

Pebblous Makes Data Tangible

Pebblous.ai

Pebblous

Better Data Makes Better AI

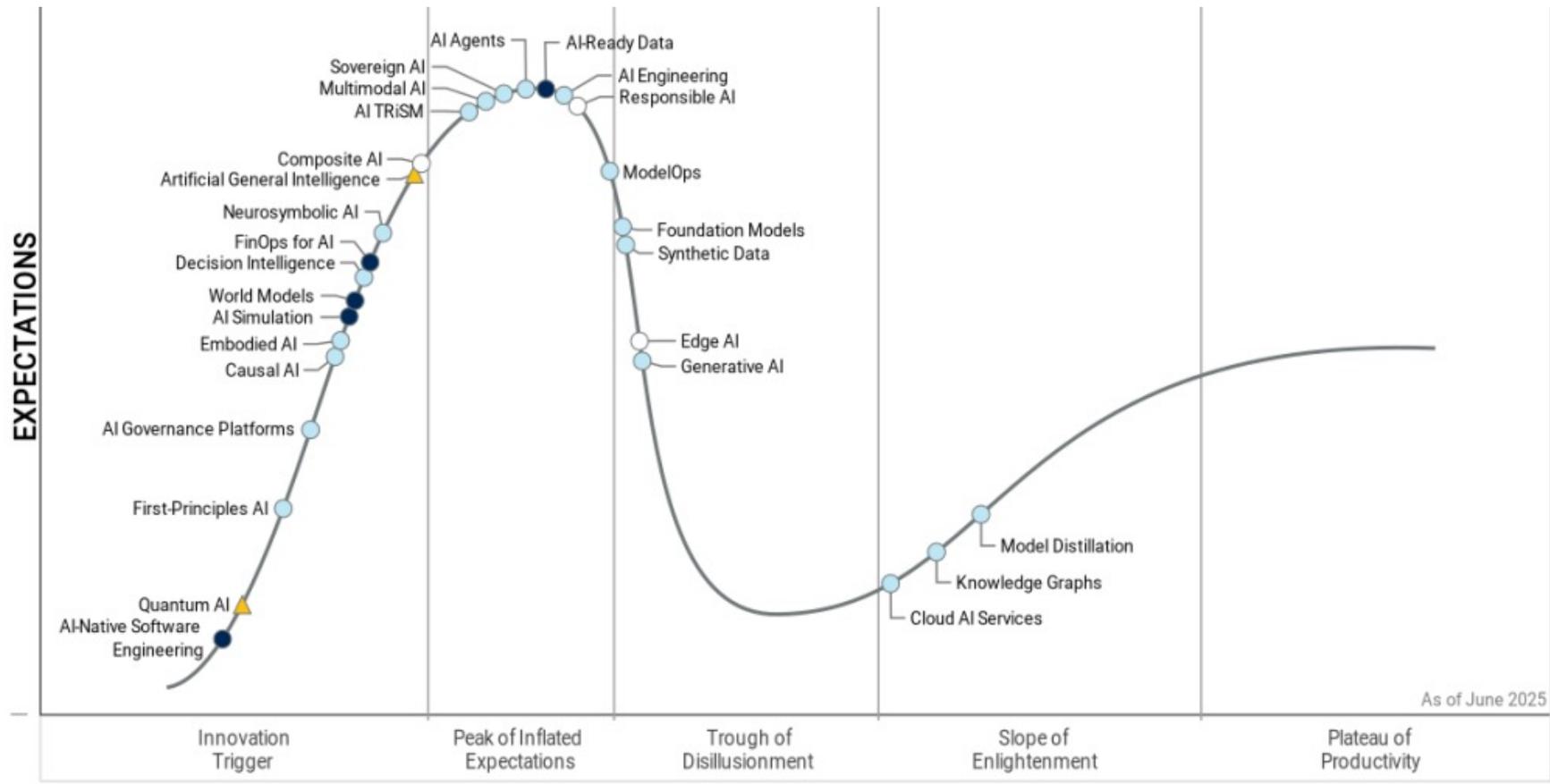


Data Makes **AI**

AI Makes Data

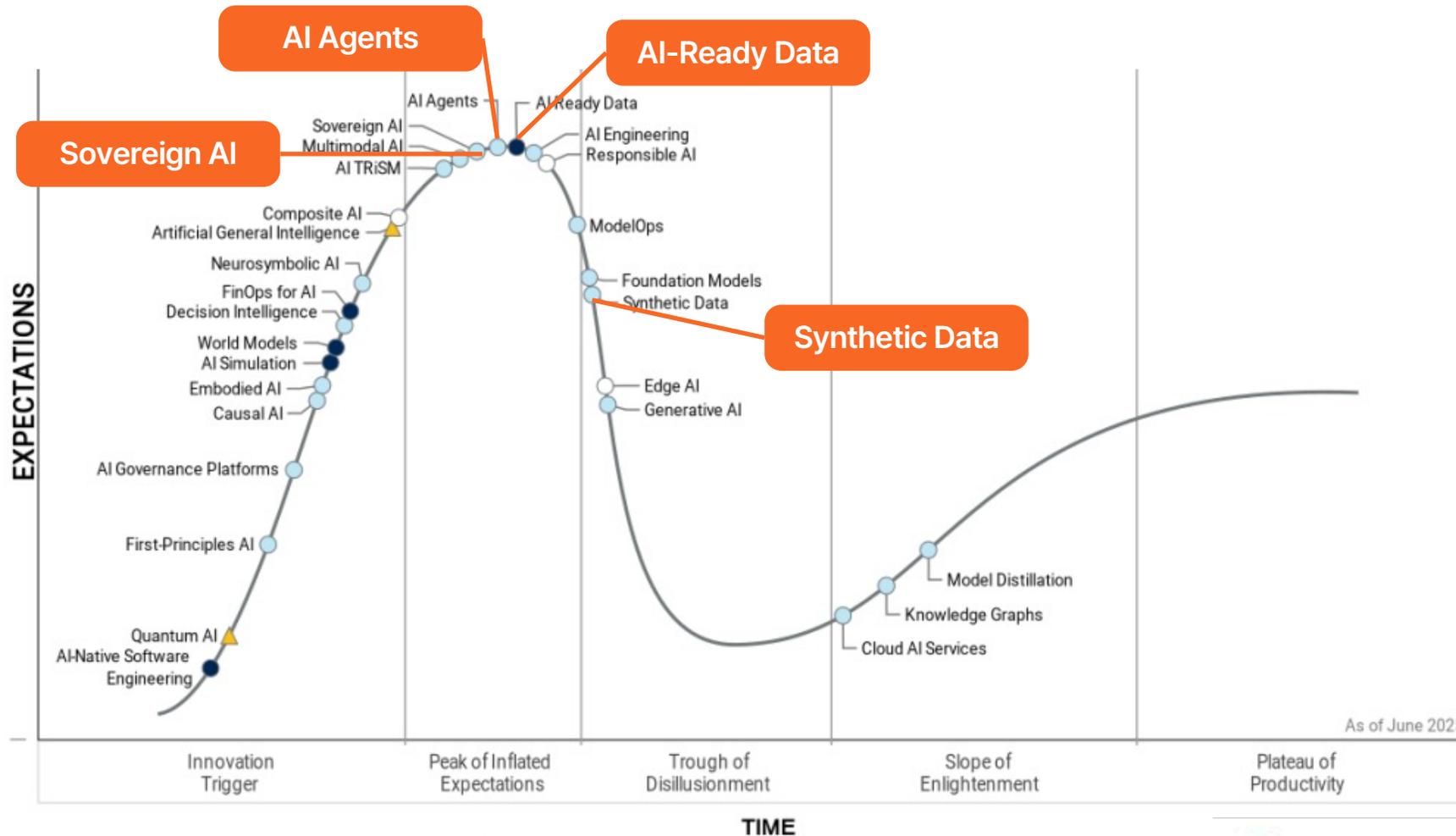
Data Makes **AI** Makes **Data**

Hype Cycle for AI, 2025



Plateau will be reached: ○ <2 yrs. ● 2-5 yrs. ● 5-10 yrs. ▲ >10 yrs. ✗ Obsolete before plateau

Hype Cycle for AI, 2025



Plateau will be reached: ○ <2 yrs. ● 2-5 yrs. ● 5-10 yrs. ▲ >10 yrs. ✗ Obsolete before plateau

AI 데이터셋의 품질로 인한 문제들

모호한 데이터 가격 체계, 모델 성능 저하, GPU 자원 낭비, 높아가는 AI 규제 대응 미흡

01



학습 데이터

품질 불량 / 고갈

02



데이터센터

운영 비용 / 효율

03



AI 관련

법 / 규제 강화

EU AI Act: 전체 매출 7% 또는 3,500만 달러 벌금 부과
한국 2026년 AI기본법 시행

통합적인 데이터 품질 관리 솔루션

데이터의 인공지능 적합성(AI-Ready)을 고려한 지속적인 품질 진단 및 개선

01



AI 학습 데이터
품질 진단/관리

모델 성능 개선

02



데이터 경량화
합성데이터

학습 효율 개선

03



AI 적합 데이터
기준 모니터링

거버넌스 구축

소버린 AI의 시대, 데이터-평가 전면 재구성이 요구된다



”

최예진 교수 | 스탠포드대학교

2025-09-25, UN 안전보장이사회

현재의 AI 모델은 영어에만 치중되어 있고 다른 언어는 성능이 떨어집니다. 그리고 문화적 가치에 있어서도 협소한 이해력을 보이는데, 이로 인해 많은 공동체들이 AI의 혜택에서 소외되고 있습니다.

이 문제는 단편적인 방식으로 해결할 수 없는, 근본적인 문제라고 생각합니다. 따라서 근본부터 바뀌어야 합니다.

학습 데이터, 학습 목표, 평가 방식까지 다 바뀌어야 합니다.

그래야만 다양한 언어, 다양한 맥락에 걸쳐 유능한 AI를 만들 수 있습니다.

History

페블러스 연혁

2021년 Pebblous의 시작은 모래알만큼 작았지만, 지속적으로 고객들의 데이터를 탐구하면서 반짝이고 멋진 조약돌로 성장하고 있습니다.



올인원 데이터 관리 솔루션

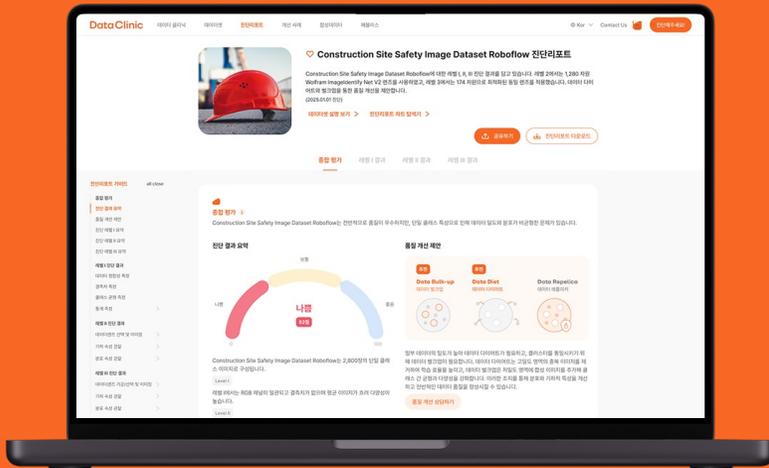
01
품질평가에서
개선까지

02
인터랙티브 가시화
커뮤니케이션

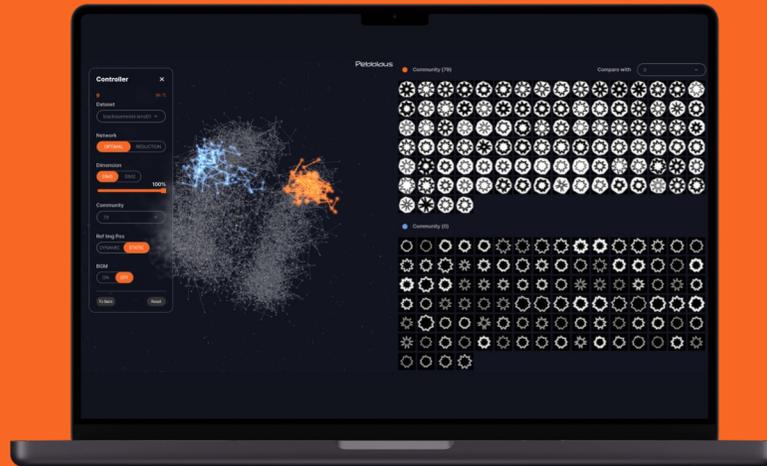
03
멀티모달
데이터셋

04
SaaS,
온프레임, API

Data Clinic



PebbloScope



Synthetic Data



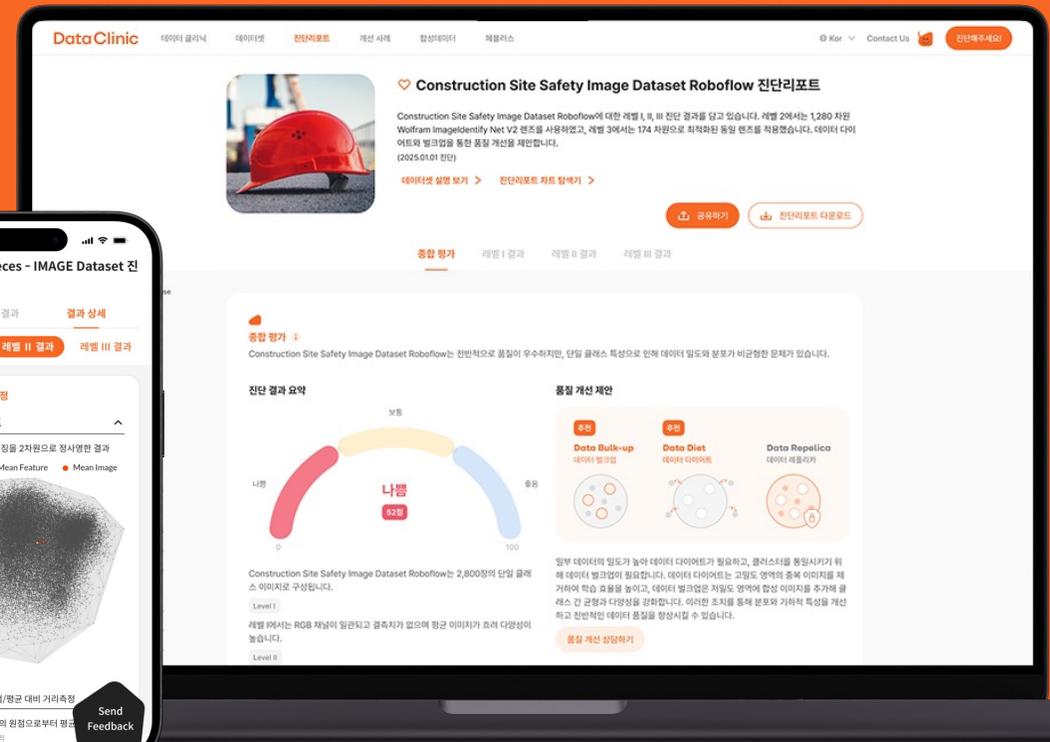
DataClinic

페블러스 데이터 클리닉은 "데이터 종합병원"입니다.

AI 학습 데이터를 위한 품질 평가에서 합성데이터 생성까지의 모든 솔루션을 제공합니다.

웹 버전

PDF 버전



모바일 버전

Product 01

데이터 품질 진단 보고서

Web version

The screenshot shows a web browser window with the URL 'dataclinic.ai'. The page header includes the 'Pebblous DataClinic' logo and navigation links: '소개', '데이터셋', '진단리포트', '개선 사례', 'Contact us', and 'Logout'. Below the header, there are three buttons: '데이터셋 이동', '메거진 이동 163', and '합성데이터 포트폴리오 이동 521'. The main content area features a bird image on the left and a report title 'Birds 450 species - IMAGE Dataset 진단리포트' on the right. The report details include: '데이터셋 주제 / 분야: 생물데이터', '데이터셋 정보: 원본 데이터 | 이미지 | 75,001개', and '진단리포트 정보: 상 | 2023.09.08'. A paragraph of text explains the purpose of the level-based diagnosis. Below this, there are tabs for '소개', '결과', and '결과 상세', with '결과' being the active tab. At the bottom of this section, there are more tabs: '대상 데이터셋 요약', '진단 절차', '진단 결과 요약', '개선 제안', '레벨 I 결과', '레벨 II 결과', and '레벨 III 결과', with '진단 결과 요약' being the active tab.

핵심 요약 Executive Summary

레벨별 진단을 통해 데이터셋의 정합성, 클래스 균형, 중복/유사 이미지 문제를 확인한 후, 경량화 데이터셋의 활용, 클래스별 데이터 밀도 균형화, 구별력 강화를 위한 합성데이터 추가 등을 통해 데이터 개선을 제안한다.

