

PebbloSim

피지컬 AI를 위한 시뮬레이션 기반 합성데이터 생성기

— 개념 설계 및 개발 전략: v2.0

-
- 작성일: 2026-01-10 (최초);
 - 작성자: 페블러스 페블로심-팩토리 TF
 - 주작성: 이주행
 - 변경이력:
 - v1.0 대비, 시나리오 관리자를 GenSim으로 이름 변경하여 기능을 포함함.
 - 자동차 시나리오를 예로 들어 모듈들의 동작을 워크플로우로 설명함

목차

1. 비전 및 상위 개념.....	2
1.1. 서론: 피지컬 AI 시대의 데이터 병목과 PebbloSim의 전략적 중요성.....	2
1.2. 핵심 개념 및 목표.....	3
1.3. 주요 적용 도메인.....	4
1.4. 핵심 차별화 요소 1: '운영 증거'로서의 PebbloSim.....	5
1.5. 핵심 차별화 요소 2: 뉴로-심볼릭 하이브리드 월드 모델.....	6
1.6. 핵심 차별화 요소 3: 자가 증식형 선순환 구조 (The Data Flywheel).....	7
2. 시스템 아키텍처 및 개념 설계.....	8
2.1. 서론: 데이터 운영체제 기반 시뮬레이션 및 합성데이터.....	8
2.2. 전체 시스템 구성 및 데이터 흐름.....	8
Step 1. 디지털 트윈 엔진 (The Base Class).....	8
Step 2. GenSim 관리자 (The Architect).....	9
Step 3. 멀티모달 합성 데이터 생성기 (The Action Engine).....	9
Step 4. PebbloScope 연동 모듈 (The Director).....	9
2.3. 페블로심 핵심 모듈별 기능 정의.....	9
2.4. 페블로심-그린하우스 연동 메커니즘.....	11
3. 시나리오 기반 워크플로우 검증.....	12
3.1. 예제 워크플로우 요약.....	12
3.2. 예제 시나리오 1: 자동차 도장 공정의 미세 스크래치 결함 데이터 확보.....	13
Step 1. 진단 및 처방 (Diagnosis & Prescription).....	13
Step 2. 시나리오 설계 및 번역 (Translation & Design).....	13
Step 3. 가상 환경 구축 및 데이터 생성 (Action & Generation).....	14
Step 4. 시각화 및 품질 검증 (Visualization & Verification).....	14
Step 5. 거버넌스 승인 및 입고 (Governance & Close).....	14
Data Flywheel 효과: 가치의 이중 나선 (The Double Helix of Value).....	15
4. PebbloSim 단계별 개발 전략.....	16
4.1. 서론: 점진적 완성을 통한 리스크 관리 및 AADS 2단계 목표 달성.....	16
4.2. PoC #1 (1~3개월): 기반 구축 및 자동차 공정 실증.....	16
4.3. PoC #2 (4~6개월): 국방 도메인 확장 및 소버린 체계 검증.....	17
4.4. PoC #3 (7~9개월): 멀티모달 데이터 고도화 및 조선/제조 적용.....	18
4.5. PoC #4 (10~12개월): 완전 자율화 및 플랫폼 완성.....	19
5. 결론: 데이터의 '수집'을 넘어 '재배'와 '증명'의 시대로.....	20
5.1. 패러다임의 전환: 수렵(Hunting)에서 경작(Farming)으로.....	20
5.2. 비즈니스 임팩트: 성장과 신뢰의 이중주.....	20
5.3. 비전: 피지컬 AI 시대를 위한 필수 인프라.....	21

1. 비전 및 상위 개념

1.1. 서론: 피지컬 AI 시대의 데이터 병목과 PebbloSim의 전략적 중요성

피지컬 AI 시장은 제조, 로봇, 국방, 조선 등 물리적 세계와 상호작용하는 AI 기술의 부상과 함께 폭발적인 성장을 앞두고 있습니다. 그러나 이 혁신의一面에는 치명적인 병목 현상, 즉 '데이터 기근(Data Famine)'이 존재합니다. 특히, AI 모델의 강건성을 결정하는 핵심 요소인 결함, 예외상황, 재난과 같은 희귀 데이터는 현실 세계에서 의도적으로 수집하는 것이 거의 불가능에 가깝습니다. 이러한 데이터 부족은 피지컬 AI 도입의 가장 큰 장벽으로 작용하고 있습니다.

페블로심(PebbloSim)은 페블러스의 핵심 자산인 '데이터 그린하우스(Pebblous Data Greenhouse)' 생태계 내에서 이러한 데이터 병목을 해결하기 위해 설계된 전략적 응용 프로그램입니다. 데이터 그린하우스가 데이터의 '관측-판단-행동-증명'을 자율적으로 순환하는 데이터 운영 체계(OS)라면, PebbloSim은 데이터 그린하우스의 '행동(Action)' 계층의 핵심 실행 엔진(Execution Engine)으로서, 데이터 그린하우스가 단순한 관측 시스템을 넘어 데이터의 품질과 생명주기에 대한 '책임(Responsibility)'을 지는 운영 체계임을 증명하는 가장 강력한 응용입니다. 이를 통해 AADS(Agentic AI Data Scientist) 2단계 과제의 핵심 목표인 산업 특화 AI 모델 개발을 가속하고, 피지컬 AI 시장의 데이터 주도권을 확보하는 데 결정적인 기여를 할 것입니다.

이 섹션은 PebbloSim의 근본적인 존재 이유(Why)와 궁극적인 목표(What)를 명확히 정의하며, 이를 실현하기 위한 구체적인 개념을 설명합니다.

1.2. 핵심 개념 및 목표

PebbloSim의 핵심 개념은 "피지컬 AI의 학습 데이터 생성을 위한 디지털 트윈 기반 시뮬레이션 및 합성 데이터 생성 플랫폼"으로 정의할 수 있습니다. 이는 단순히 가상 환경을 구축하는 것을 넘어, 데이터 그린하우스와 유기적으로 연동하여 AI 학습에 즉시 사용 가능한 고품질 데이터를 자율적으로 생산하고 공급하는 것을 지향합니다.

PebbloSim의 주요 목표는 다음과 같습니다.

