Pebblous Makes Data Tangible

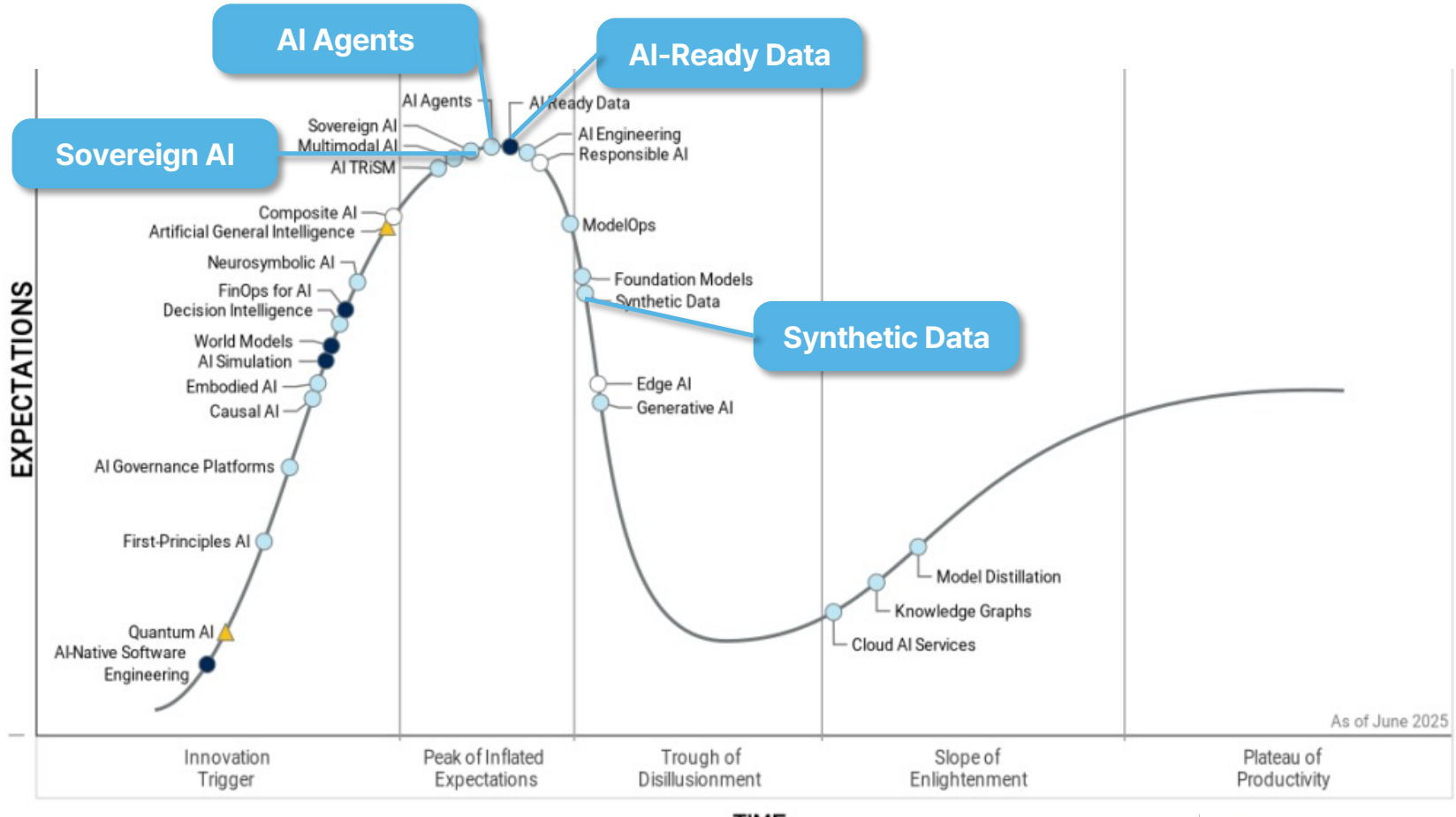
Pebblous.ai

Pebblous

Better Data Makes Better AI

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

Hype Cycle for AI, 2025



소버린 AI의 시대, '데이터-평가' 전면 재구성이 요구된다

최예진 교수 (스탠포드 대학교)

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 방식으로는 해결할 수 없는, 근본적인 문제라고 생각합니다. 따라서 근본부터 바꿔야 합니다. 학습 데이터, 학습 목표, 평가 방식까지 다 바꿔야 합니다. 그래야만 다양한 언어, 다양한 맥락에 걸쳐 유능한 AI를 만들 수 있습니다."

Pebblous

AI 데이터셋의 품질로 인한 문제들

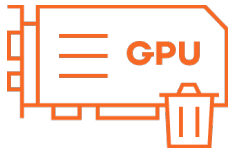
모호한 데이터 가격 체계, 모델 성능 저하, GPU 자원 낭비, 높아가는 AI 규제 대응 미흡