진단 레벨 I 기초 진단 (Basic Diagnosis)

정합성	채널은 양호, 정합성 문제 있음
결측치	특이 사항 없음
클래스 균형	클래스별 데이터 개수 차이가 있음. 모델 학습 시에 주의가 필요함. (train: 150.65±15.69)

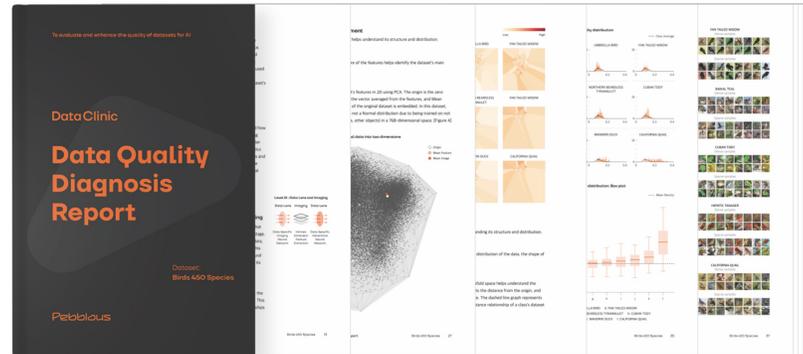
Product 01

데이터 품질 분석 보고서

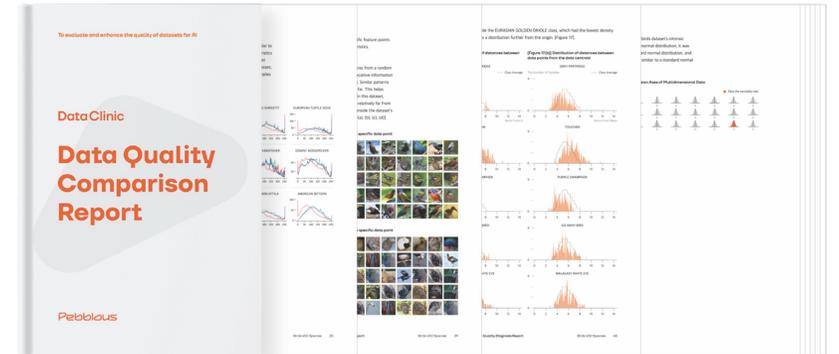
PDF / EPUB version

Pebblous

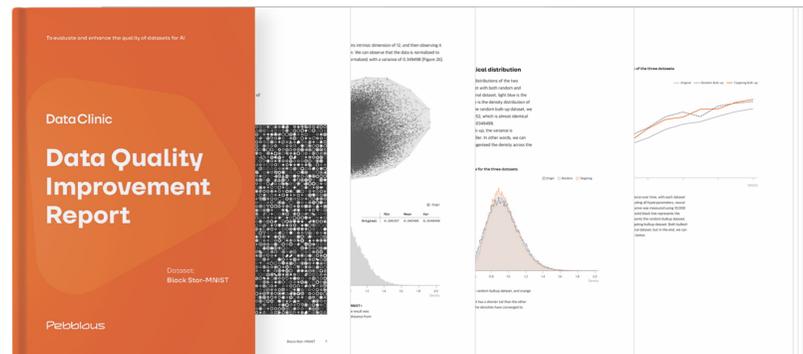
01 데이터 품질 진단 보고서



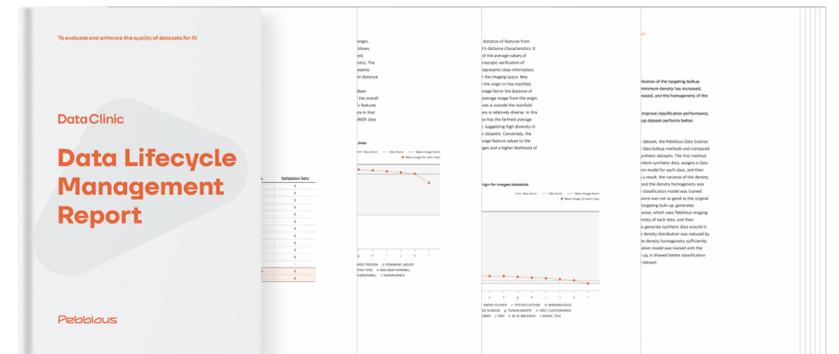
03 출시에정 데이터 품질 비교 보고서



02 데이터 품질 개선 보고서



04 출시에정 데이터 수명주기 관리 보고서



진단보고서 샘플 다운로드
<https://bit.ly/진단보고서>

Product 01

데이터 품질 진단 보고서

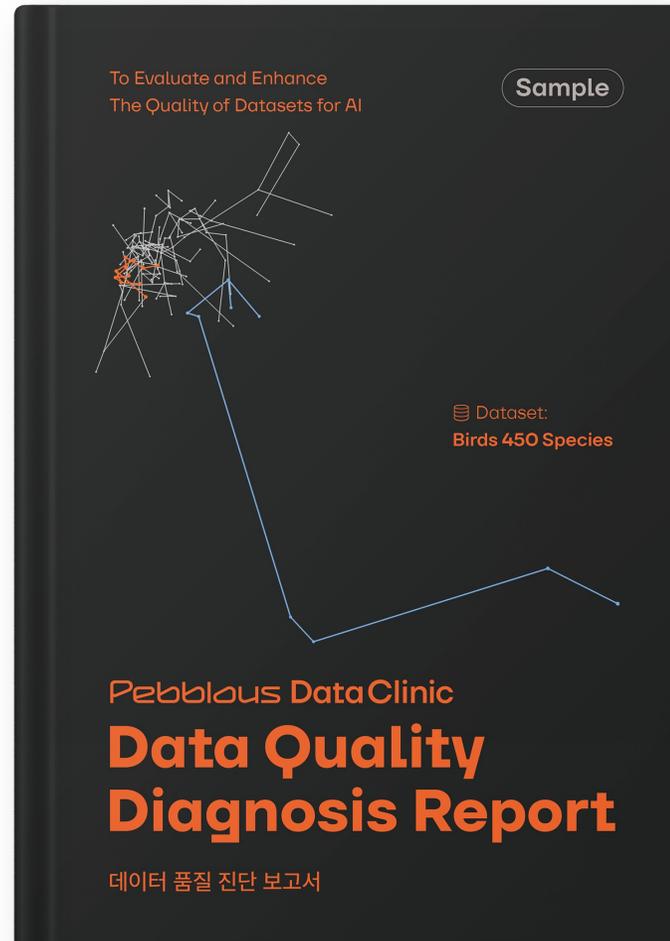
국문 샘플

AI 학습데이터셋에 대한 종합적인 품질 진단
결과와 개선방안이 담긴 보고서입니다.

Pebblous

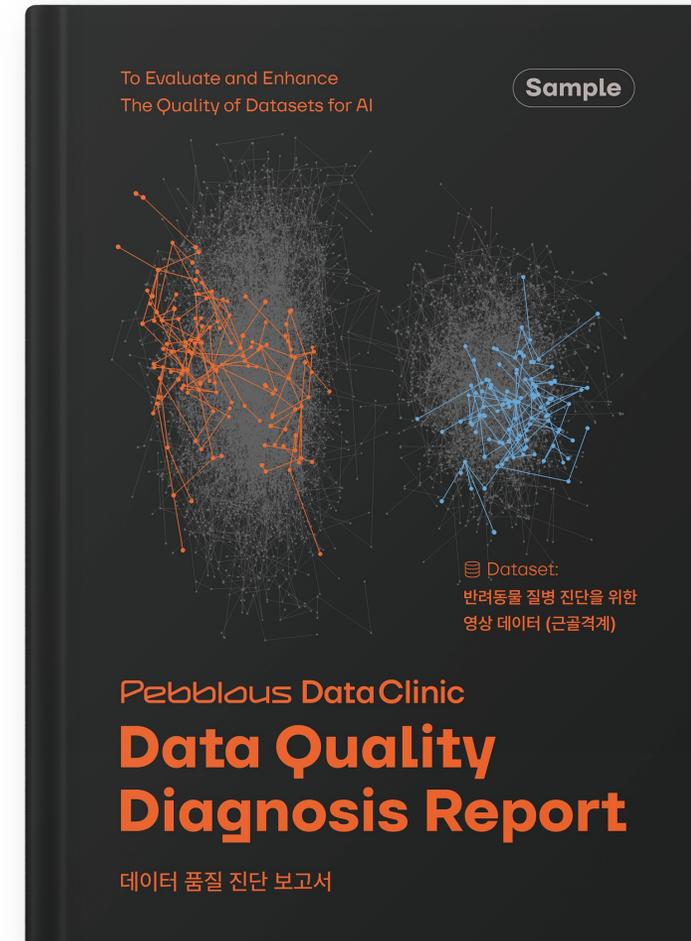
Sample 01 Birds-450 Dataset

450종에 달하는 조류 이미지 공개 데이터셋 (Kaggle)



Sample 02 Pet-bone Dataset

반려동물 근골격계 질병 진단 엑스레이 영상 공개 데이터셋 (AI Hub)



Make Your Data AI-Ready



Source: Gartner
© 2024 Gartner, Inc. and/or its affiliates. All rights reserved. CM_GTS_2952789

가트너 보고서

데이터 품질 관리 대표 기업

생성 인공지능 적합 데이터 (GenAI-Ready Data)를 지원하기 위한 비정형 데이터 관리 역량 개발, Gartner, 2025.

생성 인공지능 적합 데이터(GenAI-Ready Data)의 부족이 생성 인공지능 배포 실패의 주요 원인으로 꼽힙니다. 데이터 관리 소프트웨어 벤더의 제품 관리자들은 내부 개발과 통합 및 파트너십을 통해 비정형 데이터를 관리할 수 있는 기능을 추가해야 고객의 생성 인공지능 구현을 효과적으로 지원할 수 있습니다.



Table 1 : Parameters to Prioritize Partnerships for Unstructured DMS

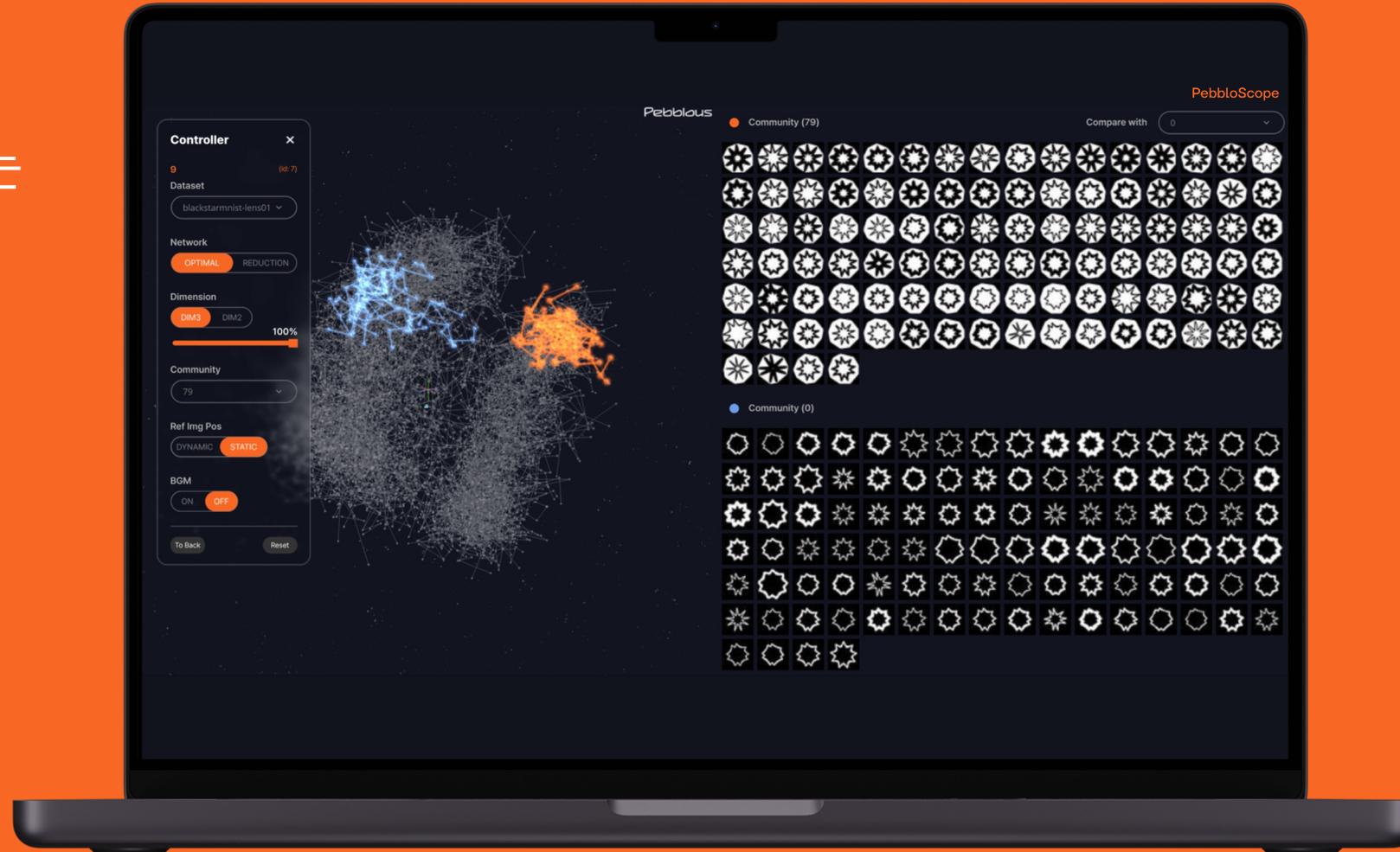
Unstructured DMS market segment	Current end-user demand	Supply of specialized vendors/tools	Degree to which traditional structured DMS vendors support this capability	Specialized unstructured DMS vendors
Data integration	High	Medium	Medium	Bem Iterative.ai, Pryon, Unstructured.io
Data quality	High	Medium	Medium	Anomalo, Pebblous , Shelf.io
Data governance	Medium	Low	Low	BigIS, DryvIQ
Metadata management	Medium	Low	Medium	Instill AI, Labelbox

Source: Gartner

PebbloScope

대용량 다차원 데이터에 대한
3D 인터랙티브 시각화를 지원하는
데이터 커뮤니케이션 도구입니다.

고차원의 데이터를 3차원 공간으로 변환하여 다양한 속성들을 인터랙티브하게
탐색하며 데이터 분석을 위한 인사이트를 얻을 수 있는 데이터 커뮤니케이션
도구입니다.



다양한 도메인의 데이터셋 목록

Controller

Dataset

- bird-lens01
- food11-lens01
- ImageNet50-lens01
- marine-lens01
- ocr4t-lens01
- premierlogo-lens01

Dim 1 100%

Community

1

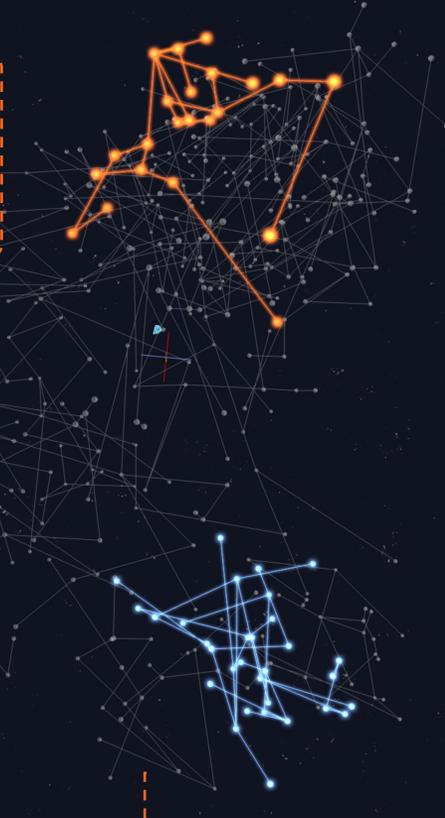
Ref Img Pos

DYNAMIC STATIC

BGM

ON OFF

Reset



시각화 패널

페블러스 데이터 렌즈로 처리된 최적 차원의 데이터를 3D 공간에서 마우스 인터랙션을 통해 탐색할 수 있습니다.

최적 차원 상에서 계산된 데이터 커뮤니티는 유사한 데이터 또는 과밀한 데이터의 분포를 확인하는 데 도움이 됩니다.

Pebblous

Community (1)

Compare with 4

Community (4)

이미지 패널

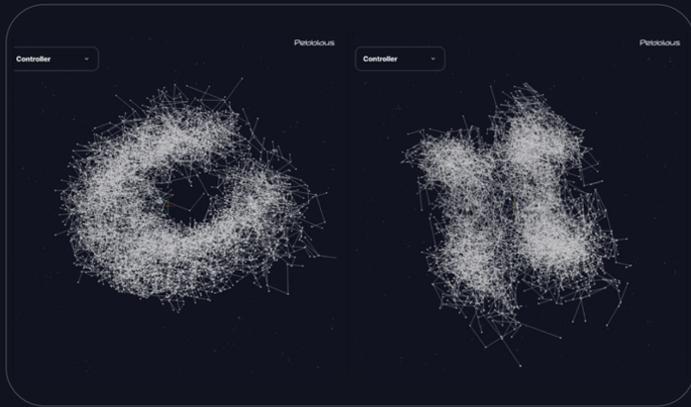
시각화 패널에서 선택한 데이터 및 데이터 커뮤니티의 원본 이미지를 확인할 수 있습니다.

데이터 커뮤니티 간 비교를 통해 각각의 커뮤니티가 갖는 이미지의 특징을 더욱 직관적으로 확인할 수 있습니다.



페블로스코프 데모 영상

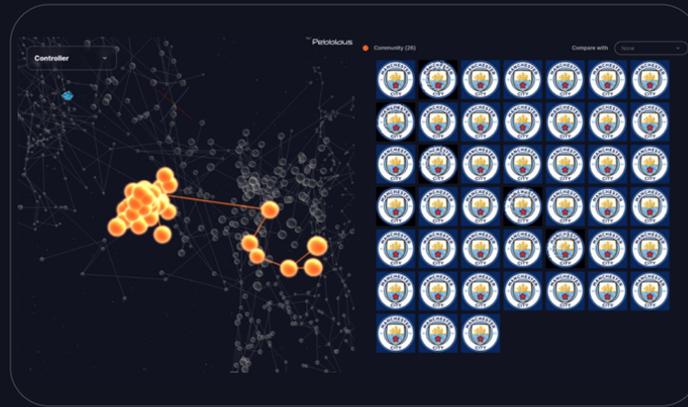
페블로스코프 특징



최적 차원으로 처리된 데이터의 3차원 시각화

페블러스 데이터 렌즈를 통과한 고차원 인공지능 학습데이터셋이 3차원 공간에 가시화 됩니다. 3차원 공간 탐색을 통해 데이터의 고유하며 독특한 분포적 특성을 관찰할 수 있습니다.

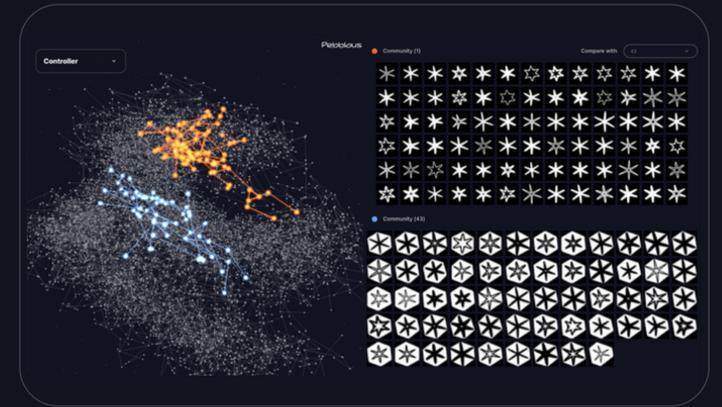
예) 관찰 시점 변경으로 클러스터의 형태 파악



데이터 커뮤니티 탐색을 통한 직관적인 문제점 파악

데이터 커뮤니티는 최적 차원 상의 거리를 기반으로 계산되기 때문에 직관적으로 유사한 데이터를 확인하거나 과밀한 데이터를 파악하는 데 유용합니다.

예) 캐글 로고 데이터셋의 과밀 클러스터의 확인



비교를 통한 데이터 특성 확인

데이터 커뮤니티 간 비교를 통해 각각의 클러스터가 갖는 특징을 확인할 수 있습니다. 독특한 분포를 갖는 데이터의 경우 더욱 심도 있는 인사이트를 얻을 수 있습니다.