- **고품질 AI-Ready 데이터 생성:** 현실에서 발생 빈도가 극히 낮거나 획득 비용이 매우 높은 엣지 케이스(Edge Case) 데이터를 생성합니다. 이는 기존 데이터를 증강하는 수준을 넘어, 현실에 존재하지 않는 시나리오의 초고품질 데이터를 창조하는 '하이퍼-합성 데이터(Hyper-Synthetic Data)' 기술을 통해 달성됩니다. 예를 들어, 현대자동차 스마트 팩토리의 용접 공정에서 발생하는 0.001% 미만의 희귀 불량 시나리오를 가상으로 재현하고 관련 데이터를 대량으로 생성하여, AI 모델이 어떤 예외 상황에서도 안정적으로 작동하도록 강건성을 확보하는 것이 핵심 목표입니다.
- **데이터 그린하우스 기반 운영 증거 확보:** PebbloSim은 단순한 시뮬레이션 도구를 넘어, 데이터 그린하우스의 핵심 아키텍처인 '자율 순환 루프(Autonomous Cycle-Loop)'를 완성하고 증명하는 핵심 실행체입니다. AADS가 데이터의 공백을 '관측(Observe)'하고 개선 계획을 '판단(Orchestrate)'하면, PebbloSim은 '행동(Action)' 계층에서 정밀 타기팅 '데이터 벌크업(Data Bulk-up)'을 수행합니다. 이 과정은 단순한 데이터 생성이 아니라, 데이터가 어떤 논리와 근거로 스스로 성장했는지를 입증하는 '운영 증거(Operational Evidence)'로서 '증명(Govern)' 계층에 기록됩니다.
- **AADS 2단계 기술 개발 가속화** PebbloSim은 총 4주기의 순차적 PoC(기술검증)를 통해 개발되며, 각 단계에서 도출된 모든 산출물은 AADS 2단계 정부 과제의 기술 보고서에 직접적으로 반영됩니다. 이는 단순한 내부 개발 프로젝트를 넘어, 정부 과제의 실증 목표를 성공적으로 달성하고 피지컬 AI 데이터 기술의 국가적 경쟁력 확보에 기여하는 실질적인 과정입니다.

1.3. 주요 적용 도메인

PebbloSim의 아키텍처는 다음 피지컬 AI 핵심 도메인을 전략적 우선순위에 따라 공략하도록 설계되었습니다. 각 도메인은 데이터 기근 현상이 심각하고, 고품질 데이터 확보가 곧 경쟁력과 직결되는 고부가가치 시장입니다.

- **자동차/제조:** 유연제조 환경에서 부품 운반 로봇이 예기치 않게 넘어지거나 충돌하는 등 이상 상황을 시뮬레이션하여, 자율 제조 시스템의 안정성을 높이는 데 필요한 데이터를 생성합니다.

- **국방:** 외부와 단절된 On-Premise 환경에서 적의 침투나 이상 물체 탐지와 같은 감시 경계 시나리오 및 전술 훈련 데이터를 생성하여, 국방 AI 모델의 신뢰성과 성능을 강화합니다.
- **조선:** 3D 설계도(CAD) 데이터와 현장에서 수집되는 센서 데이터를 결합한 디지털 트윈을 구축합니다. 이를 통해 선박 건조 공정의 병목 현상을 사전에 예측하고 최적화하는 시뮬레이션을 수행합니다.
- **로보틱스:** 보스턴 다이내믹스의 휴머노이드 로봇과 같이 복잡한 동작을 수행하는 로봇의 움직임 데이터와, 다양한 환경에서의 예외 상황 대응 데이터를 생성하여 로봇 파운데이션 모델의 학습을 지원합니다.

1.4. 핵심 차별화 요소 1: '운영 증거'로서의 PebbloSim

PebbloSim이 '핵심 운영 증거(Operational Evidence)'로서 기능한다는 것은, PebbloSim이 단순히 합성 데이터 파일(.jpg, .mp4 등)을 만들어내는 도구가 아니라, "데이터가 왜 생성되었고, 어떤 과정을 거쳐 품질이 개선되었는가"를 입증하는 인과관계의 기록(Audit Trail)을 생성한다는 의미입니다.

- '결과물'을 설명하는 '과정의 증명': 일반적인 시뮬레이터는 "요청한 데이터"만 내놓지만, PebbloSim은 "데이터 그린하우스가 자율적으로 문제를 해결한 기록"을 함께 내놓습니다. 이를 통해 시뮬레이션의 전 절차를 증명(Process Validation)할 수 있습니다.
 - **인과관계(Causality) 기록:** 단순히 "비 오는 날의 이미지"를 만드는 것이 아니라, "데이터 클리닉이 '우천 시 데이터 부족(Data Void)'을 진단했기 때문에(Why), AADS가 '강수량 10mm, 조도 50 lux'로 설정하여(How), PebbloSim이 이 데이터를 생성했다(Action)"는 전후 관계가 기록됩니다.
 - **자율 순환 루프의 완결:** PebbloSim의 가동 기록은 AADS의 **[관측(Obsserve) → 판단(Orchestrate) → 행동(Action) → 증명(Govern)]** 사이클이 실제로 작동했음을 보여주는 물리적 증거가 됩니다.
- **운영 증거를 구성하는 3가지 요소:** PebbloSim이 생성하는 '운영 증거 패키지'는 다음 세 가지가 결합된 형태입니다.
 - **진단 기반 처방전 (The Prescription):** AADS가 내린 '시뮬레이션 스크립트' 자체가 증거입니다. 데이터 클리닉이 탐지한 '데이터 공백(Vector Space Void)'

좌표를 시뮬레이션 파라미터로 역변환(Vector-to-Param)한 기록으로, 데이터 생성의 과학적 근거가 됩니다.