01

학습데이터

품질 불량/고갈

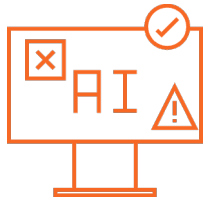
02



데이터센터

운용 비용/효율

03



AI 관련

법/규제 강화

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

Pebblous

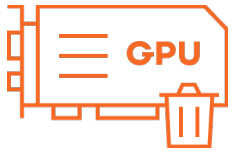
통합적인 데이터 품질 관리 솔루션

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

01

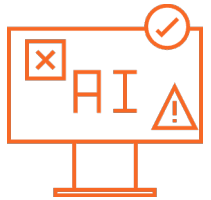
AI 학습 데이터
품질 진단/관리
> 모델 성능 개선

02



데이터 경량화
합성데이터
> 학습 효율 개선

03



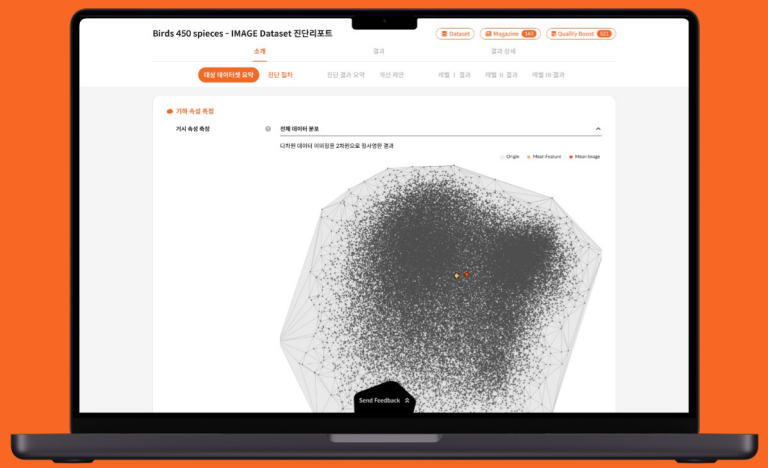
AI-적합 데이터
기준 모니터링
> 거버넌스 구축

Pebblous

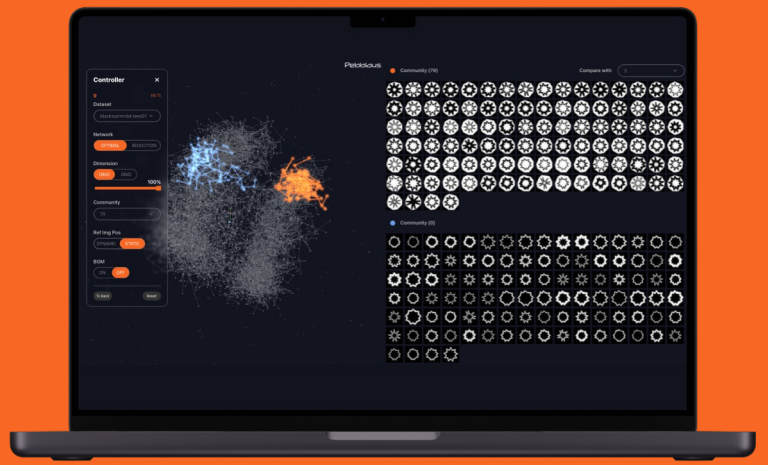
올인원 데이터 관리 솔루션

(1) 품질평가에서 개선까지,
(2) 인터랙티브 가시화/커뮤니케이션,
(3) 멀티모달 데이터셋,
(4) SaaS, 온프렘, API

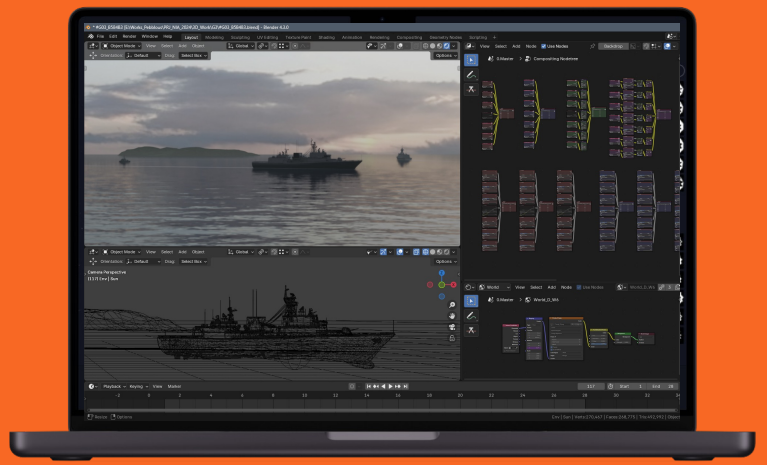
DataClinic



PebbloScope



Synthetic Data



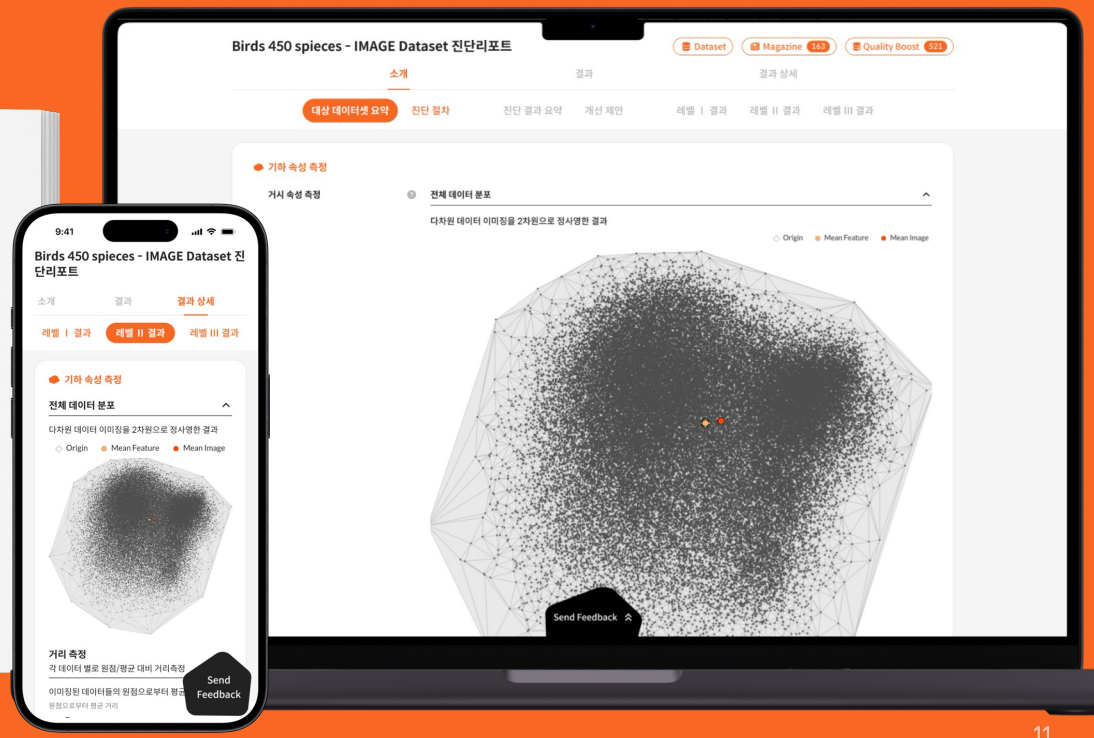
Pebblous

DataClinic

페블러스 데이터 클리닉은 **"데이터 종합병원"**입니다.
AI 학습 데이터를 위한 품질 평가에서 합성데이터 생성까지의 모든 솔루션을 제공합니다.