예) 동일 클래스 내의 유사 데이터 클러스터 비교

Controller ×

10 (id: 8)

Dataset
ocrai-lens01

Network
OPTIMAL REDUCTION

Dimension
DIM3 DIM2
100%

Community
14

Ref Img Pos
DYNAMIC STATIC

BGM
ON OFF

To Back Reset



Pebblous

Controller ×

10 (id: 8)

Dataset
ocrai-lens01

Network
OPTIMAL REDUCTION

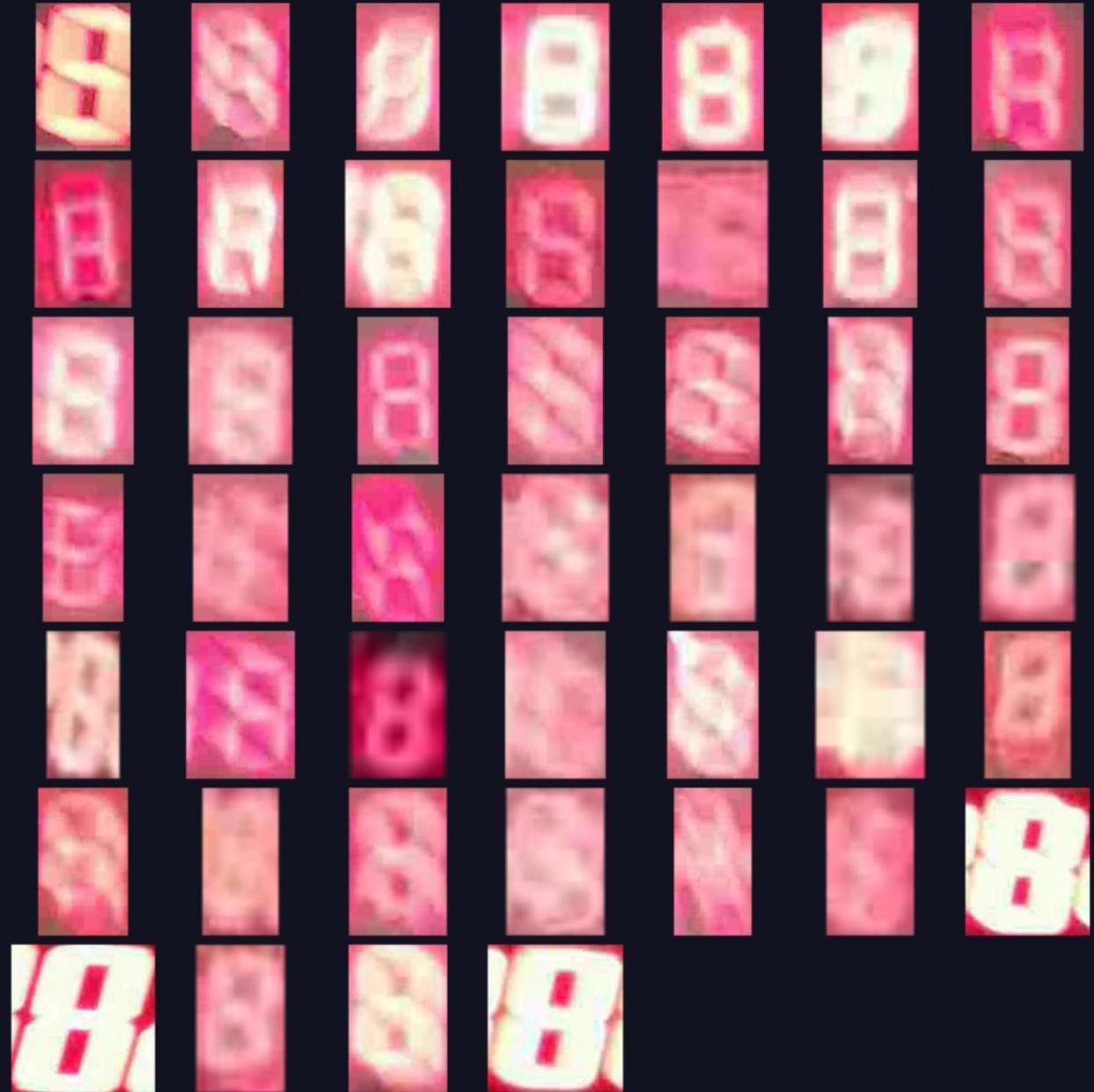
Dimension
DIM3 DIM2
100%

Community
18

Ref Img Pos
DYNAMIC STATIC

BGM
ON OFF

To Back Reset



Pet-bone Dataset

Controller

DOG_Mu03_NOR (cid:19)
Dataset
aihub062-10000

Network
OPTIMAL REDUCTION

Dimension
DIM3 DIM2
100%

Community
53

Ref Img Pos
DYNAMIC STATIC

BGM
ON OFF

To Back Reset

Pebblous

Community (53)

Compare with 0

Community (0)

<https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=501>

Pet-bone Dataset

The screenshot displays the Pebblous interface for the Pet-bone Dataset. On the left, a 'Controller' panel includes the following settings:

- Dataset:** DOG_Mu03_NOR (cid:19) with a dropdown menu showing 'aihub062-10000'.
- Network:** 'OPTIMAL' and 'REDUCTION' buttons.
- Dimension:** 'DIM3' and 'DIM2' buttons, with a progress bar at 100%.
- Community:** A dropdown menu set to '3'.
- Ref Img Pos:** 'DYNAMIC' and 'STATIC' buttons.
- BGM:** 'ON' and 'OFF' buttons.
- Buttons for 'To Back' and 'Reset'.

The central area features a network graph with nodes and edges, where a specific cluster of nodes is highlighted in orange. On the right, a grid of X-ray images shows various views of pet bones. A 'Community (3)' indicator and a 'Compare with' dropdown (set to 'None') are visible above the grid. A URL is provided at the bottom: <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=501>

Menu ⌵

Dataset
WikiArt (Lv3.2_Lens1.10/dH... ⌵

Network
Optimization Reduction

Dimensions
2D 3D

● **Node color mapping**
Normal ⌵

✦ **Selected community**
3-1 ⌵

✦ **Compare community**
3-17 ⌵

Node Tooltips
On Off

Reset



Data preview ⌵

Sort by Density: mAllm (For nearest all da... ⌵

Size 6

✦ **Community 3-17** Sample data (n=30)

Exhibition #16 Tangible Data @ 현대차 제로원 크리에이터 데모데이. 2024.10



Tangible Data:
From Data Nature
To Data Culture

Creator: 김민준, 김민준, 김민준, 김민준

이 전시회를 후원한 제로원 크리에이티브 데모데이 프로젝트
-Tangible Data: From Data Nature To Data Culture-
관객들이 직접 참여하여 데이터를 시각화하고, 이를 통해 데이터를 이해하고 활용할 수 있는 기회를 제공합니다.
관객들이 참여할 수 있는 다양한 체험 프로그램이 마련되어 있습니다.
이 전시는 다양한 연령과 관심사를 가진 관객들이 함께 즐길 수 있도록 기획되었습니다.
-Data Nature- / -Data Culture- / 2024.10.10 - 2024.10.15

