



Pebblous Data Greenhouse: 비즈니스 및 투자 유치 전략 보고서

- 기획: 페블러스 데이터커뮤니케이션팀
- 버전: 0.1
- 작성일: 2025-12-28
- 공개여부: 비공개

목차

- 1. 투자 요약 — Data Greenhouse는 “AI 데이터 운영 체계”를 제품화한다
- 2. 시장 문제 — AI의 병목은 모델이 아니라 데이터 운영에 있다
- 3. 솔루션 정의 — Data Greenhouse는 플랫폼 위의 “해석과 행동” 레이어다
- 4. 제품 차별화 — “관측”이 아니라 “개선”과 “증명”을 판매한다
- 5. 사업화 전략 — “Wedge Use Case”에서 “운영 OS”로 확장한다
- 6. 수익 모델과 스케일링 — SaaS 중심 전환과 ‘장기 고객 확보 구조’의 설계
- 7. 해자 전략 — 특허, 표준, 인증, 그리고 소버린 배포
- 8. 투자 유치 논리 — 자금의 목적은 “생존”이 아니라 “스케일업”이다
- 9. 리스크와 대응 — 레퍼런스 공백은 “첫 번째 온실 작물”로 해결한다
- 10. 결론 — Data Greenhouse는 AI 시대의 “데이터 책임”을 제품으로 만든다

1. 투자 요약 — Data Greenhouse는 “AI 데이터 운영 체계”를 제품화한다

페블러스 데이터 그린하우스(Pebblous Data Greenhouse)는 인공지능 데이터의 품질을 “측정”하는 단계에 머무르지 않고, **품질 개선, 비용 절감, 규제 대응**의 세가지 목표를 동시에 타기팅하는 **자율형 데이터 운영 체계(OS)**로 정의된다. 기업이 Snowflake, Databricks, Data Lake 등 기존 플랫폼을 이미 도입했음에도 AI 프로젝트가 반복적으로 실패하거나 비용이 급증하는 이유는, 플랫폼이 제공하는 저장·처리·실험 기능만으로는 “이 데이터가 실제로 모델 성능에 기여하는가”와 “이 데이터 운영이 규제·감사에 부합하는가”라는 질문에 답하기 어렵기 때문이다. Data Greenhouse는 바로 이 공백을 겨냥한다.

Data Greenhouse의 핵심 가치 제안은 단순하다.

- 첫째, 데이터 중복과 과밀로 발생하는 컴퓨팅·스토리지 낭비를 데이터 구조 수준에서 제거하여 **비용을 구조적으로 절감**한다.
- 둘째, 모델 성능의 변동을 데이터 분포와 품질 관점에서 해석함으로써 데이터를 **“설명 가능한 자산”**으로 만든다. 이 과정에서 필요한 경우 합성 데이터(Synthetic Data)를 활용하여 데이터 희소성 문제를 해결하고 모델 학습의 효율을 높일 수 있다.
- 셋째, ISO/IEC 5259 및 ISO 42001 수준의 증적과 로그를 운영 과정에 내장하여 **규제 대응**을 사후 문서 작업이 아니라 시스템의 기본 기능으로 전환한다.

이 세 가지가 동시에 달성될 때, 기업은 AI 도입을 “실험”에서 “산업 운영”으로 전환할 수 있다.

2. 시장 문제 — AI의 병목은 모델이 아니라 데이터 운영에 있다

AI 도입의 현실에서 가장 큰 병목은 모델 성능이 아니라 데이터 준비와 데이터 운영에 있다. 국내외 AI 전략 문서 분석에서도 데이터 준비 과정이 전체 AI 도입 절차의 대부분을 차지하며, 이 병목을 해결하지 않으면 국가적 차원의 AI 인프라 전략도 효율적으로 작동하기 어렵다는 문제의식이 명확히 제시된다. 이 구조는 개별 기업에서도 동일하게 반복된다. 모델은 빠르게 진화하지만, 데이터는 여전히 수집·정제·검증·갱신의 비용과 책임이 분산된 채로 남아 있기 때문이다.