웹 버전

PDF 버전

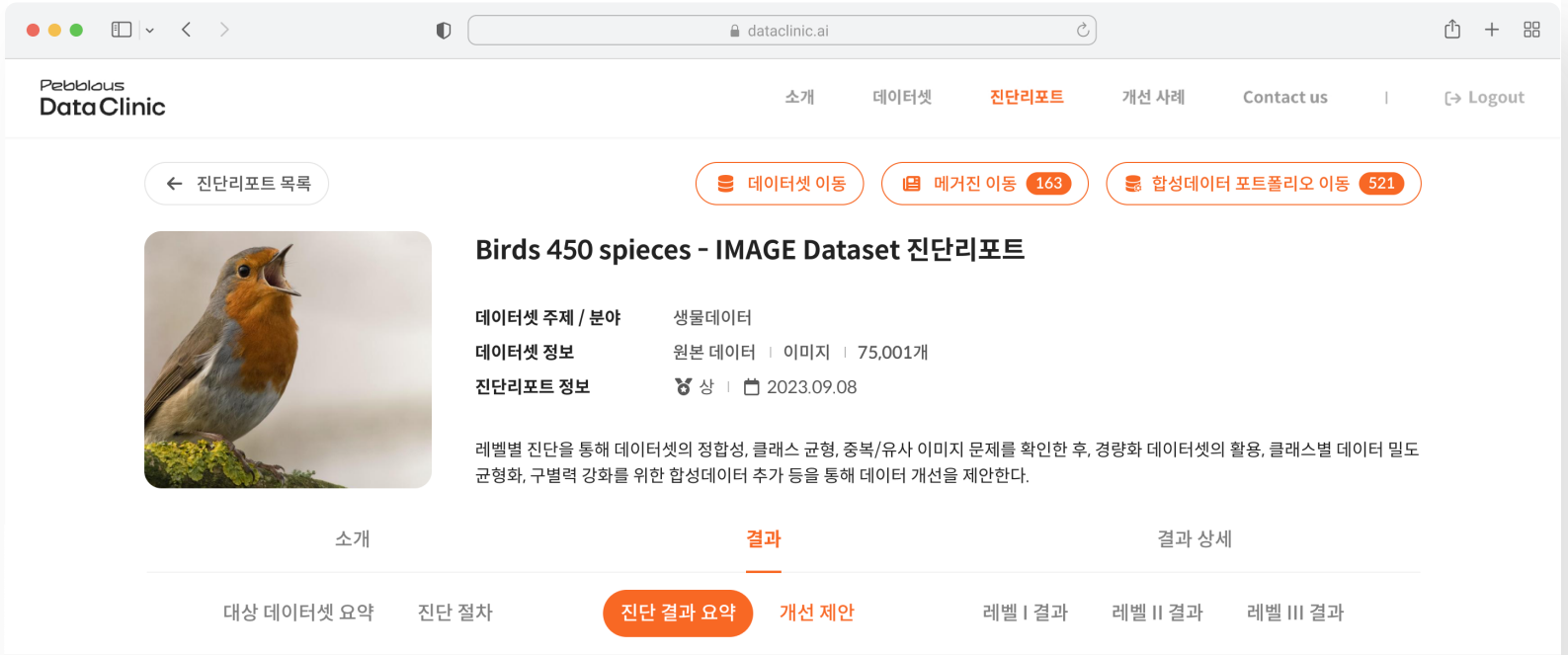


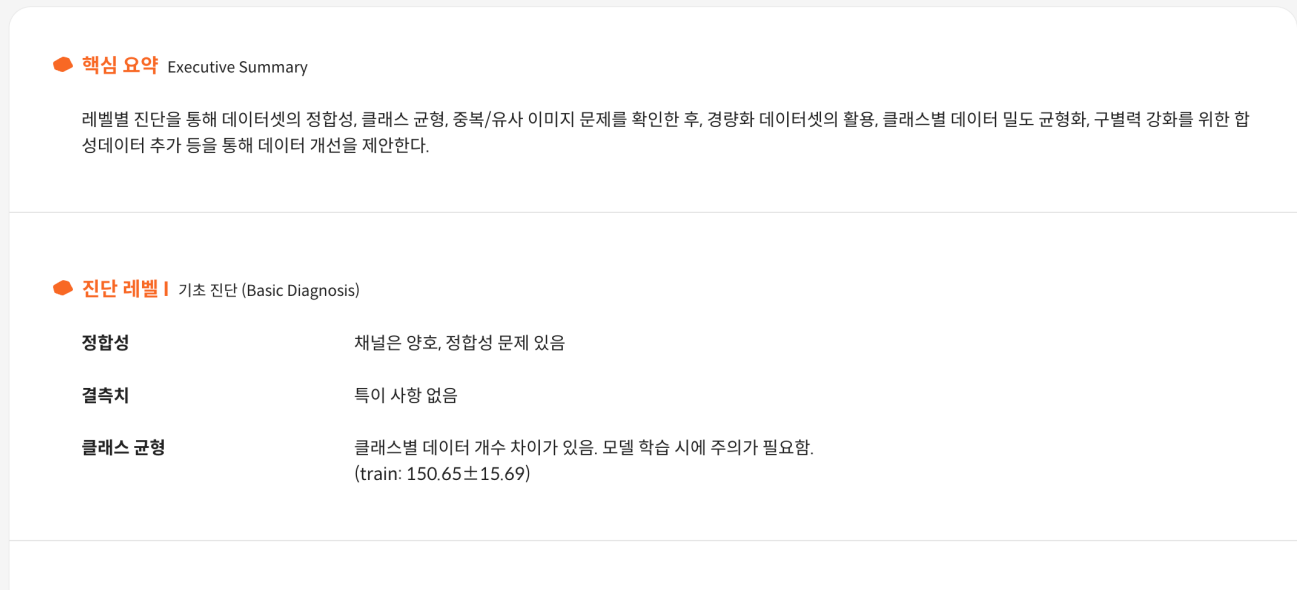
모바일 버전

Product 01

데이터 품질 진단 보고서

Web version





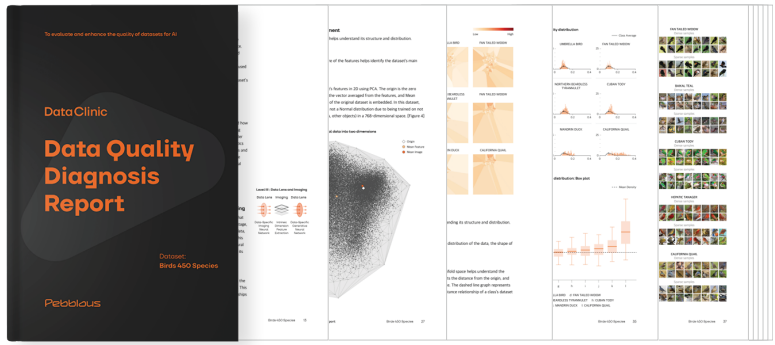
Product 01

데이터 품질 분석 보고서

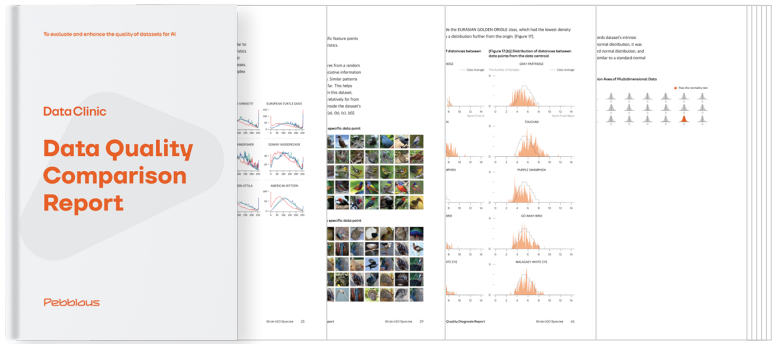
PDF / EPUB version

Pebblous

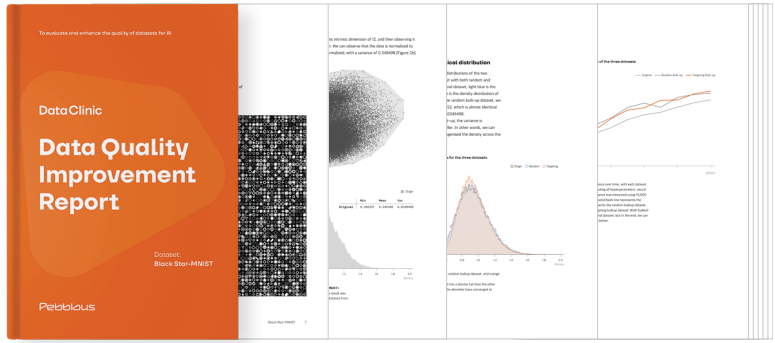
01
데이터 품질 진단 보고서



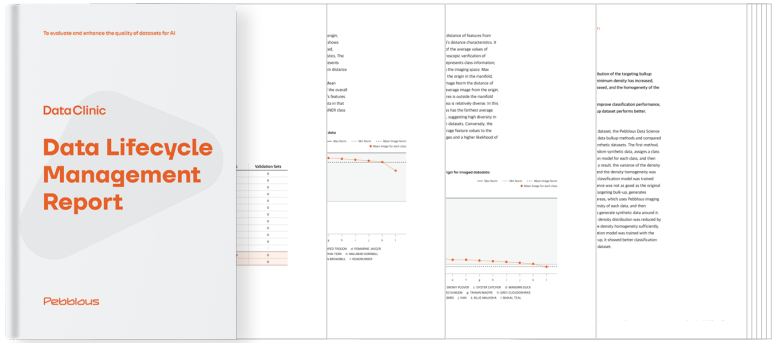
03 출시예정
데이터 품질 비교 보고서



02
데이터 품질 개선 보고서



04 출시예정
데이터 수명주기 관리 보고서



진단보고서 샘플 다운로드
https://bit.ly/진단보고서

Product 01

데이터 품질 진단 보고서

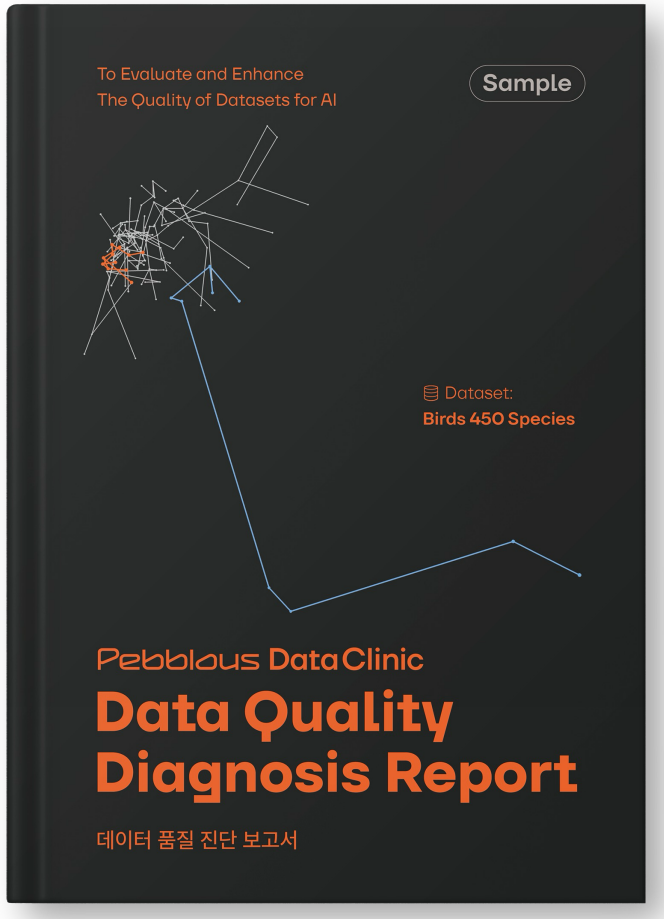
국문 샘플

AI 학습데이터셋에 대한 종합적인 품질 진단 결과와 개선방안이 담긴 보고서입니다.