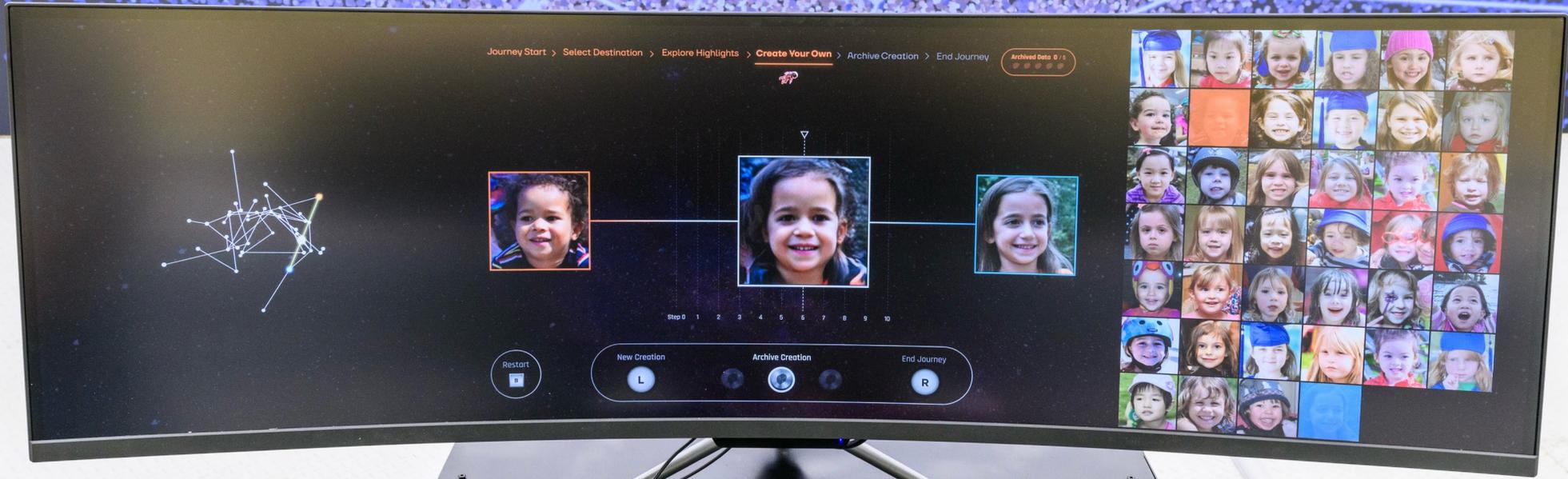
제로원 크리에이티브 데모데이
www.zerowon.com

**Pebblous
Data Art Lab**

Pebblous
Makes Data
Tangible

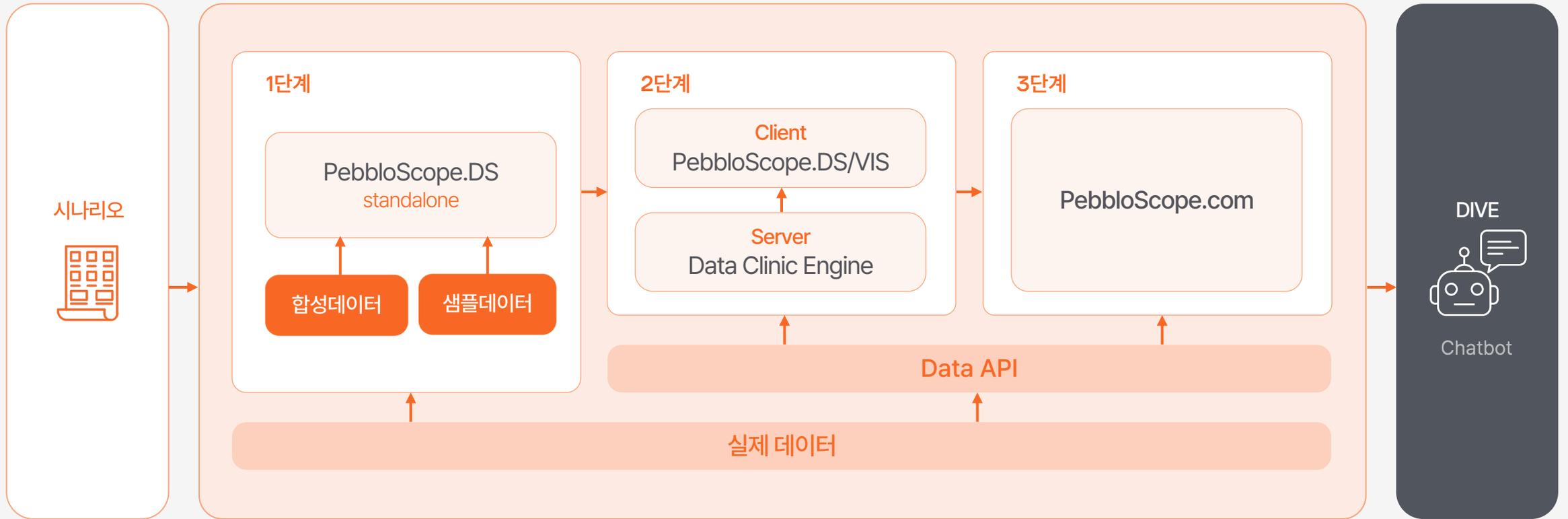




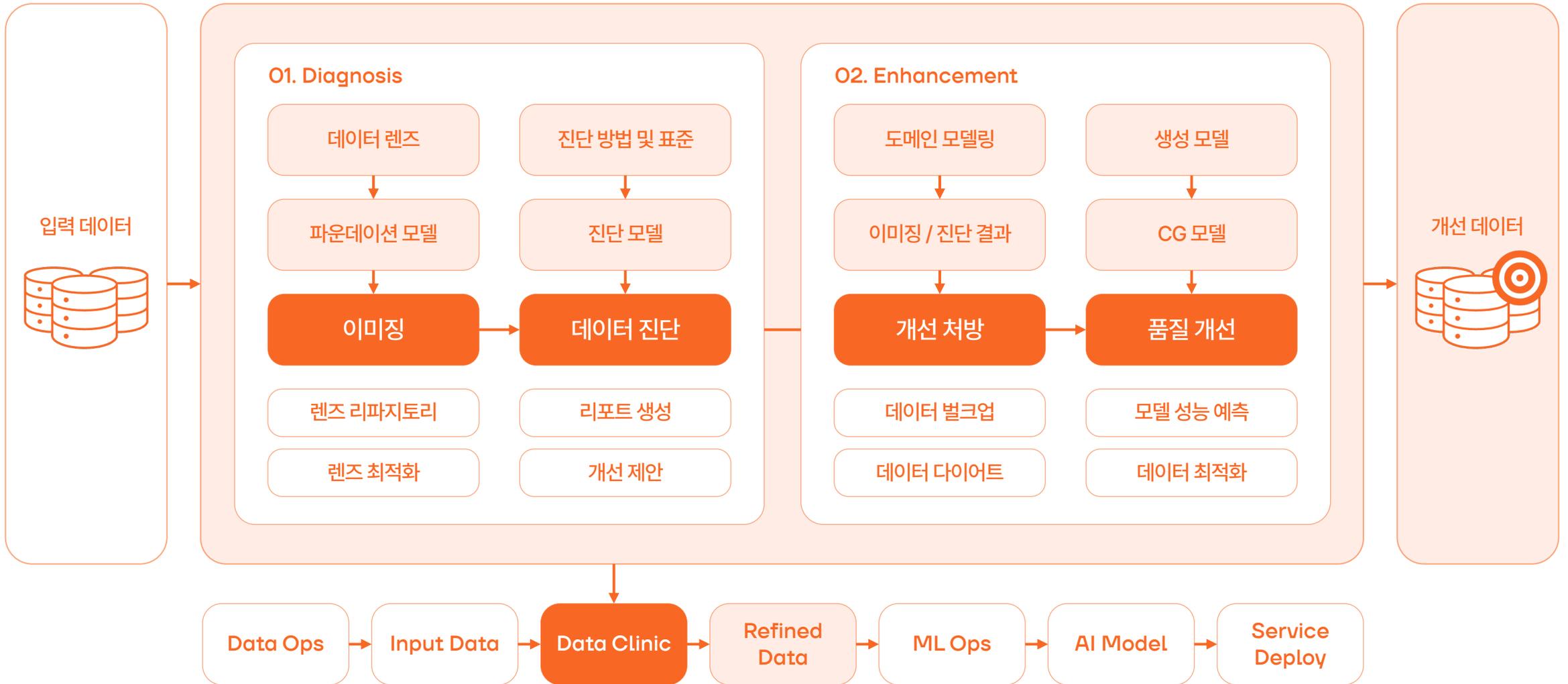




고객 정보 시각화 프로젝트



데이터 품질 진단 및 개선 솔루션

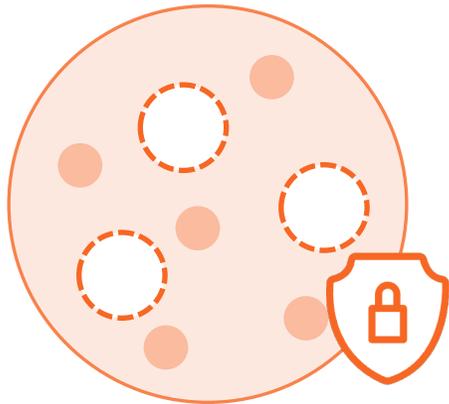


데이터 품질 개선 솔루션

○ 비어있는 데이터 영역 ● 중복 데이터 영역 ● 실제 데이터 ● 합성데이터

데이터 레플리카

원본의 내용을 보호하면서도
통계적·분포적 특성이 유사한 가상데이터를 생성

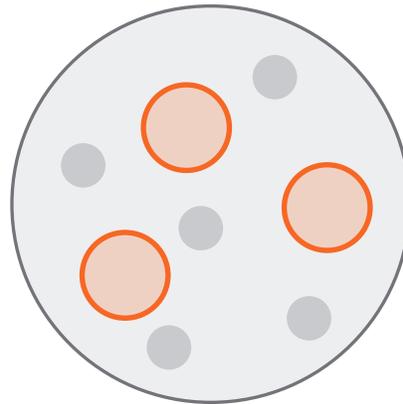


→ 개인정보 유출 리스크 예방

활용 예) 개인정보 보호를 위한 데이터 익명화

데이터 벌크업

데이터가 부족한 부분에 데이터셋에 적합한
정밀 타게팅 합성데이터를 생성하고 제공

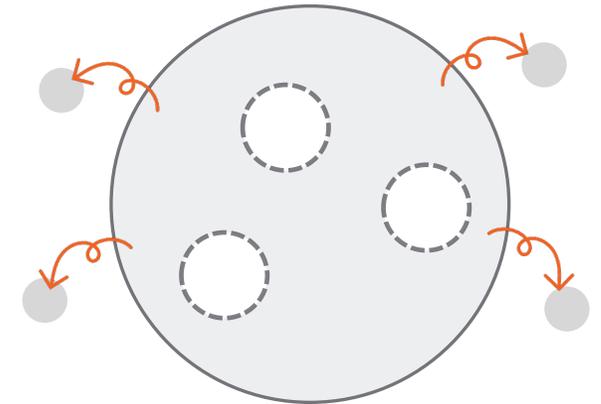


→ 데이터 수집 비용 절감, AI 성능 향상

활용 예) 예외 상황에 대한 합성데이터 생성

데이터 다이어트

AI 모델의 성능을 보장하는 선에서
데이터셋의 중복/유사 데이터를 제거하고 경량화

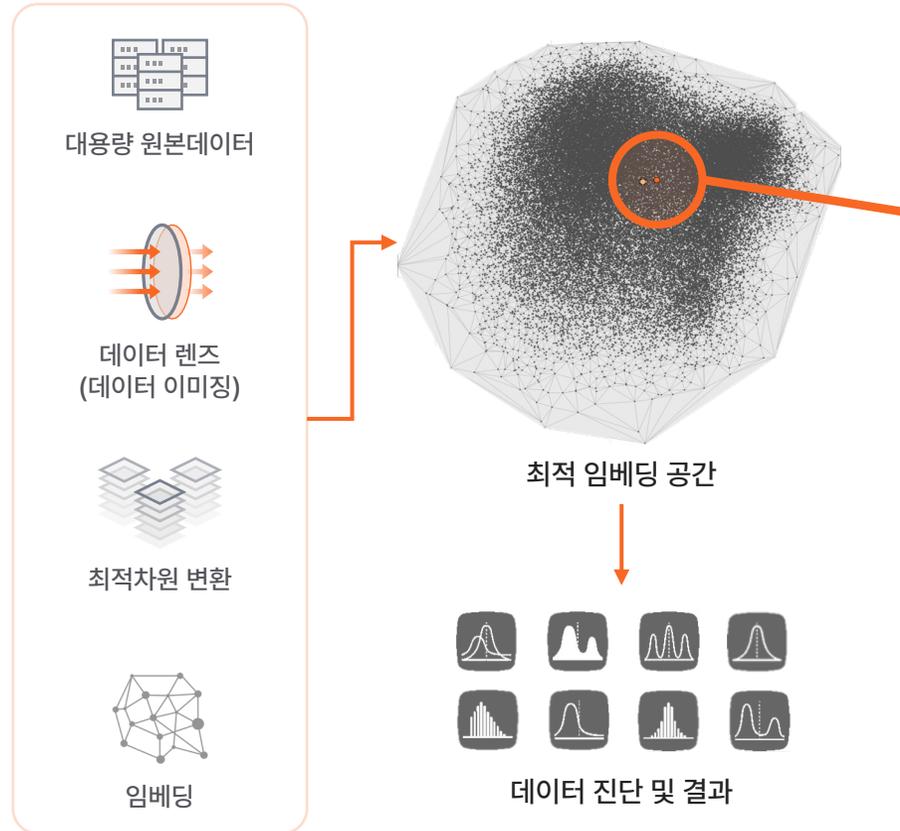


→ AI 학습 비용 절감, 개발 효율 제고

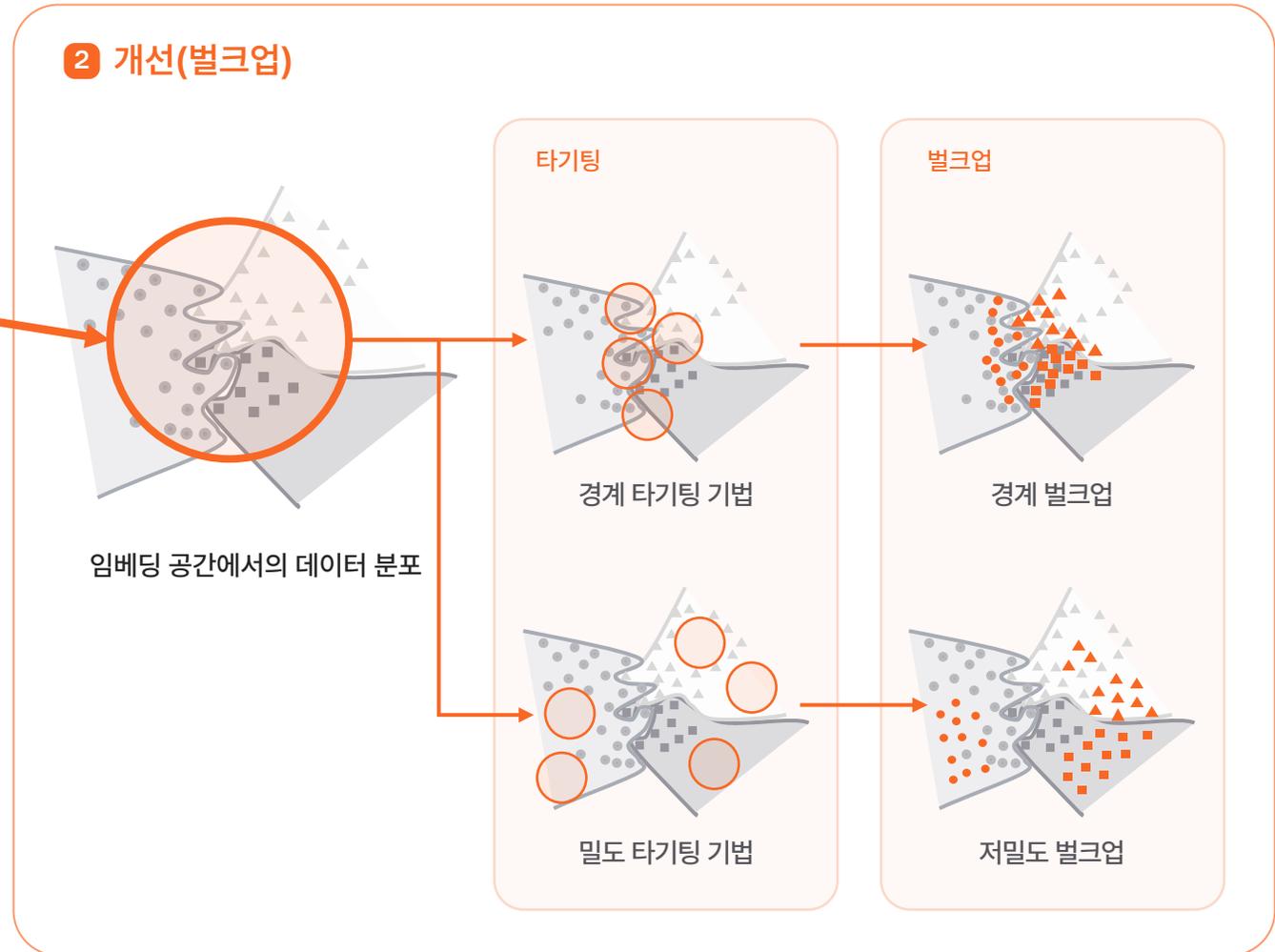
활용 예) 학습 과정에서의 데이터 경량화

정밀 타기팅 합성데이터 생성 기술

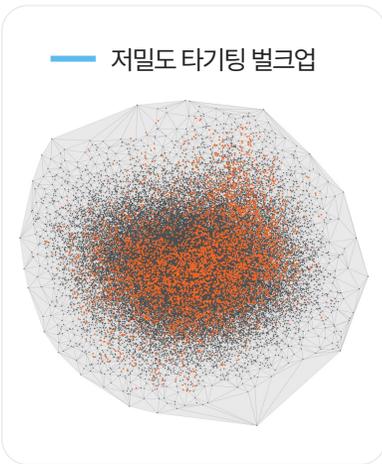
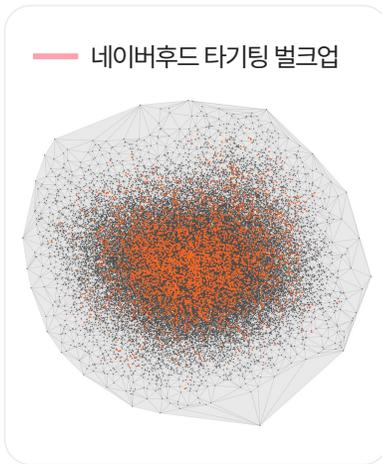
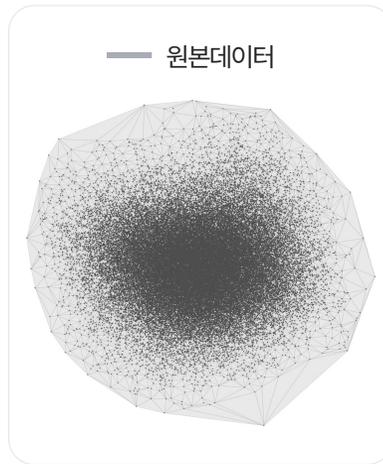
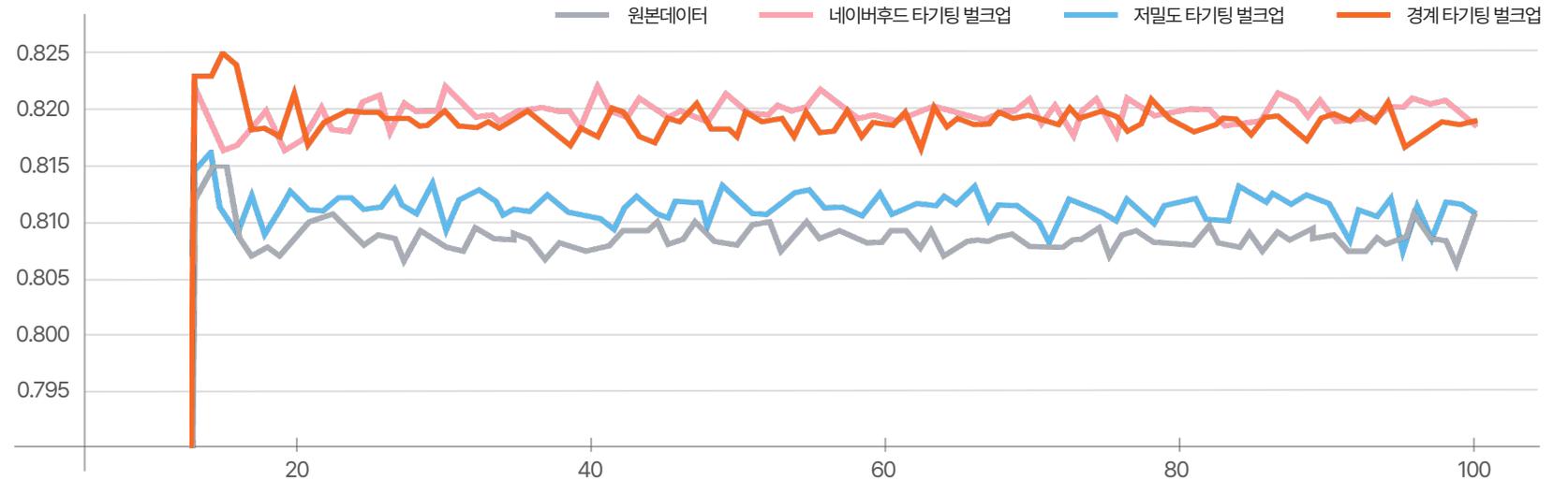
1 데이터 진단



2 개선(벌크업)

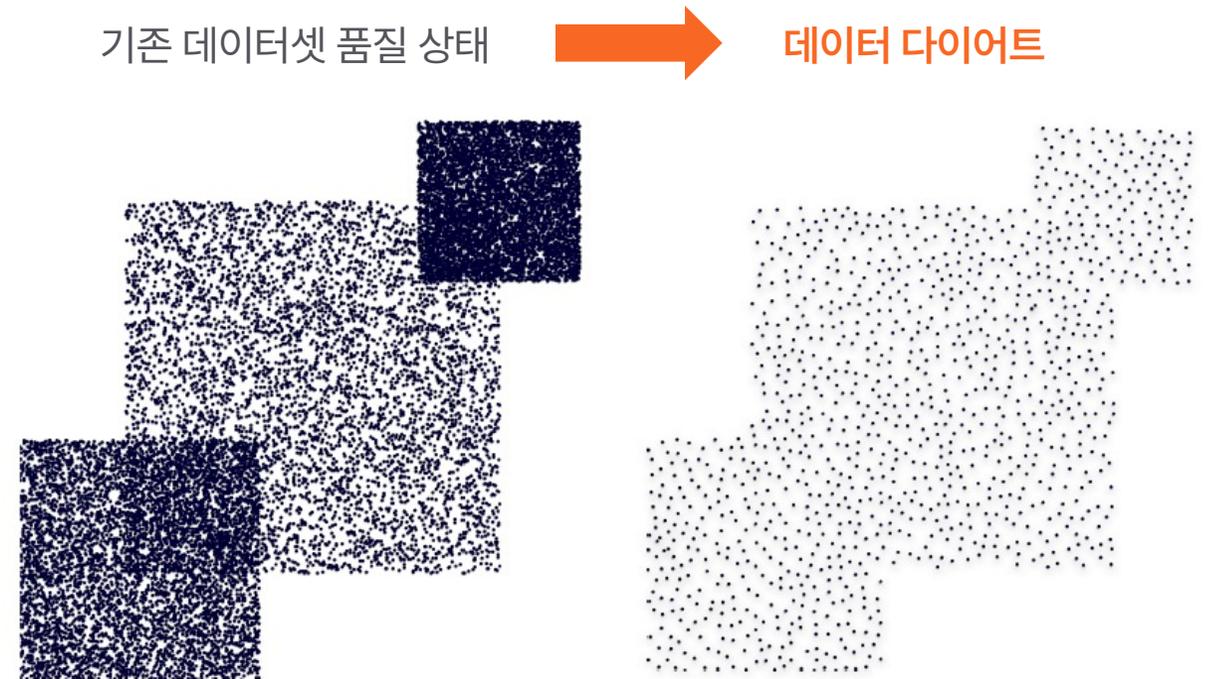
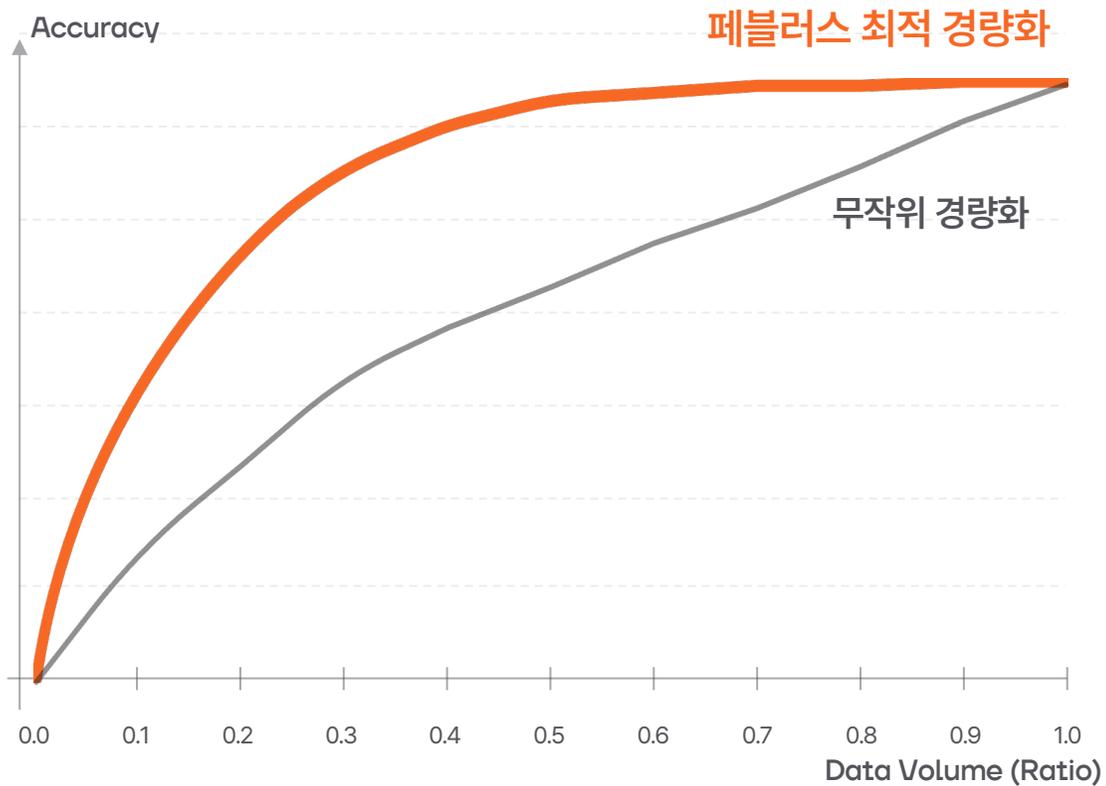


Birds-450 데이터셋에 합성데이터를 추가하여 AI 분류 성능 개선



데이터 다이어트 개념

데이터 최적 경량화 : 과밀 부분의 데이터 개수를 줄임으로써 성능 유지
인공지능 개발 효율 향상 : 학습 시간 및 GPU 비용 절감



데이터 다이어트 사례

MNIST

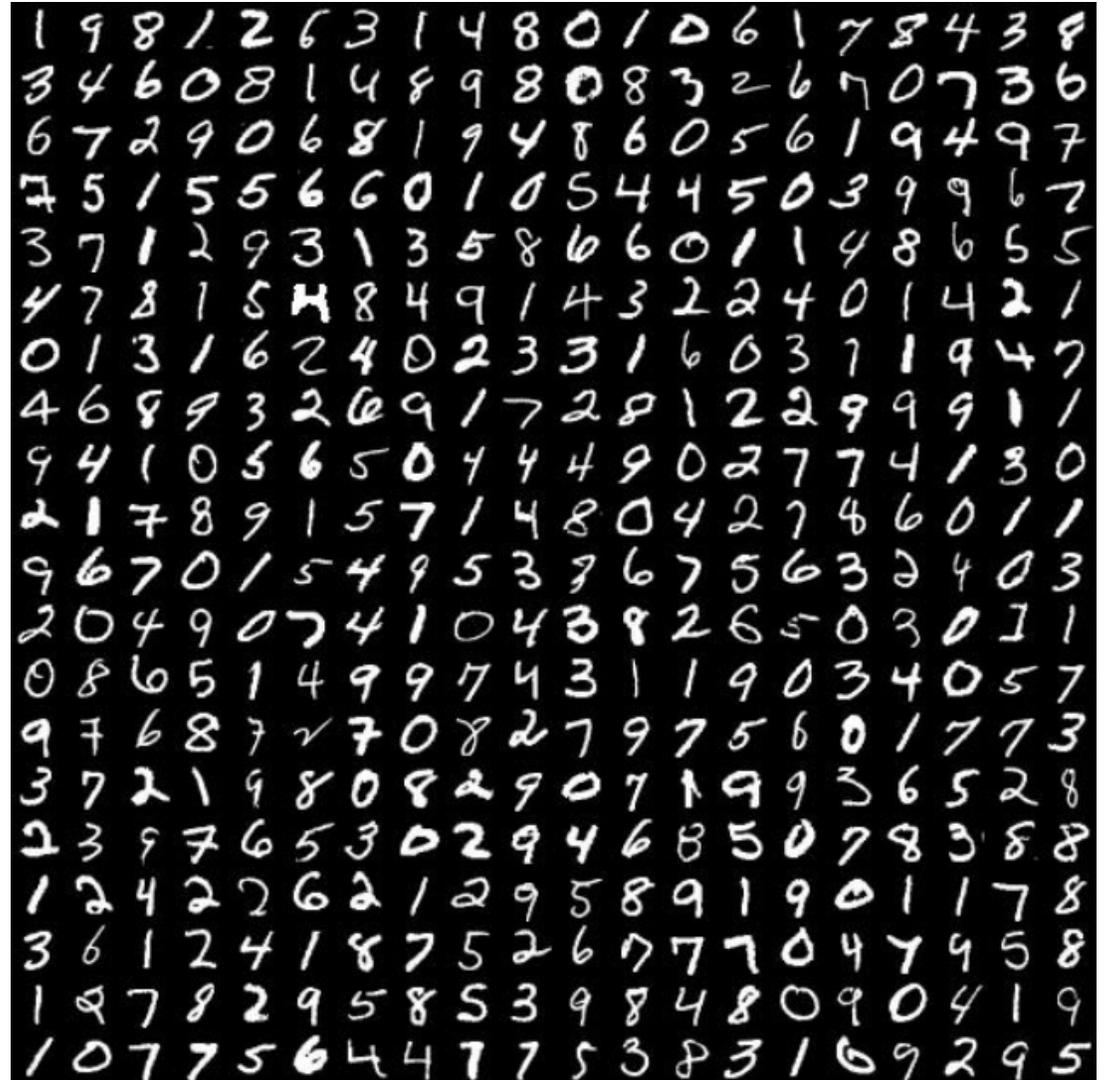


MNIST

Fire

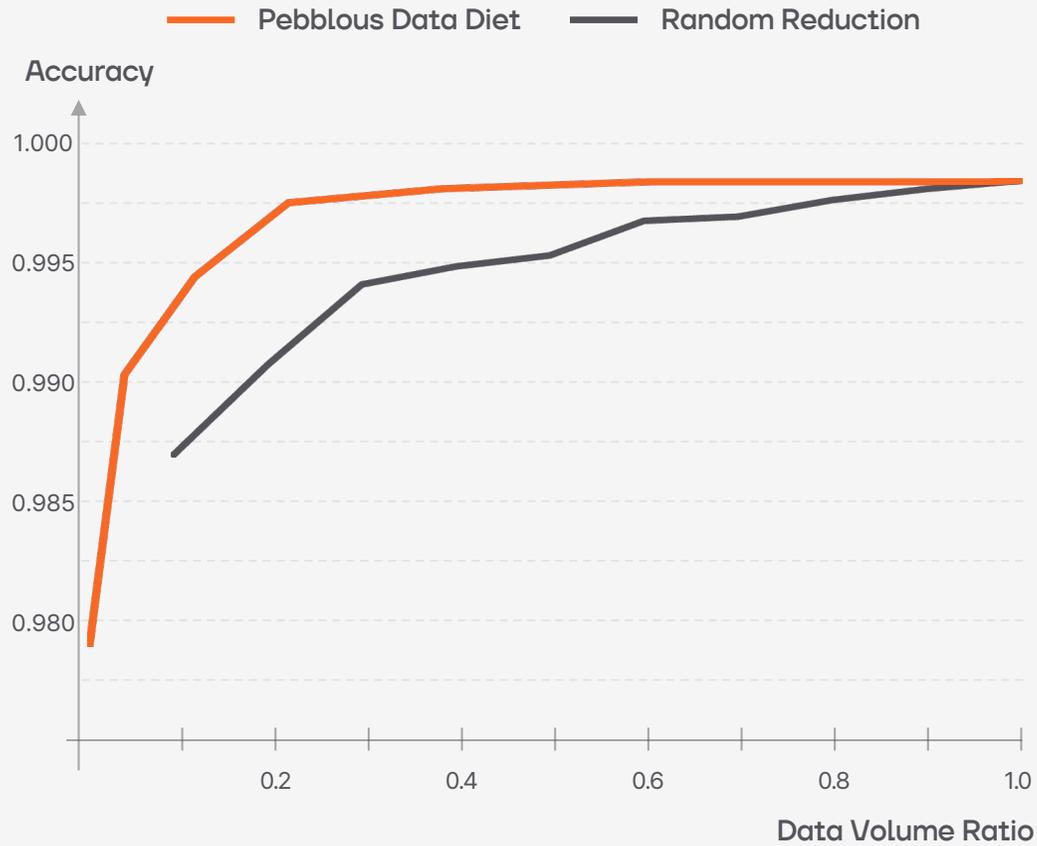
Braintumor

Eyes



데이터 다이어트 사례

Fire



MNIST

Fire

Braintumor

Eyes



Synthetic Data

인공지능 학습을 위해

- ① 데이터 수량이 부족한 경우
- ② 실제 데이터를 구할 수 없는 경우
- ③ 다양한 환경에서의 데이터가 필요한 경우

합성데이터를 제작합니다.