또한 AI 프로젝트는 점점 더 규제와 신뢰의 요구를 강하게 받는다. 규제 환경은 단순히 “데이터를 잘 관리하라”는 선언이 아니라, 데이터가 어떤 근거로 품질을 확보했고 어떤 방식으로 운영되었는지를 **감사 가능한 형태로 증명**할 것을 요구한다. 이때 데이터 품질을 표준과 연결해 정량화하고, 자동으로 감사 리포트를 생성하는 기능은 선택이 아니라 시장 진입의 조건이 된다. Pebblous는 AADS* 기술로 ‘데이터 클리닉’에 내장된 거버넌스 엔진을 통해 ISO 42001의 핵심 통제 항목에 대한 충족률을 객관적으로 입증해 왔고, 이를 규제 시장에서의 경쟁 우위로 전환할 수 있는 기반을 갖추고 있다.

*AADS: Agentic AI Data Scientist

시장 기회는 AI가 “디지털 생성”을 넘어 “물리 세계와 상호작용하는 피지컬 AI”로 이동하면서 더 커진다. 제조·로봇·국방 등 고신뢰 데이터가 필수인 분야에서는 데이터 품질이 곧 안전과 직결되며, 이 영역에서는 단순한 관측(Observability)이나 단순한 저장/처리가 아니라 **“품질의 운영”**과 **“증적의 자동화”**가 경쟁력을 좌우한다. Data Greenhouse는 이 변화에 맞춰 멀티모달 데이터 품질 관리와 산업 특화 모델을 위한 데이터 운영 인프라로 확장할 수 있는 구조를 갖는다.

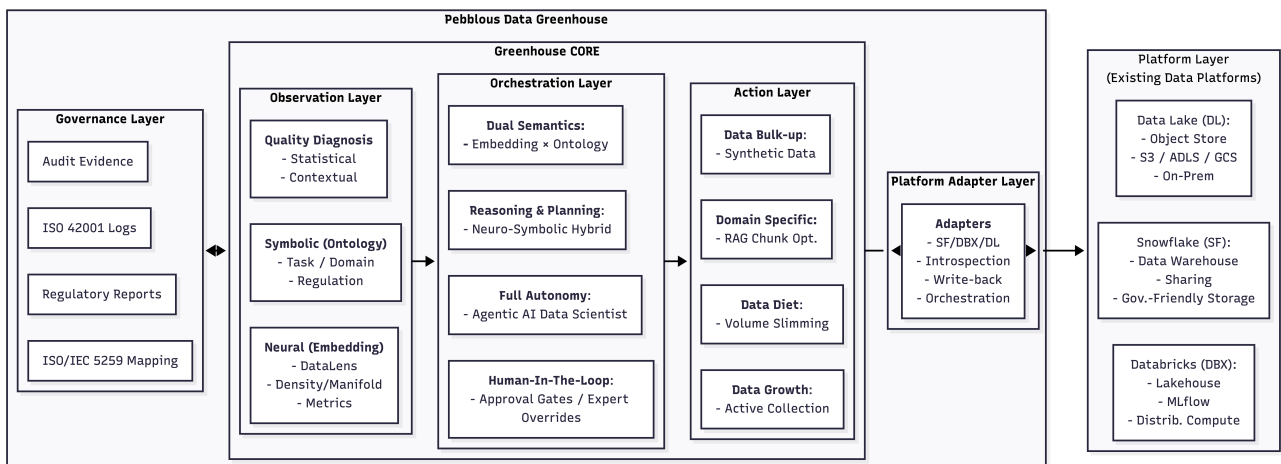
3. 솔루션 정의 — Data Greenhouse는 플랫폼 위의 “해석과 행동” 레이어다

Data Greenhouse는 Snowflake, Databricks, Data Lake를 대체하지 않는다. 오히려 이들을 플랫폼 계층으로 명확히 하부에 두고, 그 위에 **“관측-판단-행동-증명”**(Plan-Diagnose-Improve-Govern)의 순환 구조를 엮는다. 이 접근은 기업이 이미 투자한 데이터 플랫폼 자산을 존중하면서도, 플

랫폼이 답하지 못하는 질문을 제품화하는 방식이다. Data Greenhouse가 지향하는 바는 플랫폼 경쟁이 아니라 플랫폼 위에서 의사결정을 가능하게 만드는 **해석 레이어**의 장악이다.