- **실행 및 생성 로그 (The Action Log):** 실제 시뮬레이션이 수행된 환경 설정(디지털 트윈의 상태), 적용된 물리 법칙, 생성된 멀티모달 데이터가 포함됩니다. 이는 데이터가 조작되지 않고 물리적 정합성(Ground Truth)을 갖췄음을 증명합니다.
- **개선 효과 확인서 (The Validation):** 생성된 데이터를 다시 데이터 클리닉에 넣어 측정했을 때, 품질 지수(Quality Index)가 얼마나 상승했는지 보여주는 'Before/After 비교 데이터'입니다. 이는 데이터 벌크업(Bulk-up)의 실질적 효과를 입증합니다.
- **규제 대응 및 비즈니스 가치 (Trust & Compliance):** 이러한 운영 증거는 기업이 직면한 AI 규제 및 신뢰성 문제를 해결하는 핵심 열쇠가 됩니다.
 - **규제 준수 증빙:** EU AI Act나 ISO/IEC 42001(AI 경영시스템)은 AI 모델이 어떤 데이터로 학습되었는지 증명할 것을 요구합니다. PebbloSim의 운영 증거는 "우리는 부족한 안전 데이터를 이렇게 과학적으로 진단하고 보강했다"는 감사 가능한(Auditible) 자료로 활용됩니다.
 - **피지컬 AI 안전 보증:** 로봇이나 자율주행과 같은 피지컬 AI 분야에서는 사고 데이터 학습이 필수적입니다. PebbloSim은 현실에서 구할 수 없는 사고 데이터를 생성하고, 이를 학습시켰다는 '안전 보증서' 역할을 수행합니다.

요약하자면, PebbloSim이 핵심 운영 증거라는 말은, "데이터가 시스템의 자율적인 판단에 의해 과학적으로 관리되고 성장하고 있음을 제3자(감사관, 고객)에게 증명할 수 있는 블랙박스 기록 장치"라는 뜻입니다.

1.5. 핵심 차별화 요소 2: 뉴로-심볼릭 하이브리드 월드 모델

PebbloSim은 물리적 법칙을 무시하는 기존 생성형 AI(Generative AI)의 한계를 극복하기 위해, '심볼릭(Symbolic) 시뮬레이션'의 논리적 정합성과 '뉴럴(Neural) 생성 모델'의 시각적 표현력을 결합한 뉴로-심볼릭 하이브리드 월드 모델 접근 방식을 채택합니다.

- **물리적 환각(Physical Hallucination) 제로화:** '월드 모델'로서의 시뮬레이터 기준의 영상 생성 AI(예: Sora, Stable Diffusion)는 픽셀 간의 확률적 상관관계만을 학습하므로,

차가 공중에 뜨거나 그림자 방향이 어긋나는 등 물리 법칙을 위배하는 '환각'을 일으킵니다. 반면, PebbloSim은 디지털 트윈 엔진(**Symbolic**)을 통해 중력, 마찰력, 광학 등 물리 법칙이 지배하는 '**월드 모델(World Model)**'을 먼저 구축합니다. 이 '뼈대' 위에 생성 AI 기술로 '피부'를 입히기 때문에, 시각적으로 사실적일 뿐만 아니라 물리적으로 완벽하게 정합된(**Physically Consistent**) 데이터를 보장합니다.

- **설명 가능한 인과성 (Causal Explainability):** 뉴럴 네트워크 기반의 생성 AI는 블랙박스(Black-box)여서 왜 그런 이미지가 생성되었는지 설명할 수 없습니다. 그러나 PebbloSim은 심볼릭 로직(**Symbolic Logic**)을 기반으로 작동하므로, "자동차가 미끄러진 이유는 마찰 계수(Friction) 파라미터가 0.3으로 설정되었기 때문"이라는 명확한 인과관계(**Causality**)를 설명할 수 있습니다. 이는 AI 모델의 실패 원인을 분석해야 하는 피지컬 AI 개발에 있어 필수불가결한 요소입니다.
- **정밀한 제어 가능성 (Controllability):** 텍스트 프롬프트("비 오는 날 차 사고")에 의존하는 생성 AI는 구체적인 상황을 통제하기 어렵습니다. PebbloSim은 **Vector-to-Param** 기술을 통해 "강우량 30mm/h, 충돌 각도 45도, 속도 60km/h"와 같이 수치적으로 정밀하게 제어된 데이터를 생성할 수 있습니다. 이는 뉴로-심볼릭 접근만이 제공할 수 있는 '공학적 통제권'입니다.

1.6. 핵심 차별화 요소 3: 자가 증식형 선순환 구조 (The Data Flywheel)

PebbloSim은 사용할수록 똑똑해집니다. 데이터 생산이 AI 지능의 강화로 이어지고, 강화된 지능이 다시 더 정교한 데이터를 생산하는 '자가 증식형 선순환(Self-Evolving Loop)'을 구현합니다.

- **지능의 내재화 (Internalizing Intelligence):**
 - PebbloSim이 생성한 고품질 합성 데이터(Curriculum Data)는 단순히 고객에게 납품되는 데 그치지 않고, 시스템의 두뇌 역할을 하는 핵심 AI 모델을 재학습시키는 데 사용됩니다.
 - 이 '**자체 학습 루프(Self-Training Loop)**'를 통해, 시스템은 시간이 지날수록 더 복잡한 물리적 상황을 이해하고 더 정교한 시나리오를 설계할 수 있게 진화합니다.
- **진단과 생성의 상호 강화 (Reinforcement Cycle):**

- [진단] 시스템이 데이터의 공백을 찾아내면 → [생성] 시뮬레이션이 이를 메우는 데이터를 생성하고 → [학습] 이 데이터로 모델 성능이 향상되면 → [재진단] 향상된 모델은 이전에는 보지 못한 더 미세한 데이터 결합을 찾아냅니다.
 - 이 무한 루프는 기업의 데이터 자산 가치를 기하급수적으로 증대시키는 **데이터 플라이휠(Data Flywheel)** 효과를 창출합니다.,
- **자산 가치의 지속적 증대 (Asset Appreciation):**
 - 일반적인 소프트웨어는 시간이 지날수록 구식이 되지만, PebbloSim 기반의 데이터 그린하우스는 데이터가 쌓일수록 진단과 생성의 정확도가 높아지는 '**가치 증대형 자산(Appreciating Asset)**'이 됩니다. 이는 경쟁사가 단기간에 모방할 수 없는 독보적인 기술적 해자(**Moat**)를 형성합니다.,

PebbloSim의 이러한 개념들을 바탕으로, 다음 장에서는 이를 구현하기 위한 구체적인 시스템 아키텍처를 상세히 기술하겠습니다.