Pebblous

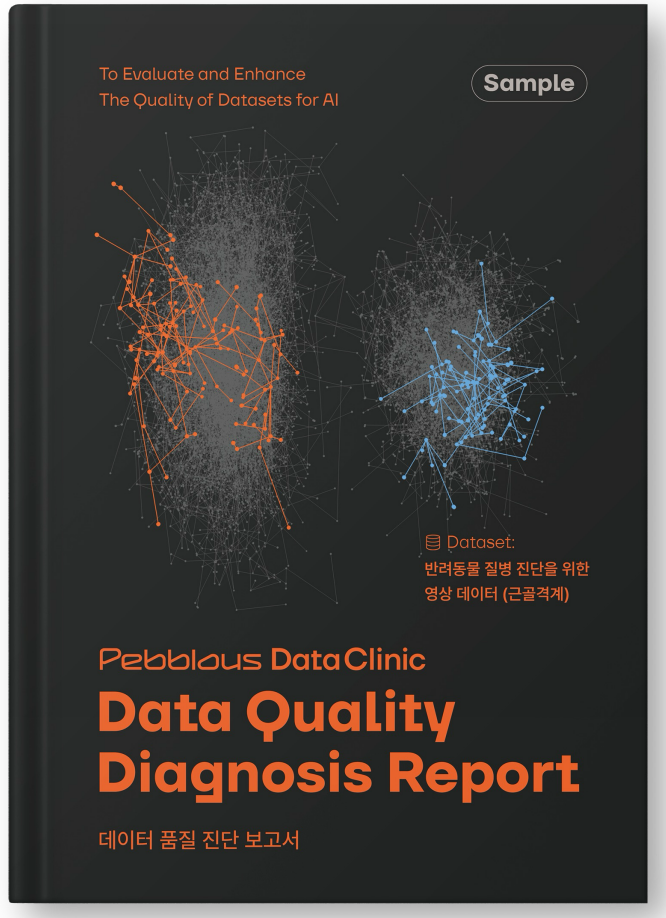
Sample 01
Birds-450 Dataset

450종에 달하는 조류 이미지 공개 데이터셋 (Kaggle)



Sample 02
Pet-bone Dataset

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



Product 01

데이터 품질 진단 상세 & 결과

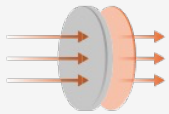
Pebblous

레벨 I 진단 : 기초 진단

데이터 정합성 측정
결측치 측정
클래스 균형 측정
통계 측정

레벨 II 진단 : 일반 진단

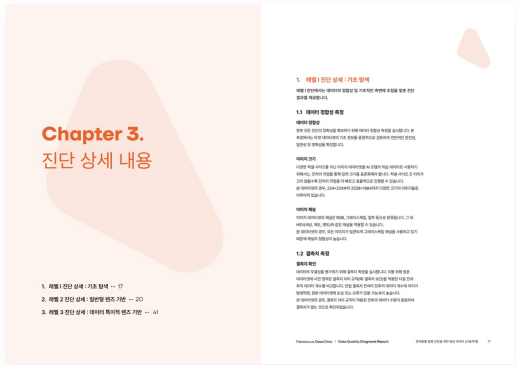
데이터 렌즈
사전 학습된 이미징 신경망

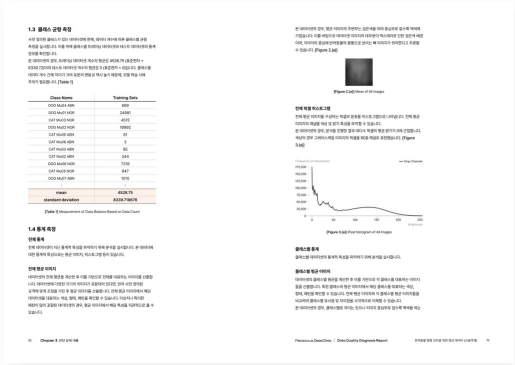


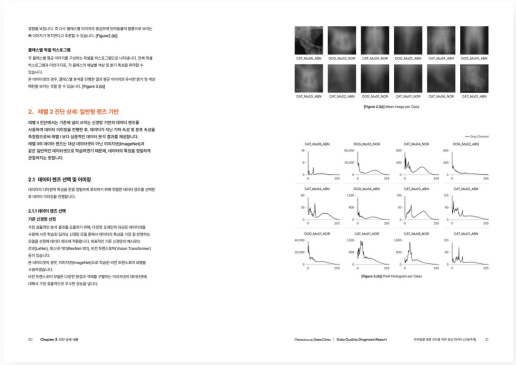
이미징
관찰 차원 특징 추출

진단 상세 내용

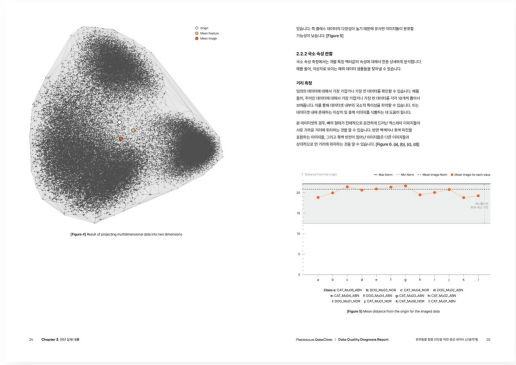
01 레벨 I 진단 : 기초 탐색

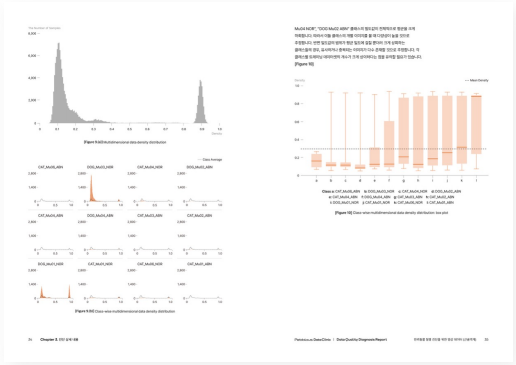






02 레벨 II 진단 : 일반형 렌즈 기반





Traction

가트너 보고서

데이터 품질 관리 대표 기업

Develop Unstructured Data Management Capabilities to Support GenAI-Ready Data, Gartner, 2025.

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.

Gartner

Table 1: Parameters to Prioritize Partnerships for Unstructured DMS

Unstructured DMS market segment	Current end-user demand[3]	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	BigID, DryvIQ
Metadata management	Medium	Low	Medium	Instill AI, Labelbox