Data Greenhouse의 기술적 핵심은 **뉴로-심볼릭(Neuro-Symbolic)** 전략의 실전 구현이다. **임베딩 기반 분석**은 데이터의 유사성, 과밀, 공백, 분포 왜곡을 “공간적 구조”로 드러내지만, 그것이 문제인지 의미 있는 이벤트인지는 말해주지 않는다. **온톨로지 기반 분석**은 태스크·도메인·규제 맥락을 제공하지만, 현상의 강도와 구조를 정량화하지 못한다. Data Greenhouse는 임베딩(Neural)과 온톨로지(Symbolic)를 결합하여 “무엇이 이상한가”와 “왜 중요한가”를 동시에 도출하고, 이를 실행 가능한 개선 계획으로 변환한다. 이 결합은 단지 분석을 정교하게 만드는 것이 아니라, 제품을 “**스스로 판단하고 실행하는 데이터 운영 시스템**”으로 만든다.

Data Greenhouse의 실행 엔진은 AADS(Agentic AI Data Scientist)로 정의된다. AADS는 **관측-판단-행동-증명**의 루프를 통해 사용자의 목표를 해석하고, 진단을 수행하며, 개선 행동을 실행하고, 그 결과를 보고서와 증적으로 정리한다. 이 구조는 기존의 **인간 주도형 Data Clinic**을 계승하면서도, **데이터 품질 관리의 전 과정을 자율화**하는 방향으로 발전한다. 또한 중요한 작업에는 Human-In-The-Loop(HITL) 관점의 승인 게이트를 둬으로써 **완전 자율성과 조직 통제 사이의 균형을** 유지한다.



4. 제품 차별화 — “관측”이 아니라 “개선”과 “증명”을 판매한다

현재의 데이터 생태계에는 관측(Observability) 중심의 도구가 풍부하지만, 관측은 문제를 “발견”할 뿐 문제를 “해결”하지 않는다. 반대로 전통적 거버넌스 도구는 규칙 기반 통제를 제공하지만, AI 데이터의 동적 품질 변화와 비정형 데이터의 구조적 특성을 정량화하기 어렵다. Pebblous는 데이터의 기하학적 현상(Vector)과 개념적 맥락(Ontology)을 통합 이해하는 전략을 통해, 관측을 넘어 직접적인 품질 개선(Data Diet/Bulk-up)을 수행하고, 그 과정을 거버넌스 증적으로 자동 변환한다. 이 결합은 단일 기능 경쟁이 아니라 “운영 체계 경쟁”의 형태로 방어 가능한 차별화를 만든다.

이 차별화의 법적·기술적 기반은 Pebblous의 핵심 특허 포트폴리오가 지탱한다. 특히 미국 등록 특허 (US 12,481,720 B2)는 데이터의 속성을 임베딩 공간에서 진단하고, 밀도·분포 분석을 통해 개선(예: 중복 제거, 공백 보강)을 수행하는 방법론을 장기간 보호한다. 이 특허는 Data Clinic, PebbloScope, Data Diet/Bulk-up, AADS로 이어지는 제품 라인을 관통하는 기술적 해자로 기능하며, 표준이 요구하

는 추상적 품질 특성을 “측정 가능하고 개선 가능한 공학”으로 전환하는 구현체로 위치한다.

5. 사업화 전략 — “Wedge Use Case”에서 “운영 OS”로 확장한다

Data Greenhouse의 시장 진입은 거대한 플랫폼 교체가 아니라, 즉시 ROI가 검증되는 문제에서 시작해야 한다.