2. 시스템 아키텍처 및 개념 설계

2.1. 서론: 데이터 운영체제 기반 시뮬레이션 및 합성데이터

PebbloSim은 페블러스 데이터 그린하우스(Pebblous Data Greenhouse)라는 AI 데이터 운영체제(OS) 위에서 구동되는 핵심 응용 프로그램(Application)입니다. 특히 데이터의 희소 영역을 물리 시뮬레이션 기반으로 보강하는 특별한 '**데이터 벌크업(Data Bulk-up)**'을 전담합니다.

이를 위해 아키텍처는 **엔진(기반시설) + 시나리오(설계서) = 시뮬레이터 인스턴스(GenSim)**라는 명확한 워크플로우로 정의되었습니다. 이 구조는 소프트웨어 공학의 '**클래스(Class)**'와 '**인스턴스(Instance)**' 개념처럼, 재사용 가능한 기반환경(엔진) 또는 프레임워크 위에 시나리오 기술하는 목적에 맞는 다양한 가상 환경(인스턴스)을 동적으로 생성합니다. 결과적으로 이 아키텍처는 PebbloSim의 정체성을 개별 도구(시뮬레이터)에서 확장 가능한 플랫폼 인프라로 격상시키는 핵심 전략입니다.

2.2. 전체 시스템 구성 및 데이터 흐름

PebbloSim은 물리적 세계를 복제하고, 그 안에서 피지컬 AI의 목적에 맞는 데이터를 생성하며, 그 과정을 표준화된 기록으로 남겨 통제를 가능하게 하며, 최종적으로 인간의 승인을 통해 데이터의 신뢰성을 확보하는 유기적인 4단계 워크플로우를 따릅니다.

Step 1. 디지털 트윈 엔진 (The Base Class)

- **역할:** 현실 세계의 물리 법칙(중력, 마찰 등)과 운영 환경의 제약 조건을 정밀하게 복제한 **디지털 기반 환경(Digital Base World)**입니다.
- **특징:** 현실의 데이터와 동기화(Reality Sync)되어 시뮬레이션의 **Ground Truth(기준점)** 역할을 합니다.

Step 2. GenSim 관리자 (The Architect)

- **역할:** 그린하우스에서 생성한 추상적인 명령(예: "우천 시 사고 데이터 필요")을 디지털 트윈을 위한 구체적인 **시뮬레이션 스크립트**로 번역(Translation)합니다.
- **특징:** 디지털 트윈 엔진 위에 시나리오를 결합하여, 목적별로 특화되고 격리된 '**GenSim 인스턴스**'를 생성합니다.

Step 3. 멀티모달 합성 데이터 생성기 (The Action Engine)

- **역할:** 데이터 그린하우스의 핵심 '**Tool**'로서, 생성된 GenSim 인스턴스 내에서 시뮬레이션을 수행하여 멀티모달 데이터(이미지, 로그 등)를 능동적으로 생산합니다.
- **특징:** 데이터 그린하우스의 '**뉴로-심볼릭(Neuro-Symbolic)** 표상 공간'에서 탐지된 **공백 영역(Void)**을 시뮬레이터 파라미터로 역변환(Vector-to-Param)하여, 필요한 데이터만 정밀 타겟(Precision Targeting)하는 **벌크업(Bulk-up)**을 수행합니다.

Step 4. PebbloScope 연동 모듈 (The Director)

- **역할:** GenSim 인스턴스의 시뮬레이션 과정을 시각적으로 모니터링하고, 스크립트를 인터랙티브하게 편집하여 데이터 생성 방향을 지휘하며 최종 승인(Approval)합니다.
- **특징:**
 - 인터랙티브 시뮬레이션 제어: 벡터 공간의 탐색뿐만 아니라, 시각화된 결과를 보며 조명, 날씨, 객체 위치 등의 GenSim 파라미터를 즉시 수정하고 재실행 결과를 프리뷰합니다.

- Human-AI 협업: 단순한 통제(Control)를 넘어, Human-in-the-Loop를 통해 AI의 제안에 인간 전문가의 직관(Insight)을 더해 데이터 품질을 완성합니다

2.3. 페블로심 핵심 모듈별 기능 정의

PebbleSim은 데이터 그린하우스라는 'OS(운영체제)' 위에서 구동되는 핵심 '**실행(Action)**' 애플리케이션으로, 영화 제작과 유사한 유기적인 4단계 워크플로우를 통해 작동합니다. 현실 세계를 정밀하게 복제한 [1. 디지털 트윈 엔진]이라는 '무대' 위에, [2. GenSim 관리자]가 구체적인 '시나리오(대본)'를 입히면, [3. 멀티모달 합성 데이터 생성기(Action Engine)]가 이를 실제 시뮬레이션으로 '연기'하여 데이터를 생성합니다. 이 전 과정은 [4. PebbleScope 연동 모듈(The Director)]을 통해 '감독(사용자)'이 실시간으로 모니터링하고 스크립트를 편집하며 최종 승인(OK)하는 구조로 완성됩니다.

단순한 데이터 생성을 넘어선 PebbleSim의 핵심 가치는 '**정밀 타겟(Precision Targeting)**'과 '**설명 가능성(Explainability)**'에 있습니다. 데이터 클리닉이 '뉴로-심볼릭 표상 공간'에서 탐지한 데이터 공백(Void)을 시뮬레이션 파라미터로 역변환(Vector-to-Param)하여 필요한 데이터만 골라내는 '데이터 벌크업(Data Bulk-up)'을 수행합니다. 또한, 벡터(현상)와 온톨로지(맥락)를 실시간 동기화하는 '**뉴로-심볼릭 통합 시각화(Interactive Link)**' 기술을 통해 데이터 생성의 인과관계를 규명하고, **Human-in-the-Loop**를 통해 AI와 인간 전문가가 협업하여 데이터의 무결성을 최종 보증합니다.