Synthetic Data Portfolio

포트폴리오 웹페이지

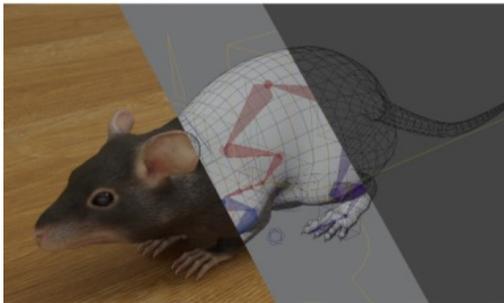
<https://dataclinic.ai/en/synthetic-data>

Characters and poses in special environments



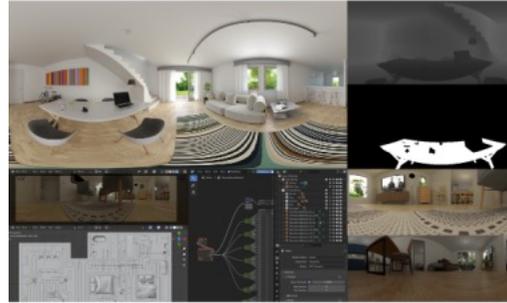
Synthesis data for AI learning for figures and attitude recognition in a special environment, such as children's perception within the vehicle

animal behavior



Synthesis data for AI learning to classify mouse behavior and recognize body parts

Robot autonomous driving field



This is a dataset required for robots such as vacuum cleaners, and AI detects objects and plans optimal driving paths.

Livestock sector



After creating a stable diffusion -based LORA model for analysis of small behavioral forms

Single image 3D creation



Single photo -based 3D model generation synthetic data

* POSTECH joint research, achievement of SOTA levels(SIGGRAPH Asia 2023)

Diet monitoring



Synthetic data for meals for monitoring. Use hybrids of 3D and generated models

Pharmaceutical field

[Download sample images](#)



In the automatic pill automatic preparation system of large hospitals, a computer vision inspection requires a pill dataset of various forms and configuration. Synthesis data for the detection and coefficient of the pill

Drone Detection



3D synthesis data for drone awareness and tracking

Logistics field

[Download sample images](#)



This is a dataset for automatically recognizing the inventory of the stand in an environment such as an unmanned shop.

Waste Plastic Recycling Classification AI Synthetic Data



Synthetic data for AI learning used to recycle resources through plastic material classification

군 경계 감시장비 시 합성데이터

자연 이미지

국방 이미지

자원 재활용

동물 행동

식단
모니터링

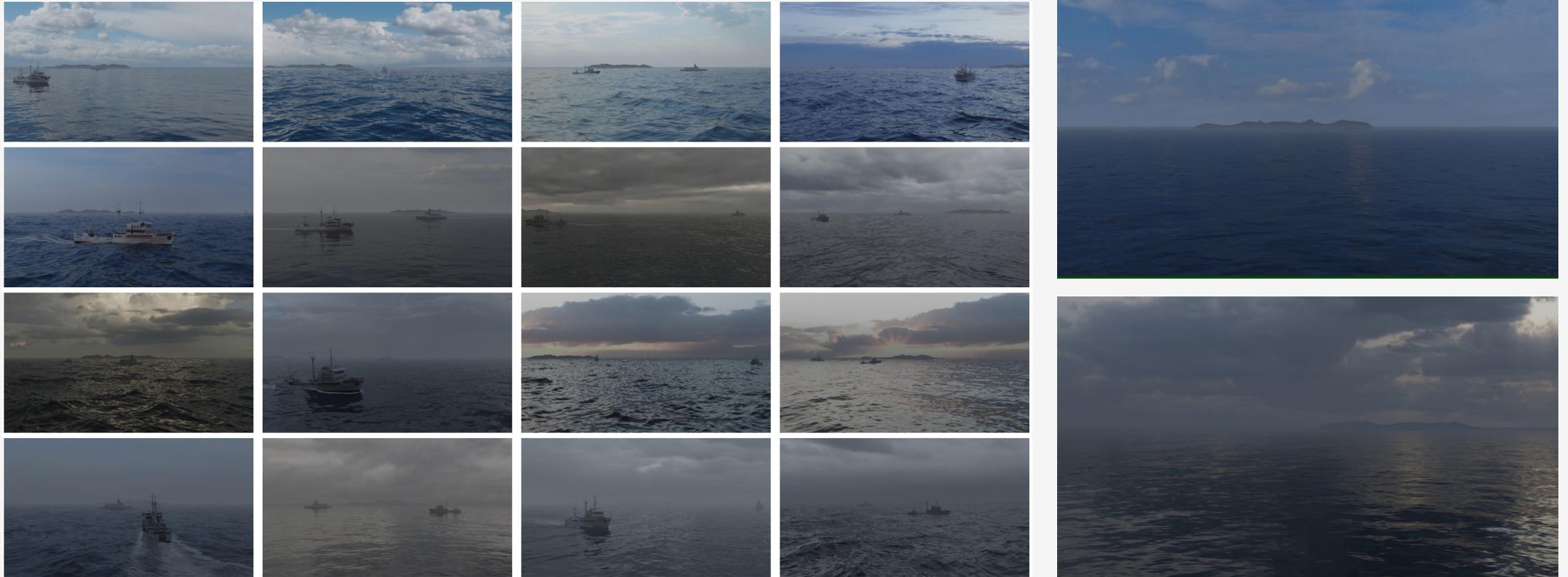
특수환경의
인물 및 자세

축산 분야

단일이미지
3D 생성



해병대 연평도/백령도 경계 시 학습용 합성데이터



NIA 초거대AI 합성데이터: TTA 품질검증 통과

실제 이미지와 합성 이미지를 구분하는 육안 테스트(VTT)에서 구분할 수 없음(50%)에 거의 근접하는 점수 49%로 품질검증 통과함

1. 종합 검증결과

항목별 검증 결과는 아래와 같다. 다만 아래의 결과값은 본 결과서에 기술된 검증환경에서 나온 것으로 검증환경이 달라지거나 데이터 구성 변경 등이 적용되는 실제 환경에서는 결과값이 달라질 수 있다.

[표1] 품질검증 결과 요약

품질특성	항목명	측정 지표	정량 목표	결과값	목표 충족여부			
다양성 (통계)	클래스 분포	구성비	분포 확인	[그림1] 참조	-			
			구분 어선 군함 상선 고정익 유인항공기 회전익 유인항공기 무인항공기 새 배라 오물복탄					
다양성 (요건)	시간대별 분포	구성비 중첩률	구성비 중첩률	50% 이상	구성비 중첩률	96.36%	달성	
			목표 구성비	구분	비율(%)	수량		
			주간	50%	주간	51.86		77,496
			야간	50%	야간	48.14		71,950
	계절별 분포	구성비 중첩률	구성비 중첩률	50% 이상	구성비 중첩률	94.35%	달성	
			목표 구성비	구분	비율(%)	수량		
			하절기	75%	하절기	72.09		107,739
			동절기	25%	동절기	27.91		41,707
	기상별 분포	구성비 중첩률	구성비 중첩률	50% 이상	구성비 중첩률	80.46%	달성	
			목표 구성비	구분	비율(%)	수량		
			맑음	14%	맑음	15.97		23,872
			흐림	14%	흐림	15.98		23,887
비			14%	비	15.98	23,875		
눈			14%	눈	4.28	6,383		
황천 급별 분포	구성비 중첩률	구성비 중첩률	50% 이상	구성비 중첩률	98.00%	달성		
		목표 구성비	구분	비율(%)	수량			
		황천 1급	14%	황천 1급	14.20		21,227	
		황천 2급	14%	황천 2급	14.29		21,355	
		황천 3급	14%	황천 3급	14.29		21,363	
		황천 4급	14%	황천 4급	14.31		21,391	
		황천 5급	14%	황천 5급	14.33		21,422	
		황천 6급	15%	황천 6급	14.28		21,338	
황천 7급	15%	황천 7급	14.29	21,350				

품질특성	항목명	측정 지표	정량 목표	결과값	목표 충족여부
구문 정확성	구조 정확성	정확도	99.5% 이상	100%	달성
	형식 정확성	정확도	99.5% 이상	99.97%	달성
의미 정확성	바운딩박스 정확성	정확도	95% 이상	95.59%	달성
	분류 태그 정확성	정확도	95% 이상	99.89%	달성
	일치성	이미지-캡션 내용 일치성	정확도	95% 이상	97.63%
유효성	Visual Turing Test(VTT)	정확도	30% 이상 ~ 70% 이하	49%	달성
	객체 탐지 성능 (EO)	mAP	65% 이상	89.70%	달성
	객체 탐지 성능 (IR)	mAP	65% 이상	93.60%	달성
	객체 분류 성능 (EO)	정확도	80% 이상	91.28%	달성
	객체 분류 성능 (IR)	정확도	80% 이상	93.24%	달성
	이미지 캡션 생성 (EO)	METEOR	30 이상	39.69	달성
	이미지 캡션 생성 (IR)	METEOR	30 이상	32.85	달성

* 상세결과 및 오류 유형은 별도로 제공하는 자료(엑셀시트)에서 확인 가능

가트너 보고서

Boost Innovation Ecosystem With **Synthetic Data**, Gartner, 2025.

합성데이터로 혁신의 생태계를 부스트하라, Gartner, 2025.

합성데이터는 개인정보 보호와 규제 대응을 가능하게 하면서, 실제 데이터 없이도 AI 훈련과 협업을 촉진하는 핵심 인프라로 떠오르고 있습니다.

공공·헬스케어 등 다양한 분야에서 빠르게 확산 중이며, 가트너는 2030년까지 합성데이터가 실제 데이터를 대체할 것이라 전망합니다. CIO들은 품질 검증과 기술 파트너 선정에 주목해 혁신 생태계 조성에 나서야 합니다.

Mostly AI	2017	Synthetic data for AI/ML training, software testing and analytics
NayaOne	2019	Synthetic data to test, train and model product for financial services industry
Parallel Domain	2017	Synthetic data for autonomy-related use cases, including training and testing of autonomous vehicles, robots, and drones
Particle Health	2018	Synthetic data solution and healthcare insight tools for data analysis and prediction
Pebblous	2021	Synthetic data for diverse computer vision training applications including pose estimation, object detection
Rendered.ai	2019	Synthetic data for training computer vision systems in industries such as defense, earth intelligence, manufacturing and logistic facility

Gartner, Inc. | G00826791

page 2A of 5A

This research note is restricted to the personal use of joohaeng@pebblous.ai.



가트너 보고서

Emerging Tech: Techscape for Startups in Synthetic Data, Gartner, 2025.

합성데이터 분야 스타트업을 위한 기술 지형도, Gartner, 2025.

페블러스는 합성 데이터 솔루션과 함께 데이터클리닉 서비스를 제공하여, 다양한 컴퓨터 비전 학습 응용을 위한 합성데이터를 생성할 수 있도록 정밀한 품질 진단을 수행합니다.

이러한 시뮬레이션을 인간과 동물의 행동까지 확장하여, 데이터 품질 진단을 기반으로 정밀 타기팅 합성 데이터를 생성함으로써 상호작용에 대한 독특한 관점을 제시합니다.

Elctric Twin offers a solution for simulating synthetic human populations, enabling the prediction of human attitudes and behaviors.

Narnia Labs produces synthetically generated product designs for the manufacturing industry.

Pebblous offers a synthetic data solution paired with a data clinic service, delivering quality diagnostics to produce targeted synthetic data for diverse computer vision training applications.

Domain-Specific Synthetic Data Startups

MDCIone provides patient medical analysis and prediction based on artificial data to overcome privacy and security risks.