Source: Gartner

Product 01

Pebblous

2025년 6월, 데이터 품질 평가 및 개선 사업 수주

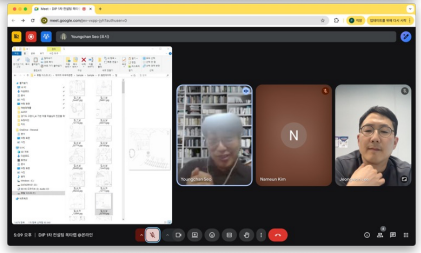
페블러스, 대구·경북 지역 기업 대상 데이터품질 개선 무료 컨설팅

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

'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



Pebblous

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

DIP 데이터 품질 진단 및 개선 사업

Dashboard

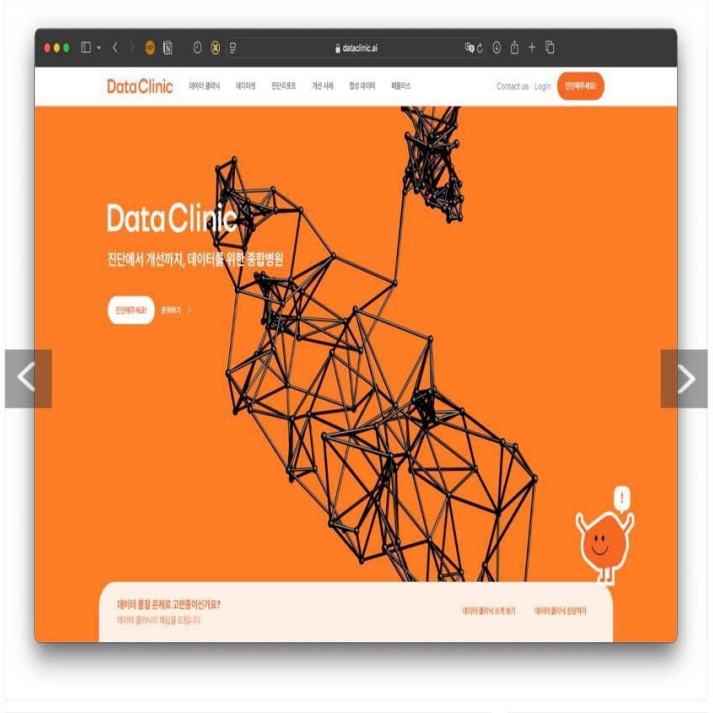
전체보기 | 컨설팅 캘린더 | 진단 진행상황 | 개선 진행상황 | DIP 공유 | New

No	Name	데이터 설명	데이터 형태	서비스 단계	진단방법	개선방법	진단보고서	개선보고서
1	엠디	3D 구강 스캔 데이터	3D데이터	컨설팅	해당없음	해당없음	해당없음	해당없음
2	인더텍	인지재활프로그램	정형	컨설팅 품질진단	AI-Ready 기초진단	해당없음	보고서 완료	해당없음
3	한국교육학술정보원	교육행정 데이터 플랫폼	정형	컨설팅 품질개선	해당없음	텍스트 합성데이터	해당없음	시작 전
4	금강방재주식회사	소방안전관리점검	정형	컨설팅 품질진단	AI-Ready 기초진단	해당없음	보고서 완료	해당없음
5	아진산업	공장 안전 사람 탐지	이미지	컨설팅 품질진단 품질개선	이미지 데이터클리닉	이미지 중복제거	진단 완료	시작 전
6	케이원에코텍	수영장 익수	이미지	컨설팅 품질진단 품질개선	이미지 데이터클리닉	이미지 합성데이터	진단 완료	개선 진행 중
7	오션라이트에이아이	반려동물 문진	텍스트	컨설팅 품질진단 품질개선	AI-Ready 텍스트 데이터클리닉	텍스트 합성데이터	보고서 완료	시작 전
8	한국로봇산업진흥원	로봇 기업 현황 데이터	텍스트	컨설팅	해당없음	해당없음	해당없음	해당없음
9	빅웨이브AI	과수 농원 야상동물 합성데이터 이미지	이미지 합성데이터	컨설팅 품질진단 품질개선	이미지 데이터클리닉	이미지 합성데이터	진단 완료	개선 진행 중
10	한국OSG	CNC설비 센서 데이터 및 작업지시서	시계열 문서	컨설팅	해당없음	해당없음	해당없음	해당없음
11	우리로봇	로봇 관련 분야 인재 데이터	문서 정리안됨 개인정보	컨설팅	해당없음	해당없음	해당없음	해당없음
12	벽진BIO텍	섬유 후가공 생산 실적 데이터	문서	컨설팅	해당없음	해당없음	해당없음	해당없음
13	나래씨앤디	산모수첩 플랫폼, 산모 인바디	텍스트 이미지 정형	컨설팅 품질진단	AI-Ready	해당없음	데이터 대기	해당없음
14	옥타랩	심리상담 그림 이미지	이미지	컨설팅 품질진단 품질개선	이미지 데이터클리닉	논의 필요	데이터 대기	시작 전
15	신세릭스	공공데이터 활용 결합데이터	정형 데이터없음 결합데이터	컨설팅 품질진단	AI-Ready	해당없음	데이터 대기	해당없음
16	씨아피주식회사	나트륨 이차전지 음극재	데이터없음	컨설팅	해당없음	해당없음	해당없음	해당없음
17	엠제이소프트	공기압축기 예지보전	정형 데이터없음	컨설팅 품질진단	AI-Ready	해당없음	데이터 대기	해당없음
18	스피어에이엑스	CCTV 산불 데이터 (&공장 사고 분류)	이미지 동영상	컨설팅 품질진단 품질개선	이미지 데이터클리닉	이미지 합성데이터	진단 완료	개선 진행 중
19	서주사이언티픽	자산관리 데이터 활용 추천서비스	정형	컨설팅 품질진단	AI-Ready	해당없음	데이터 대기	해당없음
20	생각의탄생	유아 발화 및 행동 데이터 활용 도서추천	정형 텍스트	컨설팅 품질진단	AI-Ready	해당없음	데이터 대기	해당없음
21	에어스	정형외과 초음파 뼈 세그먼테이션	이미지 개인정보	컨설팅	해당없음	해당없음	해당없음	해당없음

+ New page

Product 01

2025년 6월, 조달청 혁신제품 지정



3D&VR 상품정보

판매자 관련 상품

물품식별번호 : 25618833

클라우드소프트웨어, 페블러스, 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 : **아래 사이트 참조**

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

인증정보 | 가격정보 | 규격정보 | 속성정보


인증번호 제 2025 - 123 호

조달청

혁신제품 지정 인증서

01 기 업 명 주식회사 페블러스
사업자등록번호 584-86-02422
02 주 소 대전광역시 유성구 대학로 99, 507호(궁동, 대전팁스타운)
03 혁신제품명 AI 학습데이터의 품질진단 및 개선통합형 데이터클리닉 솔루션
04 지정기간 2025년 06월 27일부터 2028년 06월 26일까지

위 제품은 「조달사업에 관한 법률 시행령」 제33조제1항제2호 및 「혁신제품 지정 및 구매촉진 등에 관한 규정」 제11조제1항에 따라 혁신제품으로 지정되었음을 인증합니다.

2025년 06월 27일

조 달 청


인증번호 제 2025 - 123 호

조달청

제품인증 대상 규격

연번	모델명	물품품목번호 (10자리)	물품식별번호 (8자리)	등록일	비고
1	DATA CLINIC	4323299901	25618833	2025.06.27	