모듈명	역할 및 별칭	핵심 기능	기반 기술 및 가치
1. 디지털 트윈 엔진 (Digital Twin Engine)	<p>물리적 현실의 복제 (The Base Class)</p> <p>시뮬레이션이 구동되는 물리 법칙과 제약 조건이 동기화된 '기반 런타임 환경'</p>	<ul style="list-style-type: none"> • 물리 법칙 구현: 중력, 마찰, 광학 등 현실 물리학 적용 • Reality Sync: 실제 센서 데이터와 가상 환경 동기화 (실시간 또는 스냅샷) • Ground Truth: 데이터 정합성의 기준점 	<ul style="list-style-type: none"> • NVIDIA Omniverse • Reality Engine • 시뮬레이션의 물리적 무결성 보장

		제공	
2. GenSim 관리자 (GenSim Manager)	시나리오 설계자 (The Architect) 데이터 그린하우스의 추상적 의도를 실행 가능한 시뮬레이션 스크립트로 번역	<ul style="list-style-type: none"> Intent Translation: "우천 시 사고" 의도를 스크립트로 변환 Instance 생성: 디지털 트윈(Class) + 시나리오 = GenSim(Instance) 로그 관리: 감사 가능한(Auditible) 설계 로그 생성 	<ul style="list-style-type: none"> Ontology & LLM • 오케스트레이션 계층과 액션 계층의 연결 고리
3. 멀티모달 합성 데이터 생성기 (Multimodal Synthetic Data Generator)	데이터 생산 실행 (The Action Engine) 데이터 그린하우스와 연동되는 핵심 Tool 로서, 공백 영역을 찾아 데이터를 능동 생산	<ul style="list-style-type: none"> Precision Targeting: 뉴로-심볼릭 표상 공간의 공백을 타겟 Vector-to-Param: 결핍된 벡터 좌표를 시뮬레이션 파라미터로 역변환 데이터 벌크업: 엣지 케이스 정밀 보강 	<ul style="list-style-type: none"> Generative VLM • Vector-to-Param 기술 • 데이터 그린하우스의 Action Layer 수행
4. PebbleScope 연동 모듈 (Visualization Adapter)	지휘 및 승인 인터페이스 (The Director) 시뮬레이션을 시각적으로 모니터링하고 스크립트를 편집하여 최종 승인	<ul style="list-style-type: none"> Interactive Link: 벡터(현상)와 온톨로지(맥락)의 동기화 시각화 Interactive Editing: 시각화된 결과를 보며 파라미터 즉시 수정 Human-AI Collaboration: 전문가 직관을 더해 품질 	<ul style="list-style-type: none"> Unreal Engine • 뉴로-심볼릭 통합 • 시각화 기술 • 블랙박스 해소 및 신뢰성(Trust) 확보

2.4. 페블로심-그린하우스 연동 메커니즘

페블로심은 데이터 그린하우스 생태계의 각 계층과 다음과 같이 유기적으로 연동됩니다.

- **진단에서 처방으로 (Clinic → Architect):** 데이터 그린하우스 내의 '데이터 클리닉'이 현재 데이터셋의 편향이나 부족함(예: "눈 오는 날의 데이터가 5% 미만임")을 진단하면, 이 진단 결과가 '설계자(Architect)' 역할을 하는 **GenSim** 관리자에게 전달되어 정밀한 시나리오 생성의 설계도로 활용됩니다
- **실행과 공급 (Action Engine → Greenhouse):** 합성 데이터 생성기(Action Engine)는 무작위 생성이 아닌, 앞서 진단된 '뉴로-심볼릭 표상 공간'의 공백을 'Vector-to-Param' 기술로 역변환하여 정밀 타겟(Precision Targeting)합니다. 이를 통해 꼭 필요한 데이터만 생성하여 데이터 그린하우스의 균형을 맞추는 고효율 '데이터 벌크업(Data Bulk-up)'을 수행합니다
- **검증과 순환 (The Director → Flywheel):** 생성된 데이터는 바로 학습에 쓰이지 않고, '**The Director(PebbleScope)**'가 제공하는 지휘 및 승인 인터페이스의 '**승인 게이트(Approval Gate)**'를 통과해야만 자산으로 확정됩니다. 승인된 데이터는 AI 모델을 재학습시키고, 성능이 향상된 모델은 다시 진단의 기준이 되는 '자가 증식형 선순환(Self-Training Loop 또는 Data Flywheel)' 구조를 완성합니다.

3. 시나리오 기반 워크플로우 검증

3.1. 예제 워크플로우 요약

PebbleSim의 워크플로우는 데이터 그린하우스 생태계 내의 유기적인 협업 과정으로 요약됩니다. 먼저 데이터 그린하우스가 "데이터 부족해!"라며 결핍을 진단하면, **GenSim 관리자(The Architect)**가 "어떤 상황을 만들까?"를 고민하여 구체적인 시나리오를 설계합니다. 이 설계도는 **멀티모달 합성 데이터 생성기(The Action Engine)**로 전달되어 "가상 공장에서 찍어내자!"는 실행과 함께 필요한 데이터를 정밀 생성합니다. 생성된 데이터는 **PebbleScope(The Director)**에게 넘겨져 "품질 확인해줘"라는 요청 하에 시각적 검증을

거치며, 최종 승인된 데이터는 다시 데이터 그린하우스가 "OK, 창고에 저장"함으로써 자산화 및 거버넌스 절차를 완료합니다

워크플로우 요약:

1. **[Greenhouse]** "데이터 부족해!" (진단)
2. **[PebbloSim Manager]** "어떤 상황을 만들까?" (시나리오 설계)
3. **[PebbloSim Engine]** "가상 공장에서 찍어내자!" (실행 및 생성)
4. **[PebbloScope]** "품질 확인해줘." (검증)
5. **[Greenhouse]** "OK, 창고에 저장." (거버넌스 및 종료)

아래에서는 가상의 자동차 도장 공정에서 페블로심 모듈이 동작하는 워크 플로우를 설명합니다.

3.2. 예제 시나리오 1: 자동차 도장 공정의 미세 스크래치 결함 데이터 확보

배경: 자동차 제조사의 피지컬 AI가 도장면 검사를 수행 중이나, 조명이 어두운 환경에서의 '미세 스크래치' 결함 데이터가 부족하여 탐지율이 떨어지는 상황.