Data Clinic | 가치 제안

Fast and Accurate

01

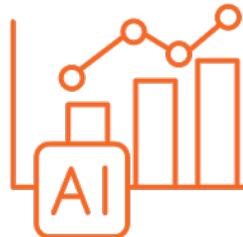


10만개 데이터셋 기준
1시간 쾌속 품질 평가

AI 훈련 데이터셋의 품질 평가 및 시각화

AI Performance

02



5% 합성데이터로
2% 성능 향상

품질 진단 기반 정밀 타기팅 합성데이터 생성

Dev Efficiency

03



80% 데이터 경량화로
x5 GPU 효율 향상

중복 데이터를 줄여 학습 효율 최적화

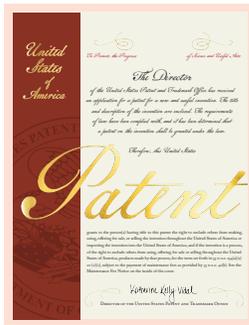
국내외 지식재산권 확보

데이터 진단

데이터 클리닉 방법, 데이터 클리닉 방법이 저장된 컴퓨터 프로그램 및 데이터 클리닉 방법을 수행하는 컴퓨팅 장치



2022-06-29



2024-01-09

출원 11건 / 등록 9건
국내 16건 / 해외 3건 / PCT 1건

데이터클리닉 아키텍처

데이터 클리닉 방법, 데이터 클리닉 방법을 저장하는 컴퓨터 프로그램과 데이터 클리닉 방법을 행하기 위한 컴퓨팅 디바이스



2025-02-05



2023-09-05

출원 4건 / 등록 2건
국내 2건 / 해외 3건 / PCT 1건

합성데이터

합성 데이터 생성을 위한 가상 환경을 제공하는 전자 장치, 전자 장치의 동작 방법 및 전자 장치를 포함하는 시스템



2023-04-03

출원 15건 / 등록 8건
국내 16건 / 해외 3건 / PCT 7건

에이전트

데이터의 현상 및 맥락을 이해하는 AGENTIC AI 기반 품질 평가 방법, 장치 및 시스템



2025-11-12

출원 3건
국내 3건

데이터클리닉 적용 도메인

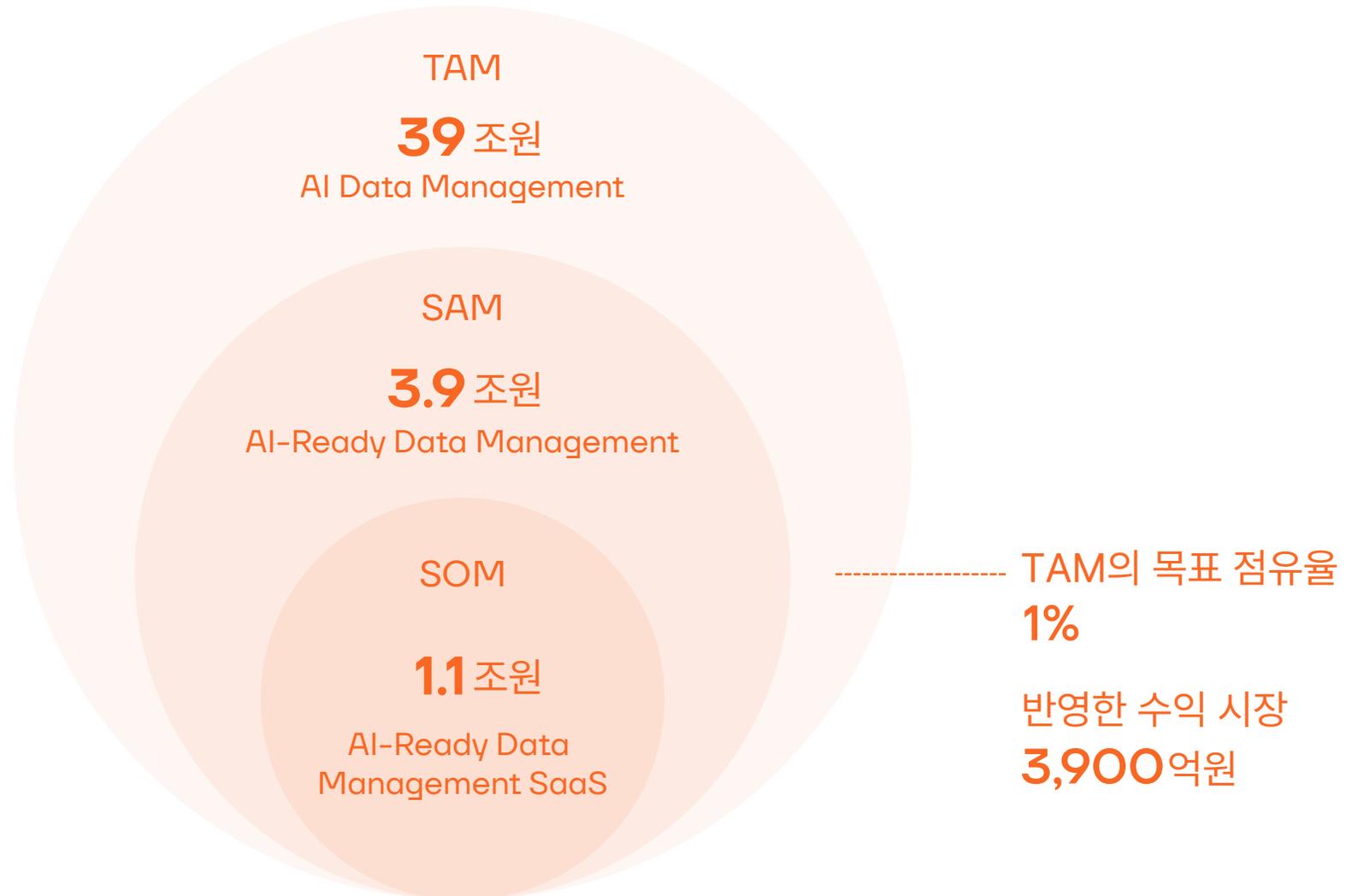


인공지능 데이터 클리닉

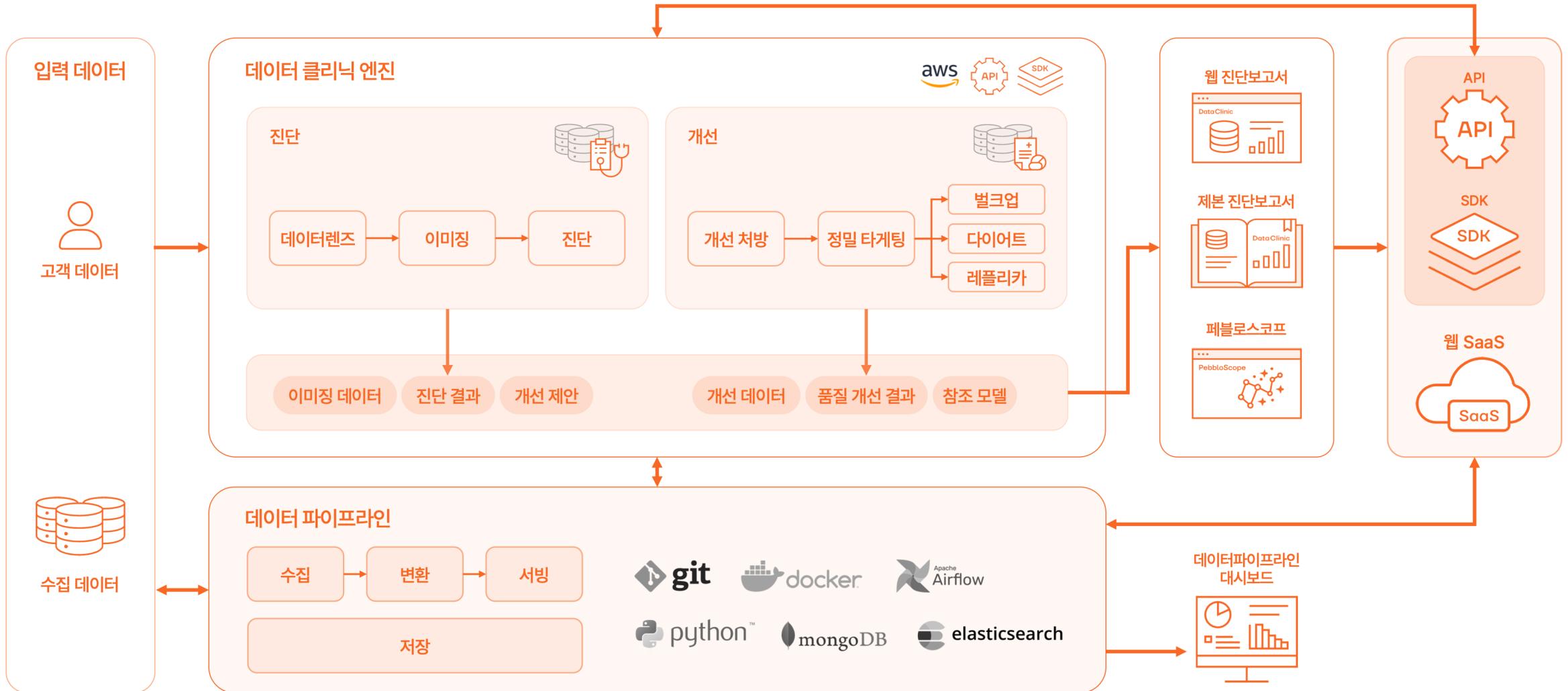
Data Clinic

Four icons representing data clinic services: a stack of data disks with 'AI' (데이터 이미징), a database cylinder with a stethoscope (데이터 품질 진단), a document with a bar chart (개선 처방), and a database cylinder with a plus sign and pill (품질 개선).

AI-Ready Data Management 시장 타기팅



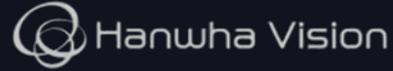
데이터 클리닉 SaaS 모델



데이터클리닉 가격 정책

	Free	Basic	Pro 페블러스 추천	Enterprise
	데이터 클리닉과 페블로스코프의 기본 기능을 체험할 수 있음	퍼블릭 데이터로 한정하여 데이터 클리닉과 페블로스코프의 고급 기능을 사용할 수 있음	고객의 데이터에 대해서 본격적인 데이터 클리닉과 페블로스코프 활용이 가능함	데이터 다이어트와 벌크업 등 데이터 클리닉의 품질 개선 기능과 기타 고급 프로젝트 가능함
	무료 체험 가능	1만원 /월 10만원 /년	50만원 /월 500만원 /년	500만원 /월 5,000만원 /년
데이터클리닉 웹	✓ 퍼블릭 데이터	✓ 퍼블릭 데이터	✓ 퍼블릭/프라이빗 데이터	✓ 퍼블릭/프라이빗 데이터
페블로스코프 웹	✓ 퍼블릭 데이터	✓ 퍼블릭 데이터	✓ 퍼블릭/프라이빗 데이터	✓ 퍼블릭/프라이빗 데이터
고급 가시화 및 인터랙션		✓ 웹 진단리포트	✓ 웹 진단리포트	✓ 웹 진단리포트
페블로스코프 스냅샷 생성		✓ 10회/월	✓ 100회/월	✓ 1,000회/월
합성데이터 체험		✓ 10회/월	✓ 100회/월	✓ 1,000회/월
Level II 데이터 진단			✓ 최대 100K 이미지	✓ 최대 1M 이미지
페블로스코프 생성			✓ 최대 100K 이미지	✓ 최대 1M 이미지
Level III 데이터 진단				✓ 최대 1M 이미지
진단 결과 다운로드				✓ 중요도, 각종 차트 등
PDF 리포트 및 인쇄				✓ PDF 리포트 150만원/건당, 인쇄 서비스 300만원/건당
데이터 다이어트				✓ 100K 이미지 기준 1천만원
데이터 벌크업				✓ 커스텀 프로젝트 진행, 최대 10K 이미지, 주석포함 기준 6천만원
온프레미스 설치				✓ 고객사 내부 사용 기준. 5억원/년. HW 별도

페블러스 주요 고객사



페블러스 주요 고객사

Narae CnD



BiG WAVE



InTheTech

Syntherixs



Rootlab

OceanLightAI



CIP Energy



SPHERE AX



우리로봇



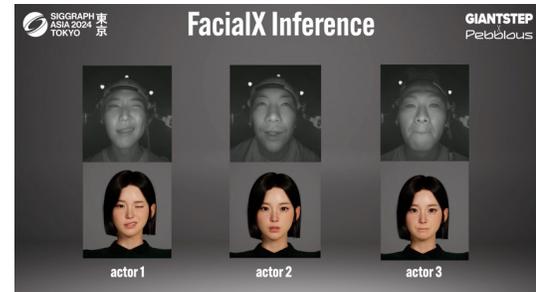
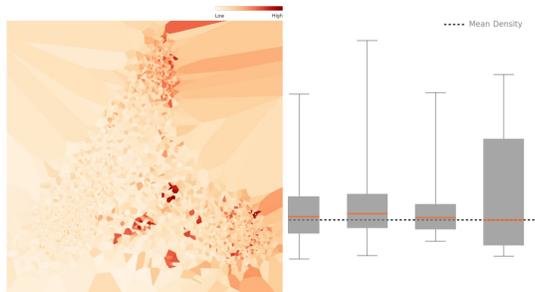
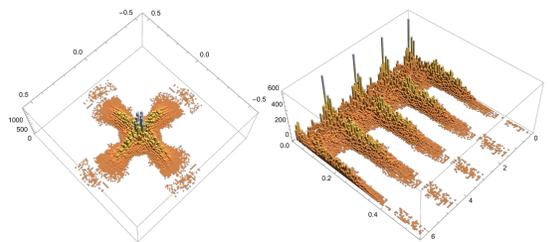
경쟁 기업 데이터 품질 관련 시각화 및 합성데이터 생성 분야

기업명	위치	최근 투자 유치	작년 매출	특징	차별점
Anomalo	미국	1170억원 (Series B, '24.11)		2025년 발간 Gartner 보고서에서 정형 데이터 품질 진단 기업으로 언급됨.	페블러스는 정평/비정형 데이터에 대해 데이터 품질 진단에서 개선 까지의 통합적 데이터 품질관리 제공
Shelf.io	미국	752억원 (Series B, '21.8)		2025년 발간 Gartner 보고서에서 정형 데이터 품질 진단 기업으로 언급됨.	페블러스는 정평/비정형 데이터에 대해 데이터 품질 진단에서 개선 까지의 통합적 데이터 품질관리 제공
SuperbAI	한국	누적 490억원 (Series C, '24.9)	?억원	Physical AI를 위한 비전 파운데이션 모델	데이터 품질, 가시화, 합성데이터, 모델 개발, 강화학습 최적화 등 스펙트럼 넓은 산업 AX 서비스 제공
Tonic.ai	미국	600억원 (Series B, '21.11)	149억원	개인정보 보호를 위한 정형 합성데이터 생성	다양한 모달리티의 고품질 합성데이터 (Hyper Synthetic Data) LLM을 위한 텍스트, 비전 모델을 위한 이미지, 정형 데이터를 위한 재현 데이터
Virtualitics	미국	400억원 (Series C, '23.8)	236억원	AI를 통한 데이터 분석 및 가시화 지원	데이터 가시화를 넘어 데이터 커뮤니케이션 도구 로. 설치형 및 SaaS 형 버전으로 세분화. AI 학습데이터 및 빅데이터 분석에 활용.

ROI 중심의 고객 사례

고객사 정보보안을 고려하여 대략의 근사치 및 내부에서 실험한 유사 사례로 설명함

	현대차  HYUNDAI	한화비전  Hanwha Vision	자이언트스텝  GIANTSTEP	해병대  대한민국 해병대
목적	차량 운행중 센서 데이터를 통한 위치 추정 AI 개발	객체 검출 및 상황인지를 위한 AI 감시 카메라 개발	가상 아이돌 제작을 위한 라이브 페이스 캡처 시스템 개발	해안 경계감시를 AI 개발
문제점	(1) 낮은 AI 성능, (2) 공간 제약에 따른 데이터 취득의 어려움	(1) 고객향 데이터 품질 입증 자료 부재, (2) 라벨링 부재	(1) AI 학습 시간 병목, (2) 중복이 심한 200만건의 합성데이터	해무, 파도 등 다양한 환경 및 북한군 객체 데이터 취득 어려움
솔루션	(1) 데이터 품질 진단, (2) 정밀 타기팅 합성데이터 생성, (3) AI 참조모델 개발	(1) 최적 신경망 선별을 통한 오토라벨링 제공, (2) 데이터 클리닉을 통한 품질 진단	(1) 데이터 클리닉을 통한 품질 진단, (2) 합성데이터 생성, (3) 인공지능 참조모델 개발	컴퓨터 그래픽스, 생성모델을 활용한 고품질 합성데이터 생성
ROI 추정	(1) 데이터 취득 과정을 15일에서 1시간 이하로 시간 단축 , (2) 공간섭외 및 인력투입에 필요한 비용 절감 , (3) 200% 이상 AI 성능 향상	(1) 고품질 자동 라벨링을 통한 시간 및 비용 절감 , (2) 데이터셋의 품질 상태 확인으로 차세대 데이터 거버넌스/파이프라인 필요성 인지	(1) 15% 데이터 경량화로 학습시간 기준 1주일에서 1일로 단축 , (2) S사 가상 아이돌 제작에 활용, (3) SIGGRAPH Asia 2025 학회 발표	(1) 실제 취득이 어려운 13.5만건 합성이미지 생성, (2) 비주얼 튜링 테스트 만점 수준 , (49점으로 통과. 실제와 합성을 구별하지 못할 때 50점)
기간	2022년. 제로원 프로젝트. 5개월.	2024년. PoC 프로젝트. 5개월.	기술개발 컨설팅 3년	2024년 과기부 프로젝트. 6개월.



현대자동차 프로젝트

프로젝트명	수행연도	적용기술	현업의 문제 및 요구사항	해결 방법 및 결과
주행 경로 추정 모델을 위한 합성데이터	2022년	시뮬레이터 합성데이터 신경망 모델 학습	대형 주차장 내에서 GPS가 끊기면서 생기는 주행 경로 오차 에 따라 정확한 주차 위치를 파악할 수 없는 문제	실제 주행 경로를 수집하는 것은 매우 어려운 일이었으므로, 합성데이터를 제작하여 모델을 학습하고 서비스 가능한 알고리즘 개발
타이어 마모 예측 모델	2022년	실험 설계 및 모델 학습	운전자의 주행 습관에 따라 달라지는 주행 센서 데이터로부터 타이어 마모 속도를 추정하여 예지보전에 활용	차량 주행 센서의 시계열 데이터로부터 타이어 마모 예측에 필요한 특징을 추출하고 타이어 마모 예측 알고리즘 개발
PBV 아동통학차량 합성데이터	2023년	CG & GenAI 이미지 합성데이터	목적 기반 차량의 내외부 안전 감시를 위한 카메라 비전 AI 학습을 위한 아동 이미지 합성데이터 생성 요청	차량 내부에 설치된 광각 카메라에 잡히는 아이들의 다양한 자세와 모습을 컴퓨터 그래픽과 생성형 AI를 동시에 사용하여 합성데이터 제작
울산공장 도장 공정 최적화를 위한 강화학습 모델	2025년	시뮬레이터 환경에서의 강화학습	도장 공정에서 칼라 전환이 최소화 되도록 다수의 레인에 차량을 인입하고 인출하는 최적화 문제 해결	시뮬레이터 상에서 수치해석 모델 대비 25% 성능을 개선하는 강화학습 알고리즘 개발
광명공장 조립 공정최적화를 위한 강화학습 모델	2025년	시뮬레이터 환경에서의 강화학습	조립 공정에서 차량 모델 전환이 최소화 되도록 다수의 레인에 차량을 인입하고 인출하는 최적화 문제 해결 요청	시뮬레이터에 복잡한 조립 공정 프로세스를 구현하고, 차량 모델 전환율을 최적화 하는 강화학습 알고리즘 개발

Traction

2025년 6월, 대구디지털혁신진흥원 데이터 품질 평가 및 개선 사업

대구 지역 30개 수요기업 대상으로
데이터 품질 진단 및 개선 수행

제조업, 헬스케어, AI개발, 로봇 등
다양한 도메인의 기업 데이터 분석

Pebblous

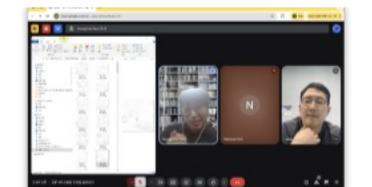
이데일리 2025년 09월 09일 화요일
세상을 올바르게 세상을 따뜻하게

페블러스, 대구·경북 지역 기업 대상 데이터품질 개선 무료 컨설팅

'데이터 품질진단 및 개선 용역' 주관 수행기관 선정
다음 달 24일까지 사업에 참여할 수요기업 모집

등록 2025-07-29 오후 2:50:36

수정 2025-07-29 오후 2:50:36



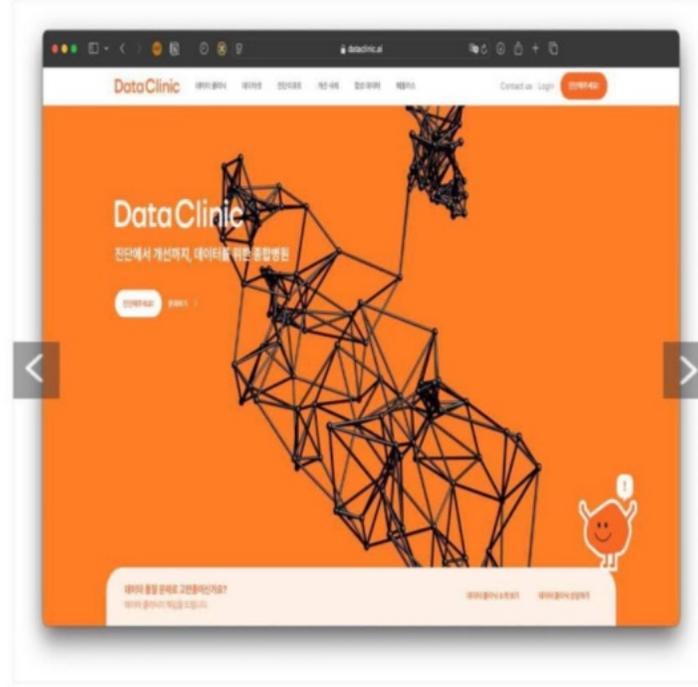
Traction

2025년 6월, 조달청 혁신제품 지정

조달청 혁신제품에 지정되어
공공기관 대상으로 정부 지원을 받아
서비스를 제공하는 방식으로 매출 확정
(2025년 11월 이후)

Pebblous

물품식별번호 : 25618833



클라우드소프트웨어, 페블
러스, DATA CLINIC, 4TB,
1 user, 분석 솔루션
55,000,000 원

※ 위 판매희망 가격은 혁신장터 운영규정(조달청 고시 제2020-36호)에 따른 견적가격 (VAT포함)이며, 혁신장터 이용약관(조달청 고시 제2020-37호) 제16조 2항에 따라, 「국가를 당사자로 하는 계약에 관한 법률 시행규칙」 제5조 1항 1호에서 규정하는 조달청장이 조사하여 통보한 가격이 아님을 알려드리오니 유의하시기 바랍니다.

기본정보

- 업체명 : 주식회사 페블러스
- 사업자번호 : 5848602422
- 모델명 : DATA CLINIC
- 세부품명번호 : 4323299901
- 제조/공급 : 제조
- 물품식별번호 : 25618833
- 단위 : 조
- 기업규모 : |벤처|소기업
- 중기간경쟁제품 : 아래사이트참조

☞ 중소벤처기업부에서 고시한 중소기업자간 경쟁제품은 <https://www.smpp.go.kr> 사이트로 접속하신 후 [정보조회]-[제품정보]-[중소기업자간 경쟁제품] 에서 조회하실 수 있습니다.



2026년 2월, 조달청 혁신제품 시범구매 기관 확정

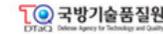
조달청 혁신제품에 지정되어
공공기관 대상으로 정부 지원을 받아
서비스를 제공하는 방식으로 매출 확정

Pebblous

Data Clinic 조달청 시범구매 3개 기관 확정!