웨지 유스 케이스(Wedge Use Case): 제품이 경쟁사보다 ‘압도적으로 우수하게 해결할 수 있는, 가장 작고 구체적인 단일 문제(Use Case)’를 뜻한다. 이는 고객의 지갑을 열게 만들고, 초기 레퍼런스를 확보하며, 경쟁사가 쉽게 따라올 수 없게 만드는 '날카로운 쐐기(Wedge)' 역할을 한다. 킬러 앱과도 유사한 의미이다.

가장 강력한 첫번째 웨지는 데이터 중복과 과밀로 인한 낭비를 줄이는 “Data Diet”다. 기업은 데이터가 많아질수록 저장 비용과 컴퓨팅 비용을 함께 지불하지만, 중복 데이터는 성능을 개선시키지 못하는 경우가 많다. Data Diet는 데이터셋의 구조적 중복을 제거해 학습 효율과 비용 효율을 동시에 개선하며, 투자 의사결정권자에게 “절감액”이라는 언어로 가치를 전달할 수 있다.

두번째 웨지는 규제와 감사가 엄격한 시장에서의 “증적 자동화”다. Data Greenhouse는 ISO/IEC 5259 기반의 품질 매핑과 ISO 42001 기반의 로그 생성 구조를 통해, 고객이 내부 감사와 외부 규제 요구에 대응할 수 있는 자동 리포트를 제공한다. 규제 대응은 기업에게 비용과 리스크로 인식되기 쉽지만, 반대로 규제 대응 능력은 고신뢰 시장에서의 진입 장벽이기도 하다. Pebblous는 별도의 사후 절차였던 규제 대응을 운영 시스템의 핵심 기술 기능으로 내재화함으로써, 규제 준수 부담을 없애고 규제 시장에서의 차별화를 실현할 수 있다.

세번째 웨지는 엔터프라이즈 AI의 필수불가결한 접근법인 RAG 환경에서의 의미 중복 제거와 커버리지 보강이다. RAG 지식 베이스의 중복 및 공백은 검색 품질과 비용을 동시에 악화시키는 핵심 병목이다. 도메인 특화 SLM 확산과 소버린 AI 환경이 중요해지면서 RAG 자산의 신뢰성은 더욱 핵심적인 요소가 되었다. Data Greenhouse는 임베딩과 온톨로지 결합을 통해 중복/공백 위험을 운영 관점에서 진단하고, 지식 자산을 “운영 가능한 품질 체계”로 변환하여 고신뢰 엔터프라이즈 RAG 시장을 개척합니다.

네번째 확장은 Physical AI 영역이다. 대한민국 국가 AI 전략에서 피지컬 AI는 차세대 핵심 영역으로 강조되고 있으며, 고품질 데이터가 산업별 파운데이션 모델과 현장 적용의 전제 조건임이 반복적으로 언급된다. 특히 제조·로봇·국방 등 고신뢰 분야에서는 데이터 품질이 곧 시스템의 안전과 직결된다. Pebblous는 이 변화에 맞춰 제조/로봇 영역의 데이터 품질 관리, 초고품질 합성데이터 생성, 자율 데이터 운영을 결합해, 고위험·고신뢰 시장에서 강력한 레퍼런스 자산을 축적하고 주도권을 확보할 수 있는 구조를 갖춘다.

6. 수익 모델과 스케일링 — SaaS 중심 전환과 ‘장기 고객 확보 구조’의 설계

Data Greenhouse는 단발성 프로젝트형 용역 모델을 넘어, 고속 성장을 위한 **반복 매출 중심의 구독형 SaaS(Software-as-a-Service)** 로 사업 구조를 본격적으로 전환하도록 설계되었다. Pebblous의 재무 로드맵은 2030년 매출 100억 원 달성을 목표로 하며, 이 중 SaaS 비중을 60%대 중반까지 끌어 올리는 공격적인 전략을 명시한다. 또한 **온프레미스 패키지 배포와 인증 서비스**가 수익 포트폴리오를 다 각화하는 핵심 축이며, 용역 비중은 장기적으로 한 자릿수로 축소된다. Data Greenhouse는 이러한 고 성장형 사업 구조에 최적화된 제품 형태이다.