Step 1. 진단 및 처방 (Diagnosis & Prescription)

- 주체: 데이터 그린하우스 (Data Clinic & AADS)
- 수행 내용:
 1. 데이터 클리닉(Data Clinic)이 현재 AI 모델의 학습 데이터를 분석합니다. "정상 데이터는 충분하나, '50 lux 이하 조도'에서의 '미세 스크래치' 데이터가 전체의 1% 미만(ISO Rep ML 1 위반)"이라는 진단을 내립니다.
 2. AADS(오케스트레이터)는 이 진단을 바탕으로 "저조도 환경의 미세 스크래치 데이터를 1,000장 생성하여 밸런스를 맞춰라"는 추상적인 명령(Intent)을 생성합니다.
- 데이터 교환: [진단 리포트 & 데이터 벌크업 명령\(Intent\)](#) → 페블로심(GenSim Manager)

Step 2. 시나리오 설계 및 번역 (Translation & Design)

- 주체: 페블로심 (GenSim Manager)

- 수행 내용:
 1. AADS의 추상적 명령을 수신하여 구체적인 시뮬레이션 스크립트로 번역합니다.
 2. 온톨로지(Ontology)를 참조하여 '저조도'를 '광원 밝기 30~50 lux', '미세 스크래치'를 '깊이 0.1mm, 길이 2cm 이내의 텍스처 패턴'으로 구체화합니다.
 3. 디지털 트윈 엔진에서 '자동차 도장 공정 라인' 에셋을 호출하여 무대를 설정합니다.
- 데이터 교환: 시뮬레이션 스크립트(파라미터 포함) ➔ 페블로심(Multimodal Generator)

Step 3. 가상 환경 구축 및 데이터 생성 (Action & Generation)

- 주체: 페블로심 (Digital Twin Engine + Multimodal Generator)
- 수행 내용:
 1. 디지털 트윈 엔진은 실제 공장의 조명 위치, 컨베이어 벨트 속도, 로봇 팔의 위치 등 물리적 환경을 완벽하게 구현(Reality Sync)합니다.
 2. 멀티모달 생성기는 설계된 스크립트에 따라 수천 번의 시뮬레이션을 수행합니다. 이때 **Vector-to-Param** 기술을 사용하여, 이미 충분한 '밝은 환경' 데이터는 생성하지 않고, 부족한 '어두운 환경' 데이터만 정밀 타격(Precision Targeting)하여 생성합니다.
 3. 단순 이미지뿐만 아니라, 로봇 팔의 센서 로그나 결함 위치 좌표(Annotation) 같은 메타 데이터도 함께 생성합니다.
- 데이터 교환: 합성 데이터셋(이미지 + 로그 + 라벨링 정보) ➔ 페블로심(PebbloScope)

Step 4. 시각화 및 품질 검증 (Visualization & Verification)

- 주체: 페블로심 (PebbloScope) & 사용자(Human-in-the-Loop)
- 수행 내용:
 1. 생성된 데이터가 의도한 대로 '저조도 스크래치' 영역(Vector Space 상의 특정 위치)에 분포하는지 시각화합니다.
 2. **Interactive Link** 기능을 통해, 데이터 분포도 상의 점을 클릭하면 해당 데이터가 'Scratch_LowLight'라는 온톨로지 개념과 연결되어 있음을 보여줍니다.
 3. 사용자(품질 관리자)는 생성된 이미지가 너무 비현실적이거나(예: 스크래치가 너무 인위적임), 표준 범위를 벗어났는지 확인합니다.

- 데이터 교환: 검증된 데이터 리포트 → 데이터 그린하우스(Governance Layer)

Step 5. 거버넌스 승인 및 입고 (Governance & Close)

- 주체: 데이터 그린하우스 (Governance Layer)
- 수행 내용:
 - PebbleScope에서 사용자가 '승인(Approve)' 버튼을 누르면, 이 데이터가 데이터 그린하우스의 품질 표준을 통과했음이 확정됩니다.
 - 데이터는 최종적으로 데이터 레이크(Data Lake)로 입고되어 AI 모델 재학습에 사용됩니다.
 - 이 모든 과정(누가, 어떤 기준으로, 어떤 데이터를 생성했는지)은 감사 로그(Audit Log)로 기록되어 향후 규제 대응(표준 준수 증빙)에 활용됩니다.
- 최종 결과: AI 모델의 결함 탐지를 향상 및 데이터 표준 준수 완료.

Data Flywheel 효과: 가치의 이중 나선 (The Double Helix of Value)

상기 5단계 워크플로우가 데이터 생산의 '실행'을 담당한다면, 이 과정이 반복될수록 시스템의 지능이 강화되는 'Data Flywheel(자가 증식형 선순환)' 효과가 발생합니다. 이는 고객과 페블러스 모두에게 지속적인 가치를 창출합니다.

- 고객 관점: "사용할수록 정교해지는 맞춤형 자산 (Customized Asset)"
 - 고객이 PebbleSim을 운영할수록, 데이터 그린하우스는 단순한 도구가 아니라 고객의 공정에 특화된 '살아있는 자산'으로 진화합니다.
 - 진단 정밀도 향상: 축적된 데이터를 통해 시스템은 고객 공정 특유의 미세한 결함 패턴이나 환경 변수를 더 깊이 이해하게 되며, 이는 시간이 지날수록 더 정확한 '데이터 공백(Void)' 탐지로 이어집니다.,
 - 자산 가치의 증대 (Appreciating Asset): 일반적인 SW는 시간이 지나면 구식이 되지만, 데이터 그린하우스는 데이터와 운영 기록(Audit Log)이 쌓일수록 진단과 생성의 정확도가 높아져, 경쟁사가 모방할 수 없는 고객만의 독보적인 데이터 해자(Moat)를 형성합니다.,
- 페블러스 관점: "핵심 두뇌의 진화 (Core Intelligence Evolution)"
 - PebbleSim이 생성한 고품질 데이터와 운영 경험은 페블러스의 핵심 AI 엔진을 강화하는 '커리큘럼 데이터(Curriculum Data)'가 됩니다.

- AADS 지능 강화: 다양한 산업 현장의 시나리오와 해결 과정은 AADS-LLM과 VLM의 재학습(Self-Training Loop)에 활용되어, 복잡한 물리적 상황을 이해하고 추론하는 능력을 고도화합니다.
- 범용성 확장: 특정 고객사에서 해결한 난제(Edge Case) 해결 능력은 페블러스 전체 플랫폼의 지능으로 흡수되어, 다른 산업이나 공정의 문제를 해결하는 데에도 기여하는 **네트워크 효과(Network Effect)**를 창출합니다.

4. PebbloSim 단계별 개발 전략

4.1. 서론: 점진적 완성을 통한 리스크 관리 및 AADS 2단계 목표 달성

PebbloSim과 같이 복잡하고 혁신적인 플랫폼을 성공적으로 구축하기 위해서는 '빅뱅' 방식의 개발이 아닌, 점진적이고 반복적인 접근이 필수적입니다. 페블러스는 1년간 총 4주기(주기당 3개월)의 PoC를 통해 PebbloSim을 순차적으로 완성하는 개발 전략을 채택합니다.