국내 유일 AI 품질관리 SaaS 혁신제품
AI 학습데이터의 품질진단 및
개선통합형 데이터클리닉 솔루션

고품질 데이터가 필수적인 국방·공공 분야의 핵심 기관



국방기술품질원
국방 규격 및 품질 보증



육군 시험평가단
무기체계 시험평가



대구디지털혁신진흥원
지역 디지털 혁신 거점



2025년 7월,
국가대표 AI 사업 출사표

평가 의견
모델 취약성 분석 기반 타겟 데이터 합성을 통한 개선 계획이 우수함

KONAN
TECHNOLOGY

독자 AI 파운데이션 모델 프로젝트

**제조 AI 파운데이션
모델 및 학습 효율화
기술개발**

2025. 7. 30.

공동연구: 고려대학교, 연세대학교, POSTECH, KAIST 한국과학기술원

공동개발: 사이냅소프트, ALCHERA, Pebblous

지원기관: 과학기술정보통신부, nipa 정보통신산업진흥원, IITP 정보통신기획평가원, NIA 한국지능정보사회진흥원

Traction

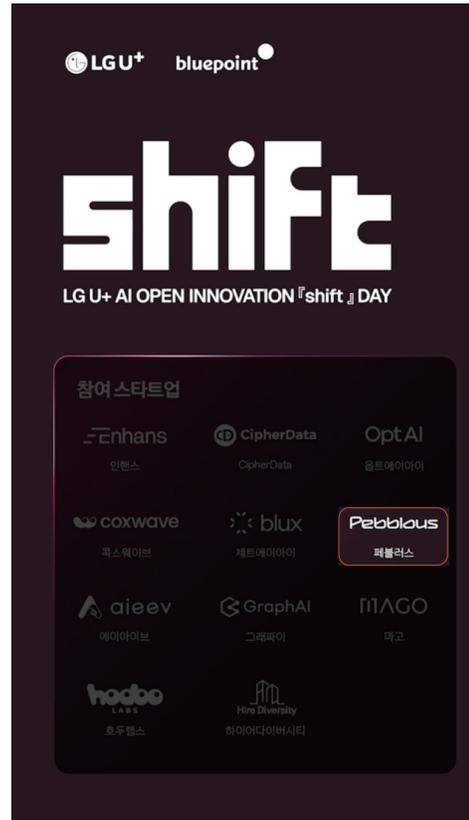
2025년 10월, LGU+ Shift 선정

LG유플러스와 블루포인트가 함께
진행하는 PoC 및 펀드 투자 연계
프로그램 『shift』 2기 선발

Pebblous

페블러스, 188:11의 경쟁률을 뚫고 LG U+ 협력 및 투자 프로그램에 선발

글로벌 AI 기업으로 성장할 우수한 스타트업 팀에 선발
2026년 LG U+ 통신망 데이터의 품질 향상 과제 수행 예정



2026년 2월, LGU+ 4개 팀 PoC 진행

LG유플러스와 블루포인트가 함께
진행하는 PoC 및 펀드 투자 연계
프로그램 『shift』 2기 선발

NW기준정보팀

RN장비 End-To-End 연결정보 **데이터의 품질 진단** 및
이상탐지 알고리즘 개발

목표	기대효과
차수별 RN장비 연결정보 E-T-E 전개 관련 데이터 불일치 검증 및 누락 검증. 그래프 데이터	실시간 데이터 불일치 및 누락 오 류 검증을 통한 데이터 이상관리

유선AX솔루션팀

AI-WG FTTH망 RN 품질 저하 딥러닝 이상탐지 모델
성능 개선을 위한 **데이터 품질 진단 및 합성 데이터 생성**

목표	기대효과
클래스 불균형이 심한 다변수 시계 열 데이터 의 품질 개선. 이를 통한 abnormal recall 개선 목표	· 데이터 선별 및 증강을 통한 모 델 성능 고도화 · 정상/비정상 데이터 시계열 패 턴 분석

무선AX솔루션팀

무선 RRH 불량 장비 AI 조치추천 모델의 성능 개선을 위
한 **데이터 품질 평가 및 개선**

목표	기대효과
데이터 불균형 문제 해결을 통한 AI조치추천 솔루션 성능 개선 및 상용화	· 조치추천 가능 카테고리수 개선 · 조치추천 정확도 개선

Foundation리서치팀

LLM 모델의 성능향상을 위한 학습 **데이터의 품질 진단
및 개선**

목표	기대효과
LLM 학습데이터를 솔루션(Data Clinic)으로 품질 평가가 가능한지 검증 및 이를 통해 솔루션의 도입 가능성 및 유용성 판단	자체 증강한 LLM 학습 데이터를 외부 솔루션을 통해 품질 진단하여, 객관적인 고품질 데이터 확보 및 LLM 모델 성능 제고

2027년 주요 사업계획 육군시험 평가단

국방 LLM/SLM 성능평가
체계 구축 용역
국방AI 시스템 구축 사전조사 사업

매출 예상: 23.1억 원

견적서

고객사	육군시험평가단	공급자	주식회사 페블러스	584-86-02422
프로젝트	국방 LLM,SLM 성능평가 체계 구축 용역	대표자	이주영	
견적일	7/28/25	주소	대전광역시 유성구 대학로 99, 507호	
수행기간	18개월	전화/이메일	044-589-3824	contact@peblous.ai

제안가격 2,310,000,000 원 (VAT포함)

코드	구분	내용	수량	단가	금액	비고
1	벤치마크 데이터셋 생성	데이터 수집 및 정제	1	100,000,000	100,000,000	
		데이터 어노테이션 및 검증	1	100,000,000	100,000,000	
		합성 데이터 생성 (API 비용 등)	1	100,000,000	100,000,000	
		데이터 경량화	1	100,000,000	100,000,000	
		데이터 보안 및 비식별화 처리	1	100,000,000	100,000,000	
소계 \					500,000,000	

코드	구분	내용	수량	단가	금액	비고
2	성능평가 체계 구축 (HW)	LLM 추론 서버 (GPU 2기 포함)	1	300,000,000	300,000,000	
		네트워크 장비 및 렉	1	30,000,000	30,000,000	
		초기 설치 및 안정화	1	20,000,000	20,000,000	
					0	
			0			
			0			
소계 \					350,000,000	

코드	구분	내용	수량	단가	금액	비고
3	품질평가 체계 구축 (SW)	국방 데이터 클리닉 플랫폼 개발	1	550,000,000	550,000,000	
		내부 개발 인건비	1	350,000,000	350,000,000	
		외부 용역 개발비	1	100,000,000	100,000,000	
		상용 SW 라이선스	1	50,000,000	50,000,000	
소계 \					1,050,000,000	

코드	구분	내용	수량	단가	금액	비고
4	사업관리 및 운영	프로젝트 관리 인력	1	150,000,000	150,000,000	
		외부 전문가 자문	1	10,000,000	10,000,000	
		문용 인력 교육 및 훈련	1	20,000,000	20,000,000	
		예비비	1	20,000,000	20,000,000	
소계 \					200,000,000	

총계(TOTAL)	(1+2+3+4)	\	2,100,000,000
절사	만원 단위 미만	\	2,100,000,000
부가가치세	10%	\	210,000,000
합계(GRAND TOTAL)		\	2,310,000,000



솔루션 확장: 데이터 그린하우스 SaaS — 데이터 품질관리에서 공급까지



AI 적합성을 만족하는 그린 데이터(Green Data) 제공

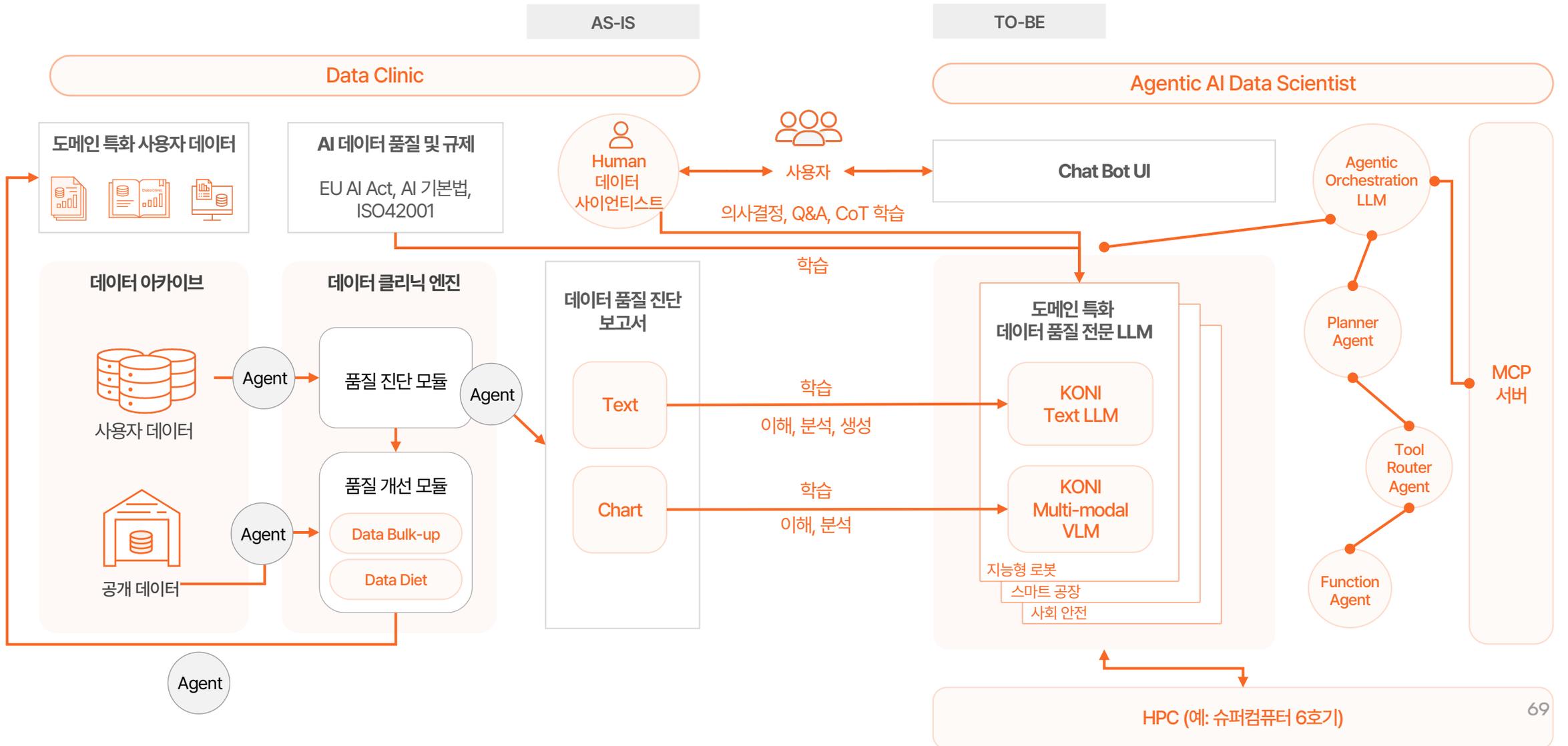
- 01 데이터 품질 평가 및 최적화를 통해
AI 적합성 기준을 만족하는 **고효율 친환경 데이터 제공**
- 02 **의미 기반 데이터 연산**을 통한 그린 데이터 생성
 - 대용량 데이터에 대한 전수조사 기반 자동 품질평가
 - 정밀 타기팅 합성 데이터 생성
 - 중복 및 편향을 제거하는 데이터 경량화
 - 개인정보 보호를 통한 안전한 데이터 유통
 - 의미 기반 탐색/추론/결합을 통한 데이터 생성



데이터 그린하우스(Data Greenhouse) SaaS 솔루션

- 01 데이터의 AI 적합성을 지속적으로 평가하고 개선하는
차세대 거버넌스를 위한 **데이터 매니지먼트 솔루션**
- 02 **의미 기반 데이터 관리 체계**를 통한 효과적인 AI 산업
규제 대응
 - 대용량 데이터에 대한 전수조사 기반 자동 품질평가
 - 데이터 수명주기 관리를 통한 유연한 작업 대응
 - 지속 가능한 AI를 위한 데이터의 안전성 및 윤리성 검증

솔루션 확장: 자율형 데이터 클리닉 — Agentic Data Clinic



글로벌 빅테크 육성사업

페블러스 주관으로 KISTI와 컨소시엄을
이루어 대형과제 수주
AI 에이전트 개발 진행중

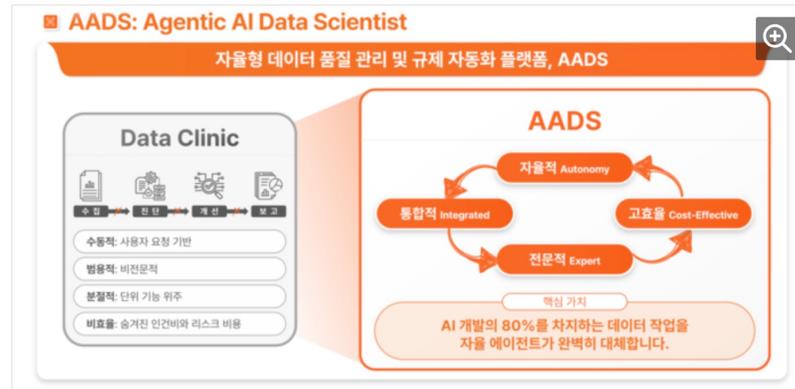
2025년 8월



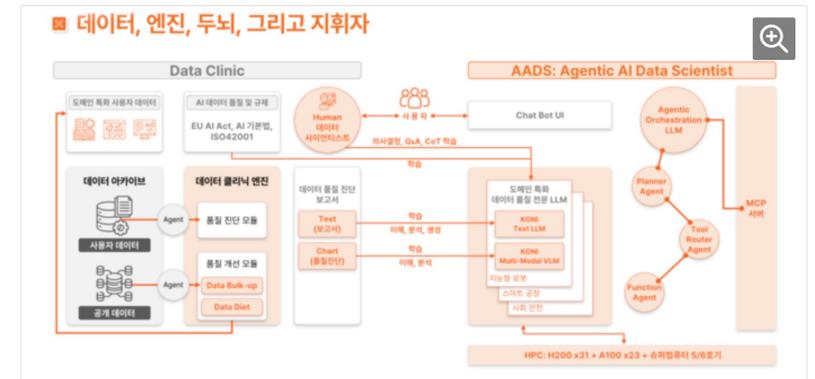
페블러스, 과기부 '글로벌 빅테크 육성사업' 선정... KISTI와 AI 에이전트 자율형 데이터 품질 관리 기술 공동개발

발행 2025-08-21 09:18

KISTI와 협력해 AADS 플랫폼 공동 개발...
총사업비 61억원 규모
데이터 품질 · 규제 대응 · 해외 시장 확장 목표



페블러스 데이터 클리닉에서 진화한 AADS



AADS는 데이터 관리 모든 단계 엔드투엔드 자동화된 시스템이다.

솔루션 확장: 피지컬AI를 위한 초고품질 합성데이터 — Hyper-Synthetic Data for Physical AI

Gartner

Emerging Tech: Tech Innovators in Hyper-Synthetic Data for High-Fidelity Imaging

23 June 2025 - ID G00825904 - 26 min read

By: Nick Ingelbrecht, Vibha Chitkara, Tuong Nguyen, Kiumarse Zamanian, Ben Lee
Initiatives: [Emerging Technologies and Trends Impact on Products and Services](#); [Increase Product Traction](#)

The surge in demand for computer vision training data is driving the adoption of hyper-synthetic data (HSD) solutions. Product leaders should capitalize on this emerging AI race opportunity by identifying and integrating high-fidelity spectral imaging innovations into their product roadmaps.

More on This Topic

This is part of an in-depth collection of research. See the collection:

- [Emerging Tech Roundup: Navigating Emerging Disruptions and Challenges Ahead](#)

PHYSICAL AI

Intelligent environments that sense, understand, and act.



Sense
Capture real-time data



Understand
Analyze and simulate



Act
Trigger intelligent actions

Physical AI의 본질적 데이터 병목



01 Data Famine

희소 데이터(Edge Case)의 절대적 부족

사고, 재난 등 AI 안전에 필수적인 '위험 상황' 데이터는 현실 세계에서 수집 자체가 불가능하거나 비용이 막대함: 데이터 기근(Data Famine)

02 Physical Hallucination

생성형 AI의 물리적 환각

기존 GenAI는 시각적으로만 그럴듯할 뿐, 중력·마찰 등 물리 법칙을 위배하여 정밀 제어 모델 학습에 치명적임

03 Data Complexity

멀티모달·3D·시계열 데이터 복잡성

이미지, 센서, 로그 등 이기종 데이터 간의 시공간 동기화(Sync) 및 인과관계 파악이 난해함

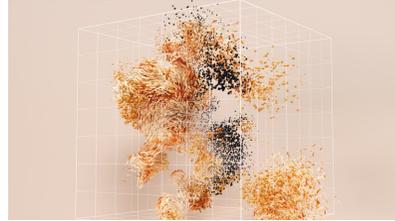
피지컬 AI 데이터 병목의 근본적 해결책: Data Greenhouse + PebbloSim

우리는 데이터를 수집하지 않고 '경작(Farming)'합니다.

① 물리 법칙을 복제한 **PebbloSim**과

② 자율 데이터 운영 체계 **Data Greenhouse**를 통해,

피지컬 AI를 위한 AI-Ready 데이터를 지속적으로 생산하고 품질을 증명합니다.