투자 유치 관점에서 Data Greenhouse의 핵심 가치는 **구조적 고객 이탈 방지(Lock-In)** 를 부작용이 아닌 강력한 **가치(Value)** 로 전환하는 데 있다. Data Greenhouse는 데이터 품질 지수, 개선 이력, 실행 로그, 규제 리포트 등을 체계적으로 축적하여, **AI 데이터 운영의 핵심 기록 시스템(System of Record)** 가 된다. 이 운영 기록은 단순한 기술 설정값이 아니라 조직의 핵심 의사결정 근거로 기능하며, **시스템의 가치를 시간이 갈수록 증대시킨다.**

이러한 **축적형 가치 구조**는 두 가지 결과를 낳는다. 첫째, 데이터와 운영 이력이 쌓일수록 고객의 제품 전환 비용(Switching Cost)은 구조적으로 커진다. 둘째, Data Greenhouse 기반으로 고객의 데이터 운영 품질은 더욱 안정화된다. 투자자는 이러한 구조를 통해 높은 고객 생애 가치(LTV)와 낮은 이탈 가능성(Churn Rate)을 기대할 수 있으며, 이것이 곧 Data Greenhouse의 장기 고객 확보 구조를 완성하는 논리이다.

7. 해자 전략 — 특허, 표준, 인증, 그리고 소버린 배포

Data Greenhouse의 방어력은 기능 경쟁이 아니라 “구조 경쟁”에서 형성된다.

첫째, Data Greenhouse의 가장 강력한 방어력은 핵심 기술에 대한 30여건 이상의 특허 포트폴리오이다. 특히 **미국 등록 특허(US 12,481,720 B2)** 는 데이터의 속성을 임베딩 공간에서 진단하고, 밀도 및 분포 분석을 통해 개선(예: 중복 제거, 공백 보강)을 수행하는 방법론을 장기간 보호한다. 이 특허는 Data Clinic, PebbloScope, Data Diet/Bulk-up, AADS로 이어지는 제품 라인의 기술적 근간이며, 경쟁사가 동일한 형태의 품질 진단 및 개선 사이클을 상업적으로 구현하기 어렵게 만드는 법적 해자(Moat)로 기능한다.

둘째, ISO/IEC 5259와 같은 국제 표준을 단순히 준수하는 수준을 넘어, **표준이 정의하는 추상적 품질 특성의 공백을 메우는 구현체를 보유하는 것이다.** Data Greenhouse는 ISO/IEC 5259가 요구하는 AI 데이터 품질 특성들을 실제 '측정 가능하고 개선 가능한 공학적 방법론'으로 전환한다. 이를 통해 시장에서 요구하는 품질 기준을 선도적으로 충족하는 포지션을 확보하고, 표준을 준수하는 것을 넘어 표준의 공학적 해석과 실행을 선도한다.

셋째, Data Greenhouse는 거버넌스 및 규제 대응 기능을 제품의 핵심으로 내재화한다. **ISO 42001** 기반의 감사 로그 생성 및 **ISO/IEC 5259** 기반의 품질 매핑과 같은 거버넌스 기능은 규제 시대의 필수 요건을 충족한다. 이는 규제 대응을 사후 문서 작업이나 별도 절차가 아닌 **운영 시스템의 기본 기능**으로 전환하는 것이며, 규제 강화가 곧 규제 준수 역량을 갖춘 Data Greenhouse의 시장 확대를 가속화하는 구조적 경쟁 우위를 만든다.

넷째, **소버린 AI와 온프레미스 배포**는 공공·국방·금융 등 고보안 시장에서의 실질적 진입 장벽을 낮춘다. 국가 전략 분석 문서는 외부 통신이 어려운 환경에서의 온프레미스 패키지가 외산 클라우드 기반 솔루션이 넘보기 힘든 경쟁 우위를 제공할 수 있음을 강조하며, Pebblous가 해당 방향성과 정합성을 갖는다고 분석한다. 이는 단순한 옵션이 아니라, 장기적으로 가장 강한 고객군을 확보하기 위한 전략적 포지션이다.