이 접근법의 핵심은 '웨지 유스 케이스(Wedge Use Case)' 전략입니다. 가장 시급하고 효과가 명확한 자동차 공정(PoC 1)을 통해 즉각적인 ROI를 증명하고, 이를 발판으로 국방(PoC 2), 조선(PoC 3)으로 확장하며 최종적으로 완전 자율화된 플랫폼(PoC 4)을 완성합니다. 각 단계는 데이터 그린하우스와의 연동 깊이를 심화시키며, AADS 2단계 정부 과제의 정량적 목표 달성에 직접적으로 기여하도록 설계되었습니다.

4.2. PoC #1 (1~3개월): 기반 구축 및 자동차 공정 실증

첫 번째 PoC는 "데이터 루프의 연결(Closing the Loop)"에 집중합니다. 기술적 완성도보다는 '진단-처방-생성-검증'의 전체 파이프라인이 끊김 없이 동작하는 것을 증명하는 단계입니다.

- **목표:** 자동차 유연제조 환경 대상 PebbloSim 기본 아키텍처 구축 및 데이터 생성 파이프라인 검증.
- **PebbloSim 핵심 개발 내용:**
 - **디지털 트윈 엔진:** 로봇 팔, 컨베이어 벨트 등 자동차 제조 라인의 기본 물리 환경(Class) 구축.
 - **GenSim 관리자 (MVP):** '로봇 넘어짐', '부품 이탈' 등 2~3개의 고정된 시나리오를 수동으로 선택하여 실행하는 기능 구현.

- **합성 데이터 생성기:** 시각 데이터(RGB 이미지) 위주의 기본 합성 데이터 생성 모듈 개발.
- **데이터 그린하우스 연동 기능:**
 - **Data Clinic (진단):** 시뮬레이션 전/후의 데이터 분포 변화를 비교 분석하는 기본 진단 모듈 연동.
 - **PebbloScope (검증):** 생성된 이미지를 단순 뷰어로 확인하고 저장 여부를 결정하는 기본 UI 제공.
- **검증 시나리오:**
 - "공정 관측 Agent가 '로봇 넘어짐' 데이터 부족을 보고하면, 운영자가 PebbloSim에서 해당 시나리오를 실행하여 데이터 500장을 생성한다. 생성된 데이터가 Data Clinic의 진단을 통과하여 최종 데이터셋에 병합되는가?"
- **AADS 2단계 기여점:**
 - 연차 기술보고서 내 '합성 데이터 생성 파이프라인 실증' 섹션의 핵심 증거 자료 확보.

4.3. PoC #2 (4~6개월): 국방 도메인 확장 및 소버린 체계 검증

두 번째 PoC는 "데이터 주권과 보안(Sovereignty & Security)"에 집중합니다. 외부 클라우드 의존 없이, 폐쇄된 국방망 내부에서 데이터 생성부터 모델 학습까지 완결되는 '독립형 데이터 그린하우스'를 검증합니다.

- **목표:** 국방 감시 경계 도메인 확장 및 외부망 차단 환경에서의 완전한 기능 작동(On-Premise Integrity) 검증.
- **PebbloSim 핵심 개발 내용:**
 - GenSim 관리자 (Defense): 국방 특화 '침투', '배회', '유기' 등 복합 시나리오 에셋 추가 및 작전 교리(Ontology) 기반 시나리오 검증.
 - 거버넌스 모듈 탑재: 모든 시뮬레이션 및 데이터 생성 이력(Who, When, What)을 위변조 불가능한 로그로 기록하여 보안 감사(Audit)에 대응.
- **데이터 그린하우스 연동 기능:**
 - Independent Core: 외부 인터넷 연결(Gemini, Claude, OpenAI 등) 없이 페블러스 자체 소형언어모델(sLLM)과 렌더링 엔진만으로 작동하는 패키징 기술 검증.
 - **검증 시나리오 (Defense Scenario):**

- "외부와 통신이 차단된 작전 통제실(벙커) 환경에서, 적군의 새로운 야간 침투 패턴 데이터를 즉시 생성(Generation)하고, 이 데이터가 내부 보안 규정(ISO/IEC)을 준수하며 승인(Approval)되어 전술 AI 모델 업데이트에 활용되는가?".
- **AADS 2단계 기여점:**
 - 소버린 AI(Sovereign AI) 기술력 입증: 외산 플랫폼 종속 없는 국산 기술 자립 실현.
 - 투명한 수행 과정 보장: 국방 분야에서 필수적인 데이터 처리 과정의 투명성 및 ISO 인증 충족률 달성.

4.4. PoC #3 (7~9개월): 멀티모달 데이터 고도화 및 조선/제조 적용

세 번째 PoC는 "데이터의 깊이(Depth of Data)"에 집중합니다. 단순 이미지를 넘어선 복합 데이터 처리 능력과 비정형 정보를 이해하는 지능형 생성 능력을 고도화합니다.

- **목표:** 3D CAD/센서 융합 멀티모달 처리 능력 확보 및 조선 도메인 적용.
- **PebbloSim 핵심 개발 내용:**
 - **멀티모달 생성기 (Pro):** 선박 3D 도면(CAD)과 크레인/로봇의 센서 로그(IoT)를 결합하여 '시공간 동기화(Spatio-temporal Sync)'된 복합 데이터 생성 엔진 개발.
 - **산업용 VLM 통합:** **VLM(Vision Language Model)**을 탑재하여, 설계 도면 내의 비정형 텍스트 주석(Note)이나 작업 지시서를 해석하고, 이를 시뮬레이션의 물리적 제약 조건(Constraints)으로 자동 변환하여 반영.,
- **데이터 그린하우스 연동 기능:**
 - **Observation Layer (심층 분석):** 이기종 데이터(이미지 vs 센서) 간의 시간적 동기화(Time-Sync) 오류를 탐지하고, 인과관계를 분석하는 고급 품질 진단 기능.
 - **PebbloScope (Interactive):** 핵심 특허 기술인 'Interactive Link' 적용. 사용자가 3D 공간상의 데이터 분포(Vector)를 클릭하면, 연관된 설계 도면의 위치와 부품 정보(Ontology)가 즉시 하이라이트되는 '뉴로-심볼릭 양방향 시각화' 제공.
- **검증 시나리오:**
 - "조선소 블록 조립 공정의 시뮬레이션 중 센서 이상 데이터가 발생했을 때, PebbloScope의 Interactive Link를 통해 해당 센서의 '물리적 위치'와 연관된 '3D CAD 도면 정보'를 즉시 역추적(Back-trace)하여 원인을 규명할 수 있는가?"