O1 Agentic Data Synthesis

데이터 기근을 해결하는 '자율 데이터 벌크업'

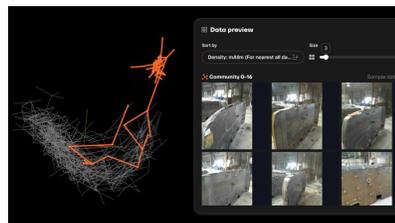
AADS 에이전트가 데이터 공백을 스스로 진단하고, 필요한 엣지 케이스를 정밀 타격(Targeting)하여 합성데이터 생성: On-Premise/Sovereign 환경 지원



O2 Physically-Grounded

물리적 환각 없는 '뉴로-심볼릭' 합성 데이터

시뮬레이션의 물리적 정합성 위에 생성 AI의 시각적 사실감을 입혀, 물리적 환각을 구조적으로 원천 차단



O3 Explainable Verification

인과관계를 규명하는 '설명 가능한 검증'

데이터의 분포(Vector)와 비즈니스 의미(Ontology)를 실시간으로 상호 연결하여, 데이터 생성의 근거를 투명하게 시각화하고 설명함

Physical AI의 최전선: 산업별 난제를 해결하는 자율 데이터 운영 체계 Data OS

산업 도메인	주요 요구사항	구체적 페인 포인트	페블러스 솔루션
 조선	인력 절벽 공기 단축 압박	AI 도입이 생존 문제로 직결 로봇 용접 AI OCR(도면 분석) 자율 운항 수십만 개의 부품·도면이 디지털화 되지 않았음	도면과 센서를 통합 이해하는 멀티모달 융합
 국방	데이터 주권 폐쇄망 환경	데이터의 외부 반출이 불가능한 폐쇄망 환경 소버린 AI 온프레미스 설치 한 번의 오류가 치명적인 고위험 환경	데이터 주권(Sovereignty)을 보장하는 온프레미스(On-Prem)
 모빌리티·제조	안전 필수 물리적 정합성	현실에서 1만 번 사고를 낼 수는 없는 '데이터 기근' 희소상황 데이터 안전규제 준수 기존 생성 AI는 물리 법칙을 무시하는 '환각' 발생	물리적 환각(Hallucination) 없는 고정밀 합성 데이터

Physical AI에서 페블러스가 갖는 결정적 우위



01

특허 기반 기술 해자

- 미국 특허 Data Imaging (US 12,481,720 B2)
- 단순 관측을 넘어 구조적 개선(Diet/Bulk-up)까지 보호받는 기술적 해자



02

단절 없는 End-to-End 데이터 OS

- 진단(Clinic) → 치료(Bulk-up) → 검증(Sim)의 유기적 통합
- 파편화된 도구가 아닌, AADS 기반의 완전 자율 운영 체계



03

공신력과 진입장벽

- KOLAS 기반의 국제 공인 시험성적서 발급 권한 (품질 보증)
- SaaS가 진입 불가능한 '국방/제조 폐쇄망(On-Prem)' 구축 역량 (진입 장벽)



04

글로벌 표준 내재화

- ISO/IEC 5259 (데이터 품질)
- ISO/IEC 42001 (AI 거버넌스)
- 운영 로그 자체가 ISO 42001/5259 감사 증거가 되는 Embedded Governance

데이터 그린하우스 OS: 데이터가 스스로 성장하는 지능형 운영 체제

Zero
Physical
Hallucination

ISO
Certification
Ready



PebbloSim

스스로 생성하고

- 핵심 실행 엔진. 디지털 트윈 기반
- 물리 법칙 기반의 합성 데이터 생성
- 부족한 Edge Case를 정밀 타격하여 보강



Data Clinic

데이터를 진단하고

- 데이터의 결핍과 과밀을 정밀 진단
- '무엇을 고쳐야 하는가' 자동 처방
- 개선을 위한 진단리포트 발행

Data
Communication



PebbloScope

투명하게 통제함

- Human-in-the-Loop 게이트웨이
- 뉴로-심볼릭 양방향 시각화
- 시뮬레이션 과정 실시간 모니터링

Data Diet

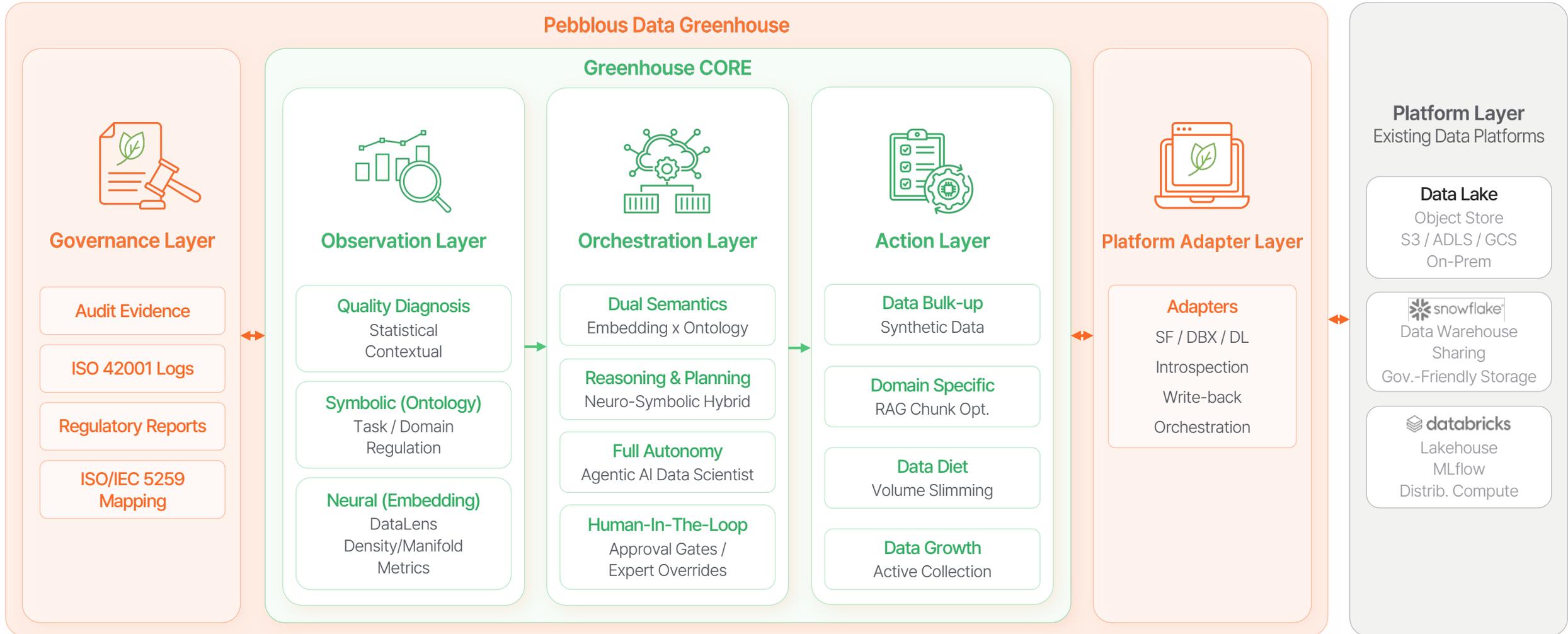
Data Bulk-up

데이터 그린하우스
OS

데이터 그린하우스 OS: AADS 기반의 'End-to-End' 자율 운영 파이프라인



솔루션 확장: 데이터 그린하우스 시 데이터 플랫폼의 '책임 레이어'



솔루션 확장: 데이터 그린하우스 적용 시나리오



모빌리티

현대자동차 스마트팩토리 용접 불량 검출

문제 상황

연간 희귀 불량 패턴 발생률 0.001% 미만
→ 데이터 부족, 결함 검출 한계, 비용 손실

솔루션 적용

- 현장 데이터 수집
- 합성 불량 데이터 생성
- AI 모델 재학습 & 배포

기대 효과

- 불량 검출 정확도 ▲ 94% 향상
- 불량률 ▼ 78.75% 감소
- 데이터 수집 시간 ▼ 99% 단축
- 수동 검사 시간 ▼ 83% 단축



조선

한국 조선소 건조 일정 예측 & 최적화

수천 개의 공정이 복잡하게 얽힌 초대형 과제
→ 데이터 복잡성, 소통 장벽, 비용 손실

- 멀티모달 데이터 수집
- 데이터 통합 & 디지털 트윈 구축
- PebbloScope 3D 시각화 & 현장 소통 강화

건조 일정 예측 정확도 ▲ 76% 향상

- 현장 관리자 데이터 활용률 ▲ 300% 향상
- 병목 발견 시간 ▼ 95% 단축
- 일정 지연 건수 ▼ 67% 감소



국방

육군/해병대 전술 훈련 합성데이터 생성

극도로 제한된 환경에서 이뤄지는 군사 훈련
→ 데이터 희소성, 보안 제약, 비용 손실

- 기존 데이터 수집 & 품질 진단
- Hyper-Synthetic 전투 시나리오 생성
- AI 모델 학습 & 시뮬레이션 통합

AI 모델 성능 ▲ 50% 향상

- 데이터 생성 기간 ▼ 99.9% 단축
- 훈련 비용 ▼ 95% 감소
- 데이터 라벨링 비용 ▼ 100% 감소



투자 유치 및 투자금 사용 계획



Plan #1
SaaS 매출 10억 원 달성 기반 마련

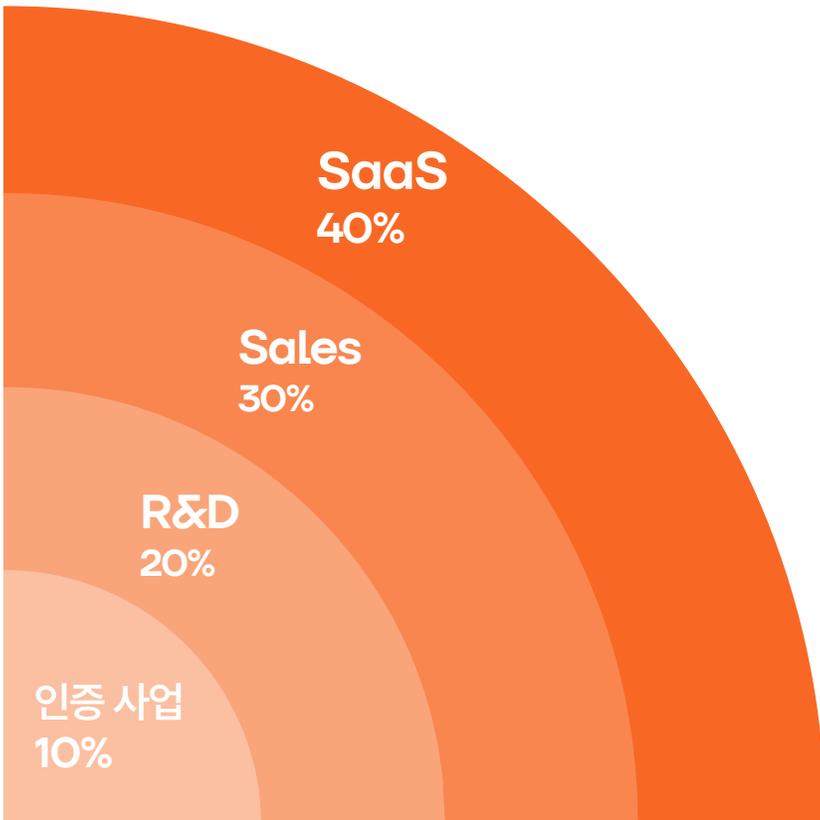
Plan #2
미국 상표권 및 특허 기반 시장 진출 가속화
국내 및 글로벌

Plan #3
자율형 AI 데이터 과학자 (AADS) 상용화
데이터 그린하우스, 페블로심

Plan #4
공공 조달 시장 진입 및 신규 인증 수익 창출
KOLAS 인증



2026년 신규 투자금 사용 계획



01

SaaS

2026년 SaaS 매출 10억 원 달성 기반 마련

- GPU 서버 및 클라우드 인프라 확충
- SaaS 운영 및 유지보수 전문 인력 채용
- 주요 버티컬 특화 솔루션: 제조, 국방, 조선 등

02

Sales

미국 상표권 및 특허 기반 시장 진출 가속화

- 글로벌 세일즈 파트 신설 및 미국/유럽 영업 활동 강화
- Biz Dev 전문가 영입을 통한 대형 고객 파이프라인 확장

03

R&D

자율형 AI 데이터 과학자 (AADS) 상용화

- 핵심 기술 고도화: 데이터 품질진단, 합성데이터, 데이터 그린하우스, 산업AI
- 연구개발 인력 신규 채용
- 특허 포트폴리오 강화 (기술적 해자 구축)

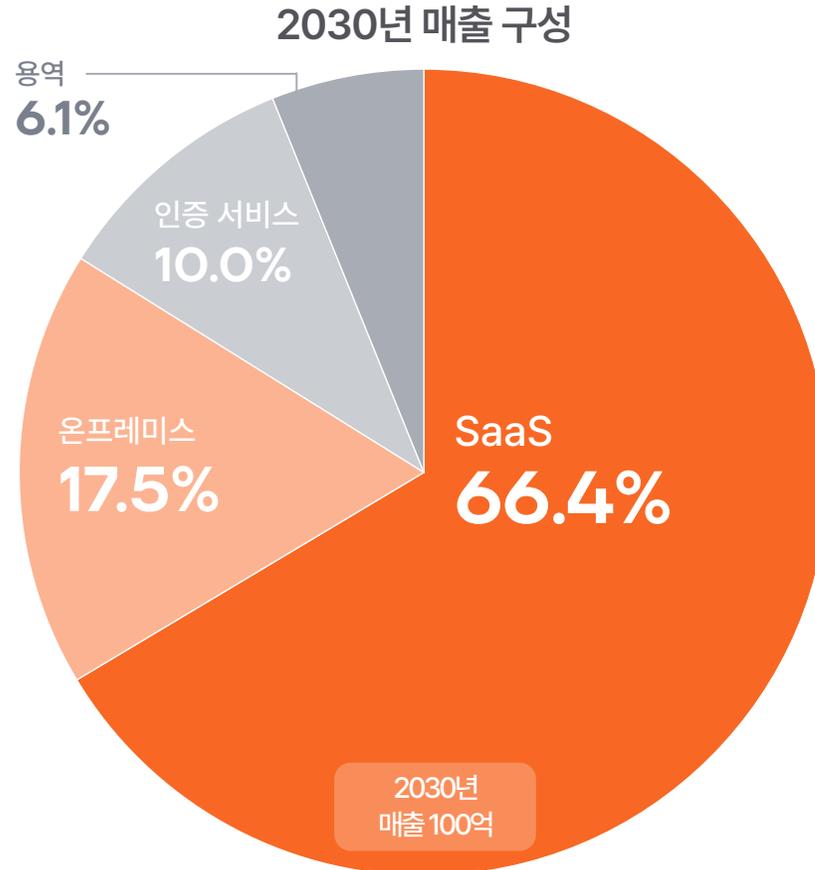
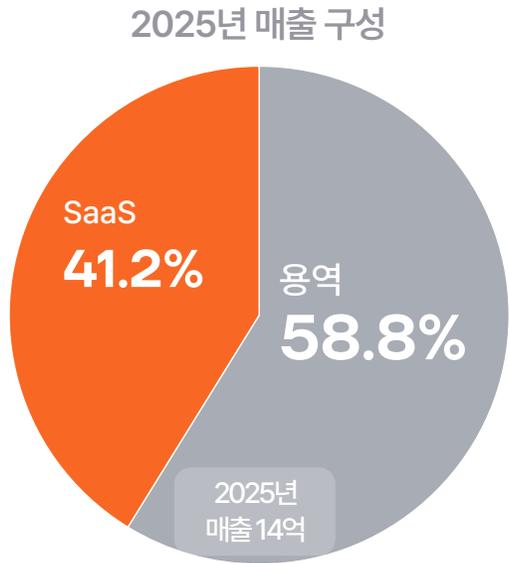
04

인증 사업

공공 조달 시장 진입 및 신규 인증 수익 창출

- 인증사업(예, KOLAS) 전담 인력 배치
- 공인 시험 성적서 발급 시스템 구축

2030년 재무 로드맵



Plan #1 25'
구독형 비즈니스의 시장성 검증

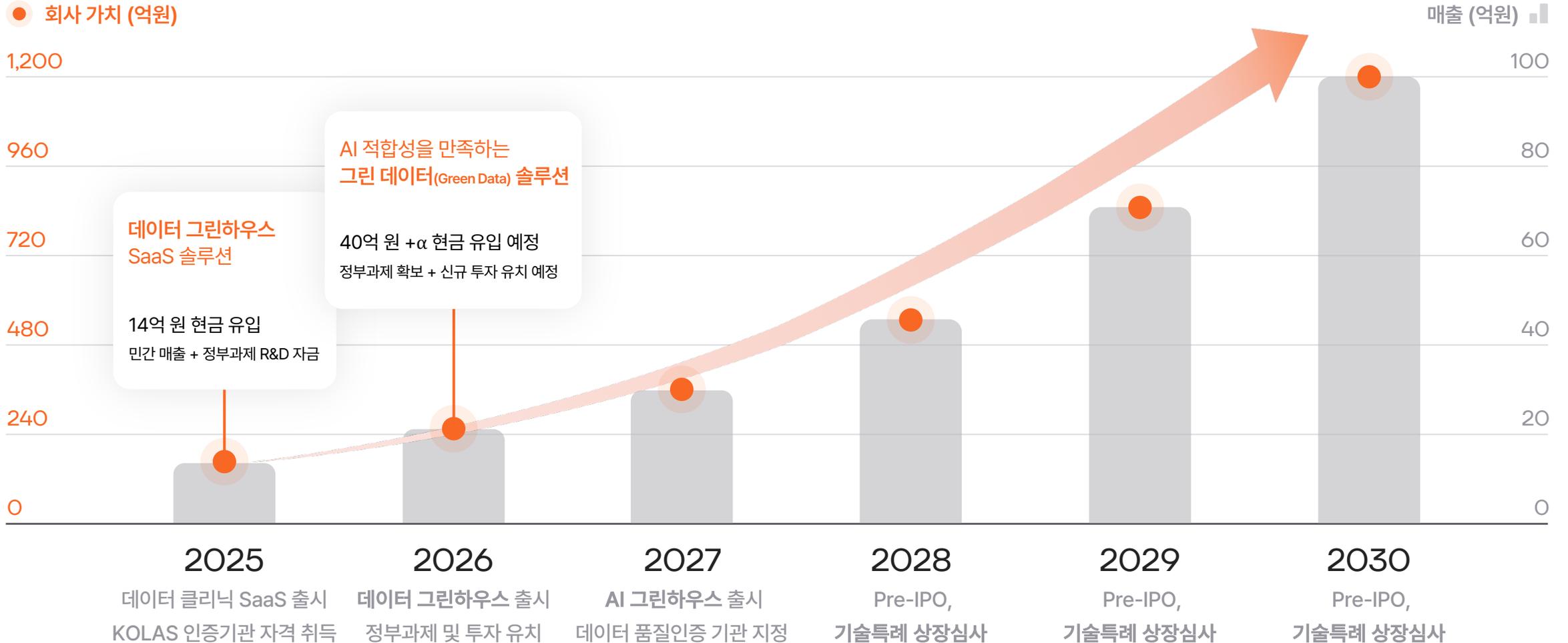
Plan #2 26-27'
온프레미스와 KOLAS 기반 인증 서비스
수익 구조를 플랫폼 중심으로 재편

Plan #3 28-29'
예측 가능한 반복 매출 창출
기술 중심의 고수익 구조 확립

Plan #4 30'
안정적이고 수익성 높은 최종 포트폴리오 완성
최소의 운영비용과 고마진 추구



SaaS based Revenue



데이터 및 AI 분야에서 합산 50년 이상의 연구개발 능력 인재들로 구성된 Perfect 팀



CEO | Co-Founder
이주행

Profile

POSTECH 컴퓨터공학 박사(Ph.D)
ETRI 책임연구원(CG, VR, Robotics, AI) 경력 23년
UST 교수
대전 비엔날레 초청 작가



COO | Co-Founder
이정원

Profile

KAIST 바이오및뇌공학과 박사(Ph.D)
서울대학교 전기공학부/의공학과 석사
ETRI 책임연구원(AI) 경력 20년
의료 AI 개발(개발 제품 서울대병원 설치 및 수출)

BD & Marketing 2

(미국) 구글 B2B 마케터 출신
(한국) UC 버클리 경제학과 졸업

AI/Data Engineer 4

POSTECH 컴퓨터공학 박사
서울대 공학박사 수료
KAIST 전산학과 졸업

CG Engineer 1

SW Engineer 2

UI/UX Designer 1

Back Office 3

Fabulous Data With

Pebblous

Better Data Makes Better AI



[Pebblous.ai](https://pebbulous.ai)