혁신제품

Pebblous

Make Your Data AI-Ready

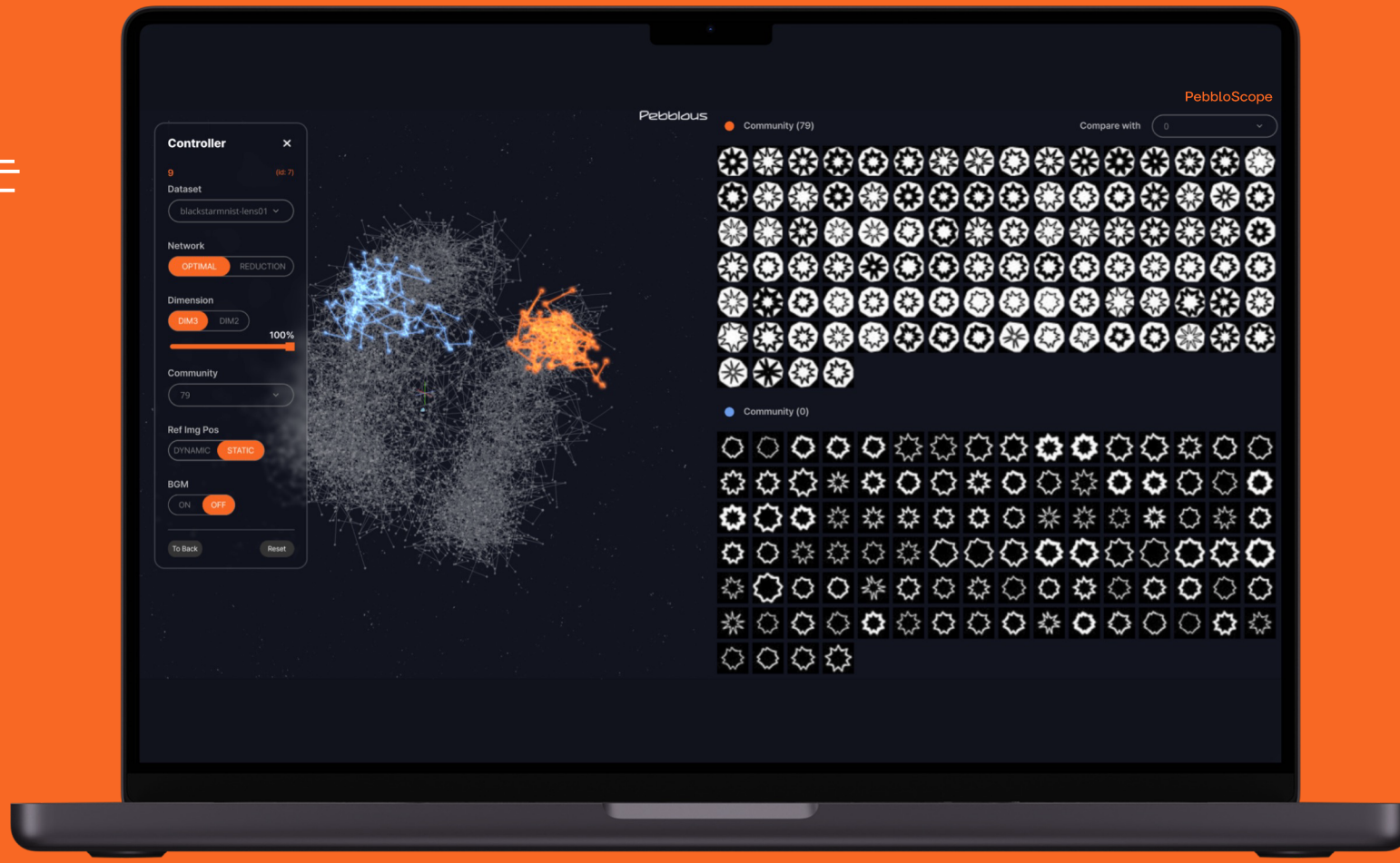
Source: Gartner

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.

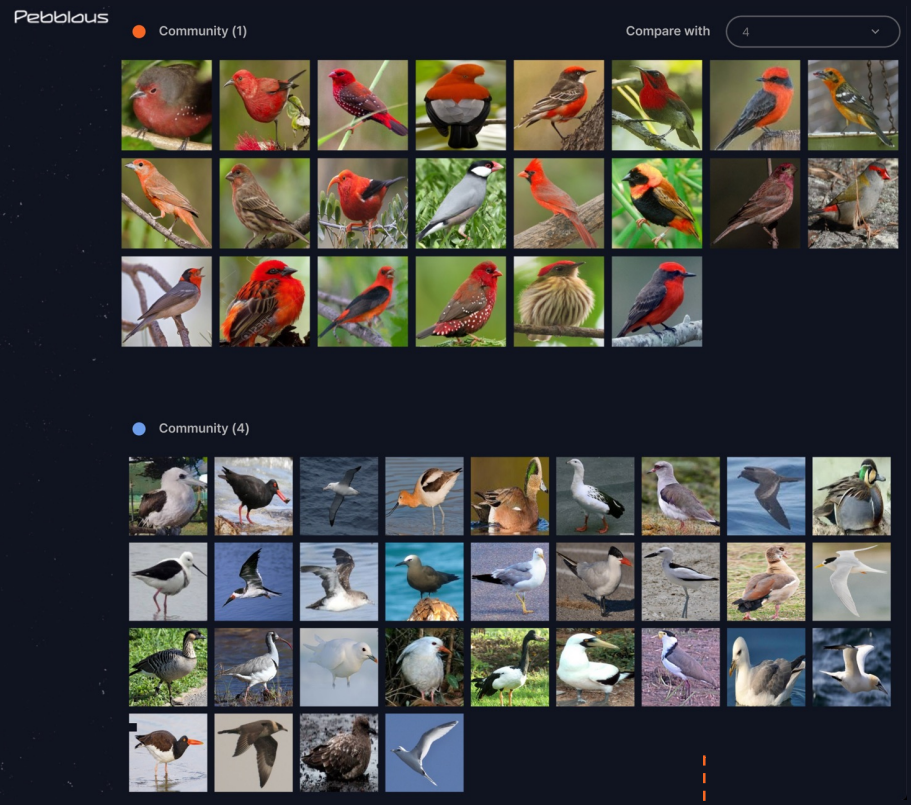


Product 02

페블로스코프

Pebblous

다양한 도메인의 데이터셋 목록



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.

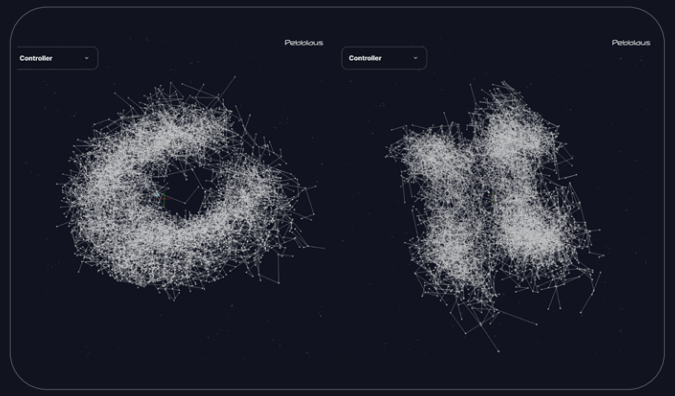
이미지 패널

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

Pebblous

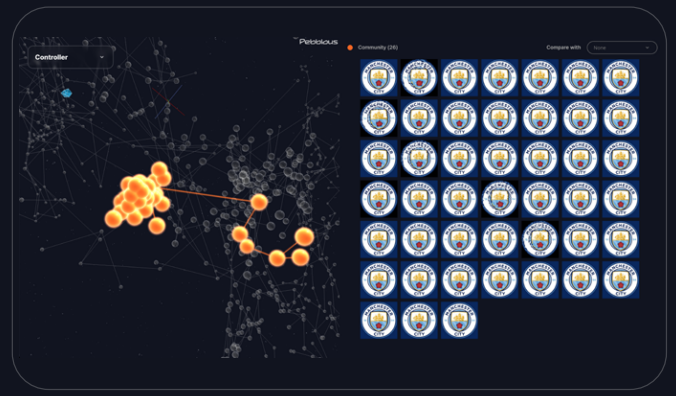
페블로스코프 특징



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.

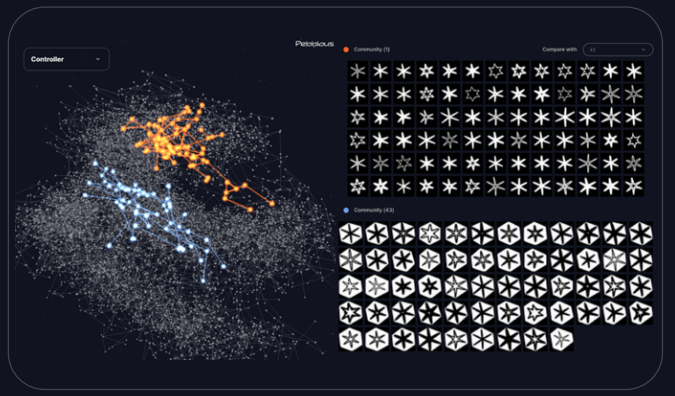
예) 관찰 시점 변경으로 클러스터의 형태 파악



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.