- **AADS 2단계 기여점:**

- 산업 특화 멀티모달 VLM 실증 및 고부가가치 제조 시장(조선, 반도체 등) 진입을 위한 '복잡계 데이터(Complex Data) 처리 상용 패키지' 확보.

4.5. PoC #4 (10~12개월): 완전 자율화 및 플랫폼 완성

마지막 PoC는 "자율성(Autonomy)과 연결"의 완성입니다. 인간은 목표(Goal)만 설정하고, AI 에이전트가 데이터 생산과 개선의 전 과정을 주도하는 'Self-Driving Data Ops' 환경을 구축합니다.

- **목표:** AADS Agent 주도의 '완전 자율 운영 루프(Fully Autonomous Loop)' 완성 및 AADS v2.0 플랫폼 출시.
- **PebbloSim 핵심 개발 내용:**
 - Agentic API Gateway: 단순한 RESTful API를 넘어, AADS 에이전트가 PebbloSim을 자신의 손발처럼 제어할 수 있도록 '**상태 조회(State Monitoring)**'와 '**명령 실행(Command Execution)**'이 통합된 에이전트 전용 인터페이스 완성.
 - Vector-to-Param 자동화 엔진: 데이터 공백(Vector Space Void) 좌표를 입력받으면, 이를 채우기 위한 최적의 시뮬레이션 환경 변수(Parameter)로 **자동 역변환(Inverse Mapping)**하는 핵심 특허 기술(US 12,481,720)의 완전 자동화 구현.
- **데이터 그린하우스 연동 기능:**
 - Orchestration Layer (PDIG 루프): Planner Agent가 [계획(Plan) → 진단(Diagnosis) → 개선(Improve) → 증명(Govern)]의 4단계 루프를 스스로 설계하고 실행합니다. 데이터 부족이 감지되면 PebbloSim을 **반복 호출(Retry Loop)**하여 목표 품질에 도달할 때까지 멈추지 않습니다.,
 - Human-in-the-Loop (Smart Gate): 일상적인 개선은 자율적으로 수행하되, 대규모 리소스 소모나 임계치 이상의 데이터 변경 시에만 인간에게 '**승인 요청(Approval Request)**'을 보내는 지능형 협업 모드.,
- **검증 시나리오:**
 - "AADS에게 '야간 주행 데이터 품질 지수 95점 달성'이라는 목표만 부여했을 때, 에이전트가 스스로 취약한 시나리오(예: 비 오는 야간)를 분석하고, PebbloSim을

24시간 가동하여 인간의 개입 없이 목표 점수를 달성하고 리포트를 제출하는가?"

- AADS 2단계 기여점:
 - AADS v1.0 플랫폼의 '완전 자율화(Full Autonomy)' 기능 실증 완료 및 피지컬 AI 데이터 인프라 솔루션으로서의 최종 제품 경쟁력 확보,.

5. 결론: 데이터의 '수집'을 넘어 '재배'와 '증명'의 시대로

PebbloSim은 단순한 가상 환경 시뮬레이터가 아니라, 데이터 그린하우스(Data Greenhouse)라는 운영 체계 위에서 데이터의 결핍을 능동적으로 해결하는 '행동하는 지능(Actionable Intelligence)'입니다. 본 설계서를 통해 정의된 PebbloSim의 전략적 가치와 비전은 다음과 같이 요약됩니다.

5.1. 패러다임의 전환: 수렵(Hunting)에서 경작(Farming)으로

우리는 데이터를 현실 세계에서 우연히 발견하고 '수집'하던 시대에서, 필요한 데이터를 의도적으로 설계하고 '재배(Cultivation)'하는 시대로의 전환을 선언합니다. PebbloSim은 물리적 세계의 위험과 비용, 시간의 제약을 넘어, AI-Ready 데이터를 무한히 공급하는 화수분 역할을 수행하여 피지컬 AI(Physical AI) 산업의 고질적인 '데이터 기근(Data Famine)'을 근본적으로 해결할 것입니다.,

5.2. 비즈니스 임팩트: 성장과 신뢰의 이중주

PebbloSim은 고객에게 두 가지 차원의 결정적인 가치를 제공합니다.

- 성장(Growth)
 - 자가 증식형 선순환(Data Flywheel): PebbloSim은 사용할수록 똑똑해집니다. 축적된 데이터가 진단 모델을 고도화하고, 고도화된 모델이 더 정교한 데이터를 생산하는 선순환을 통해, 고객의 데이터 시스템은 시간이 지날수록 가치가 상승하는 '가치 증대형 자산(Appreciating Asset)'으로 진화합니다.
- 신뢰(Trust)
 - 핵심 기록 시스템(System of Record): 모든 데이터 생성과 개선 과정은 ISO/IEC

5259 및 ISO 42001 표준에 부합하는 감사 가능한 증거(Audit Trail)로 기록됩니다. 이는 규제가 강화되는 글로벌 시장에서 기업이 AI 모델의 안전성과 투명성을 입증하는 '**운영 증거(Operational Evidence)**'가 될 것입니다.,

5.3. 비전: 피지컬 AI 시대를 위한 필수 인프라

결국 PebbloSim은 페블러스가 지향하는 "Makes Data Tangible" 비전을 피지컬 AI 세상에서 실현하는 가장 강력한 실행 도구입니다. 우리는 자동차, 국방, 조선 등 대한민국 주력 산업이 AI와 결합하여 초격차 경쟁력을 확보하는 데 필수적인 '**데이터 인프라(Infrastructure)**'로 자리매김할 것입니다. PebbloSim은 데이터가 흐르는 모든 곳에 신뢰와 가치를 심으며, 다가오는 대전쟁(Great Competition Wars)의 시대에 고객을 승리로 이끄는 핵심 무기가 될 것입니다