8. 투자 유치 논리 — 자금의 목적은 “생존”이 아니라 “스케일업”이다

Data Greenhouse는 연구용 데모가 아니라, 스케일업 가능한 B2B 제품으로 정의된다. Pebblous의 계획은 투자금을 생존 자금이 아니라 글로벌 확장과 SaaS 스케일업을 위한 재원으로 사용하며, 구체적으로 SaaS 스케일업(인프라 및 운영 인력), Biz/Sales(글로벌 세일즈와 파이프라인 확장), R&D 고도화(AADS 및 합성 데이터 기술), 인증 사업(KOLAS 기반 인증 역량)에 투자하는 구조를 제시한다. 이 사용 계획은 Data Greenhouse가 “제품-영업-규제”의 삼각 구도를 동시에 밀어야 한다는 점에서 합리적이다.

투자자는 Data Greenhouse가 제공하는 세 가지 성과 지표에 주목할 수 있다. 첫째, **성능 개선**은 데이터 품질 지수(QI) 변화와 모델 성능 변화의 상관을 통해 설명 가능해진다. 둘째, **비용 절감**은 Data Diet/Bulk-up의 전후 비교로 정량화될 수 있다. 셋째, **규제 대응**을 통한 신뢰성 확보는 ISO 기반 매핑과 감사 로그의 완결성으로 증명된다. 이 지표들은 단지 성과 측정이 아니라, 고객의 의사결정을 움직이는 “구매 근거”로 작동한다.

9. 리스크와 대응 — 레퍼런스 공백은 “첫 번째 온실 작물”로 해결한다

초기 단계에서 가장 현실적인 리스크는 레퍼런스의 절대량이 부족하다는 점과, 고객이 “플랫폼 내부 최적화”로 문제를 해결하려고 시도하는 경향이 있다는 점이다. Data Greenhouse는 이 리스크를 오히려 제품 전략으로 전환해야 한다. 즉, 플랫폼 튜닝과 경쟁하지 않고 “이 쿼리를 돌린 데이터 중 실제로 정보 기여도가 있는 샘플은 몇 %인가”라는 질문을 전면에 내세워, **비용 절감과 성능 개선의 원인을 데이터 구조에서 설명**하는 방식으로 판매해야 한다. 이 질문에 답할 수 있는 조직은 거의 없고, 이 공백이 곧 시장 기회다.

또한 레퍼런스 공백은 “범용 성공 사례”보다 “산업용 성공 사례”에서 더 빠르게 메워질 수 있다. 제조·국방·공공과 같은 **고신뢰 시장은 도입 장벽이 높은 대신, 한 번 검증되면 확산 파급력이 크고 락인이 강하다**. 국가 전략 문서가 지적하는 산업 AX와 피지컬 AI 전환의 흐름 속에서, Data Greenhouse는 데이터 품질 관리와 자율 개선을 필수 과업으로 끌어올리는 정책·시장 레버리지를 활용할 수 있다.

10. 결론 — Data Greenhouse는 AI 시대의 “데이터 책임”을 제품으

로 만든다

Pebblous Data Greenhouse는 데이터 품질을 측정하는 도구를 넘어, 데이터가 스스로 성장하고, 그 성장의 근거가 축적되며, 규제와 산업 요구를 충족하도록 만드는 산업용 운영 인프라다. 이 제품은 Snowflake와 Databricks 같은 플랫폼을 대체하지 않고, 플랫폼 위에서 의사결정을 가능하게 만드는 책임 레이어를 장악한다. 또한 특허로 보호되는 데이터 이미징 방법론, 표준 기반 품질 매핑, 감사 로그 자동화, 온프레미스·소버린 배포 전략은 제품의 방어력과 확장성을 동시에 확보한다. 따라서 Data Greenhouse는 “기술 제품”이 아니라 “운영 체계”로서 투자 논리를 갖추며, 투자금은 이 운영 체계를 시장 표준으로 밀어붙이기 위한 스케일업 자원으로 정의될 수 있다.

Pebblous

Pebblous Makes Data Tangible

contact@pebblous.ai