대 Physical AI 시장 선점:** 지능형 로봇, 스마트 제조, 자율주행 등 차세대 Physical AI 분야는 가상 환경에서 생성된 초고품질의 데이터를 필수로 요구합니다. 페블러스는 로봇 및 제조 분야의 PoC를 통해 이 시장에 대한 전략적 교두보를 이미 확보했으며, 향후 이 분야 최고의 데이터 생산자로서의 입지를 조기에 구축하여 미래 고성장 시장을 선점할 계획입니다.

결론적으로, 페블러스는 AADS라는 강력한 기술 엔진을 통해 데이터 중심 AI 시대를 선도할 것입니다.

우리는 데이터 병목 현상을 해결하는 것을 넘어, 데이터의 생성, 관리, 공급 방식 자체를 혁신할 것입니다. 이를 통해 고객에게는 압도적인 생산성과 신뢰를, 투자자에게는 지속 가능한 성장과 가치를 제공하는 글로벌 리더 기업으로 성장할 것을 확신합니다. 페블러스는 데이터 문제의 해결사를 넘어, AI 시대의 표준을 정의하는 기업이 될 것입니다. 우리는 데이터가 흐르는 모든 곳에 신뢰와 가치를 창출하며, AI 산업의 미래를 이끌어 나갈 것을 약속합니다.

Pebblous