



피지컬 AI를 위한 데이터, 사냥할 것인가 재배할 것인가?

《데이터 그린하우스》가 제안하는 데이터 경작의 패러다임

- 기획: 페블러스 데이터커뮤니케이션팀
- 감수: 이주행 (페블러스 대표)
- 작성일: 2026-01-02

Exec Summary

피지컬 AI 시대, AI가 '안전'을 배우려면 '위험한 순간'의 데이터가 필요하지만 현실에서 사고를 일부러 낼 수는 없다는 역설이 있습니다. 페블러스는 이러한 '데이터 기근'을 물리 법칙을 완벽히 복제한 디지털 트윈(PebbloSim), 물리적 환각 없는 뉴로-심볼릭 합성 데이터, 그리고 자율 에이전트가 관리하는 데이터 품질 체계로 해결합니다. 마치 숙련된 농부가 스마트 온실에서 초고품질의 토마토를 재배하듯, 페블러스는 《데이터 그린하우스》에서 AI 학습 데이터를 경작합니다.

1. 씨앗 없이는 수확도 없다

자율주행차, 휴머노이드 로봇, 스마트 팩토리. 피지컬 AI(Physical AI)의 시대가 도래했습니다. 하지만 화려한 비전 뒤에는 치명적인 병목이 숨어 있습니다.

"정작 가장 중요한 데이터는 수집할 수 없다"는 역설입니다.

자율주행차가 보행자와 충돌하는 순간, 로봇 팔이 오작동하는 상황, 재난 현장의 데이터. 이런 '희귀 데이터(Edge Case)'가 있어야 AI는 안전을 배웁니다. 하지만 현실에서 사고를 고의로 낼 수는 없습니다.

우리는 이것을 '데이터 기근(Data Famine)' 이라 부릅니다. 수렵(Hunting)하듯 데이터를 찾아다니는 방식으로 이 기근을 해결할 수 없습니다.

2. 온실에서 경작하는 데이터

농부가 겨울에도 토마토를 수확할 수 있는 건 온실이 있기 때문입니다. 온실은 자연을 거스르는 것이 아니라, 통제된 환경에서 성장을 최적화하는 것입니다.

페블러스의 데이터 그린하우스(Data Greenhouse) 는 이 철학을 AI 데이터에 적용합니다. 현실의 물리 법칙을 완벽히 복제한 가상 환경(Digital Twin)에서, 필요한 상황을 의도적으로 만들어냅니다. 디지털 트윈은 단순한 3D 모델이 아니라, 중력, 마찰, 충돌, 관성까지 현실과 동일하게 시뮬레이션하는 물리 엔진입니다.

"데이터는 우연히 발견하는 것이 아니라, 목적에 맞게 재배하는 것입니다."

3. 물리적 환각이 없는 데이터

일반적인 영상 생성 AI는 그럴듯한 이미지를 만들지만, 물리 법칙을 모릅니다. 차가 공중에 뜨거나 그림자가 제각각인 '물리적 환각(Physical Hallucination)' 은 피지컬 AI에게 치명적인 독이 됩니다.

물리적 환각이 없는 합성데이터를 만들기 위한 페블러스의 실행 엔진, 페블로심(PebbloSim) 은 '뉴로-심볼릭(Neuro-Symbolic)' 접근을 취합니다.

구분	일반 생성 AI	PebbloSim
핵심 원리	픽셀 확률 예측 (그림 그리기)	물리 시뮬레이션 + 렌더링 (구조 만들기)
제어력	"빗길 사고" (모호한 텍스트)	마찰계수 0.3, 충돌각 45° (정밀한 수치)
결과물	환각이 포함된 이미지	설명 가능한 물리적 데이터

쉽게 말해, 뼈대(물리 시뮬레이션) 를 먼저 세우고 그 위에 피부(생성 AI) 를 입힙니다.

뼈대가 정확하기에 데이터는 물리적으로 완벽하며, AI가 실패했을 때 그 원인을 공학적으로 설명할 수 있습니다.

4. 자율 에이전트가 관리하는 데이터

데이터 그린하우스는 단순한 공장이 아닙니다. 데이터의 건강 상태를 관리하는 **자율 운영 체계(OS)** 입니다.

1. **관측(Observe)**: 데이터의 부족함과 편향을 실시간으로 진단합니다.
2. **처방(Orchestrate)**: "우천 시 야간 데이터가 부족하다"는 처방을 내립니다.
3. **행동(Act)**: PebbloSim이 부족한 데이터를 정밀 타격하여 생성(Bulk-up)합니다.
4. **증명(Govern)**: 생성 전후의 품질 변화를 입증하고 기록합니다.

EU AI Act나 ISO 42001과 같은 규제가 "당신의 AI는 안전한가요?"라고 물을 때, 데이터 그린하우스는 대답합니다.

"우리는 부족한 위험 데이터를 진단했고, 물리적으로 검증된 데이터로 보강했습니다. **여기 그 과정이 담긴 '운영 증거(Operational Evidence)'가 있습니다.**"

5. 파트너십 제안

피지컬 AI 시장의 성패는 모델이 아니라 '**누가 더 가치 있는 데이터를 보유했는가**'에서 갈립니다. 페블러스는 현재 자동차, 국방, 조선, 로봇틱스 분야의 선도 기업들과 함께 데이터의 미래를 검증하고 있습니다.

특히 과학기술정보통신부의 '**AI 글로벌 빅테크 육성사업**' 주관으로 수행 중인 **AADS(Agentic AI Data Scientist) 2단계 과제**에서는 피지컬 AI 분야의 데이터를 핵심 타겟으로 삼고 있습니다. 멀티모달 합성데이터 생성 기술과 제조 산업 특화 **소버린 VLM(Vision-Language Model)** 개발을 통해, 대한민국 제조 강국의 AI 경쟁력을 한 단계 끌어올리고자 합니다.


물리적 정합성이 보장된 고품질 합성데이터가 필요한 제조 산업의 주요 기업들, 그리고 신뢰할 수 있는 AI 데이터 인프라를 함께 구축할 파트너분들의 연락을 기다립니다.

"우리는 데이터를 수집하지 않습니다. 재배합니다."

Pebblous

Pebblous Makes Data Tangible

 contact@pebblous.ai

 pebblous.ai