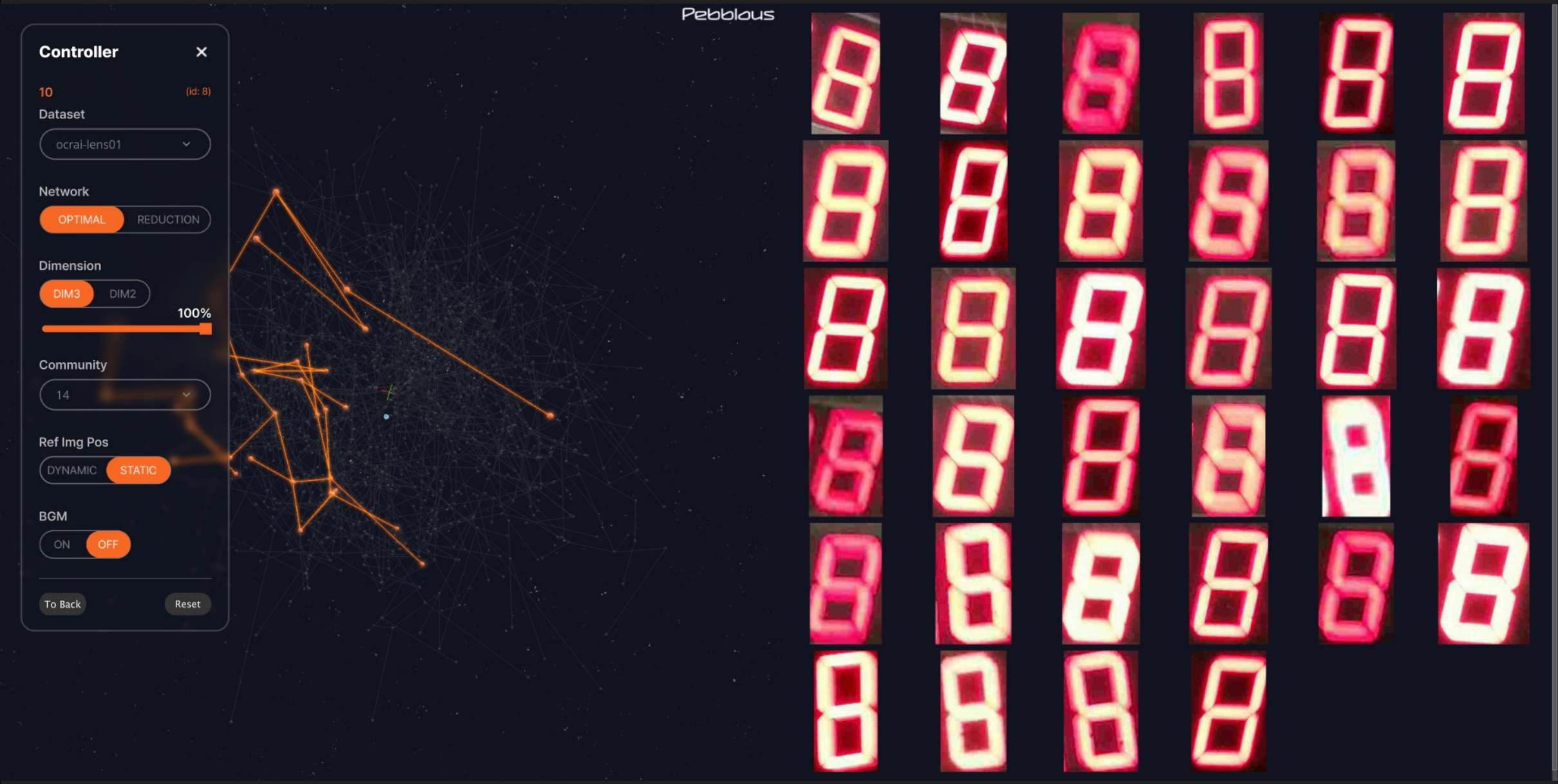
예) 캐글 로고 데이터셋의 과밀 클러스터의 확인

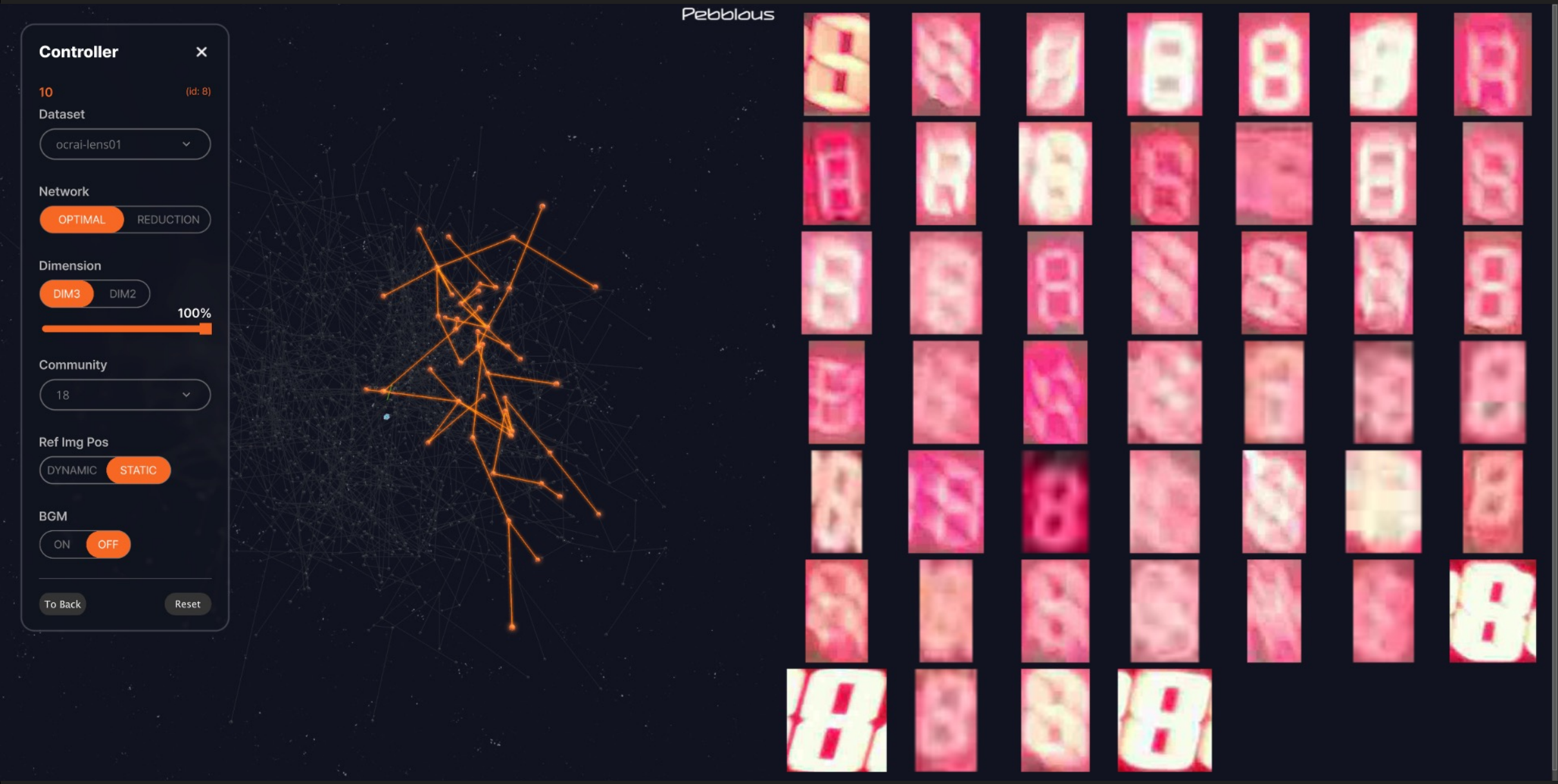


비교를 통한 데이터 특성 확인

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

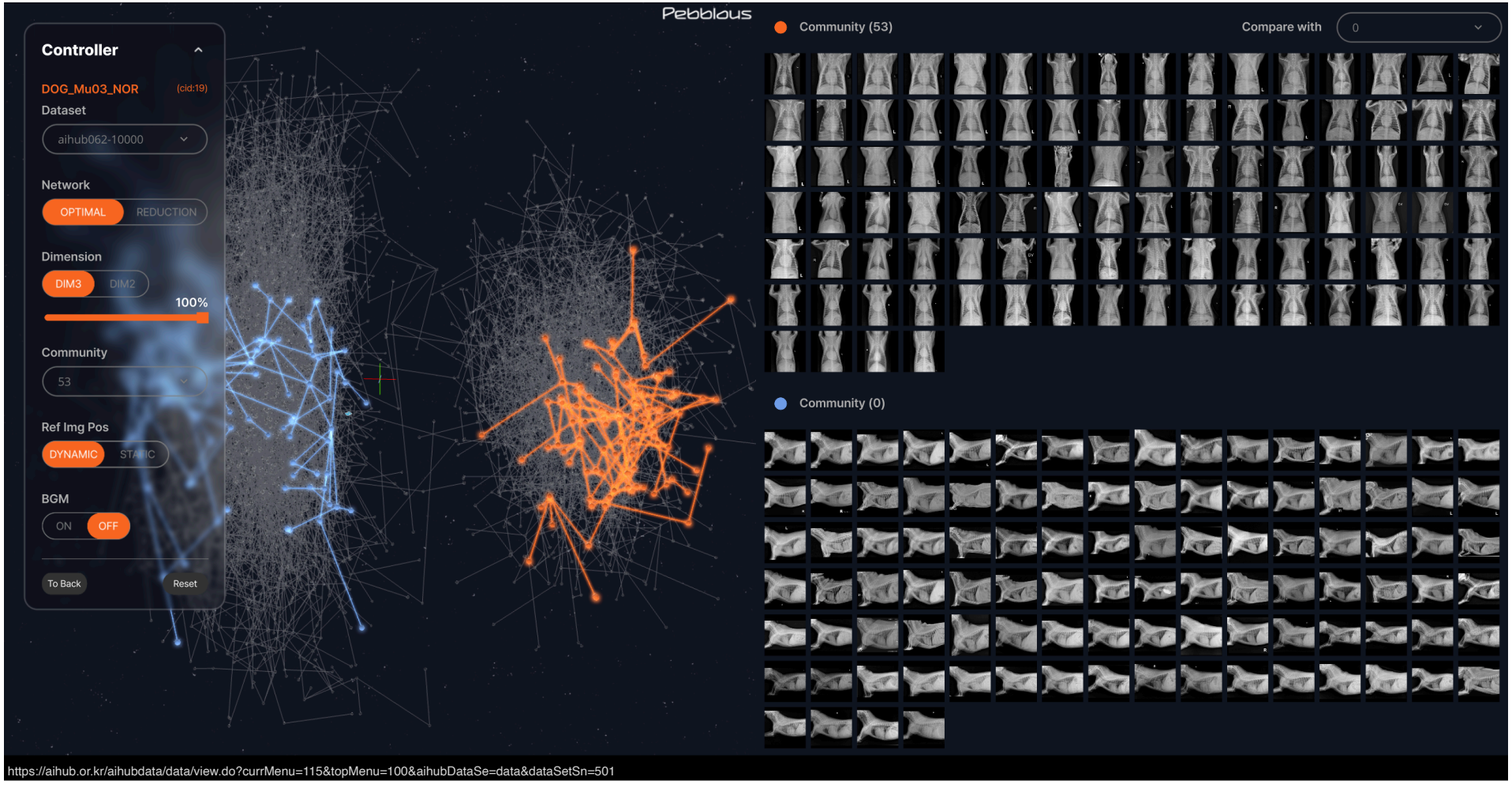
예) 동일 클래스 내의 유사 데이터 클러스터 비교


Pebblous
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

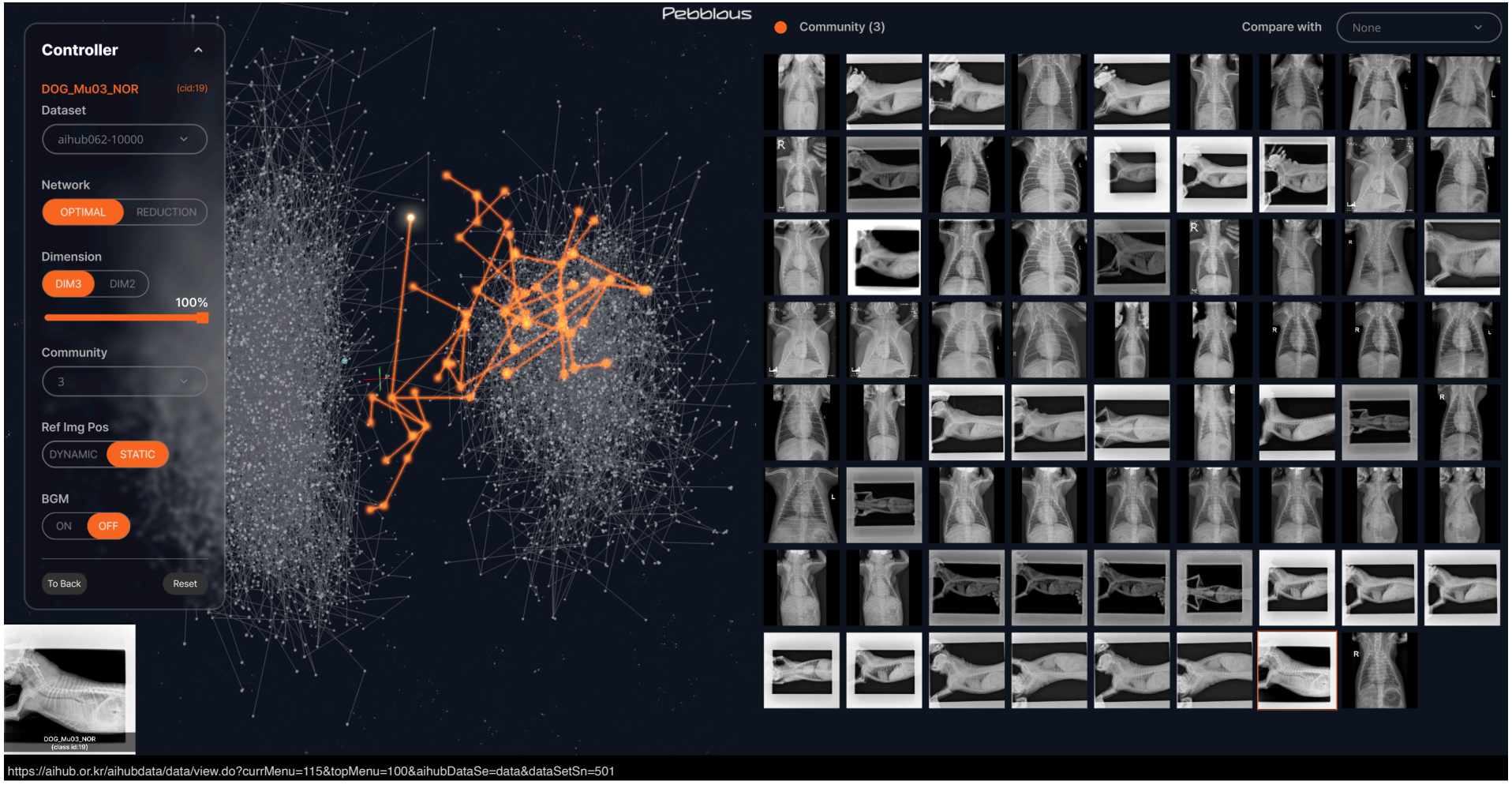
Pebblous

Pet-bone Dataset



Pebblous

Pet-bone Dataset



Menu

Dataset
Birds 450 species - IMAG...

Network
Optimization Reduction

Dimensions
2D 3D

Node color mapping
Normal

Selected community
4-4

Compare community
4-0

Node Tooltips
On Off

Reset

FLAME BOWERBIRD
(Class id : 206)

FLAME BOWERBIRD
(Class id : 206)

Data preview

Sort by
Density: mAlm (For nearest all da...

Size 5

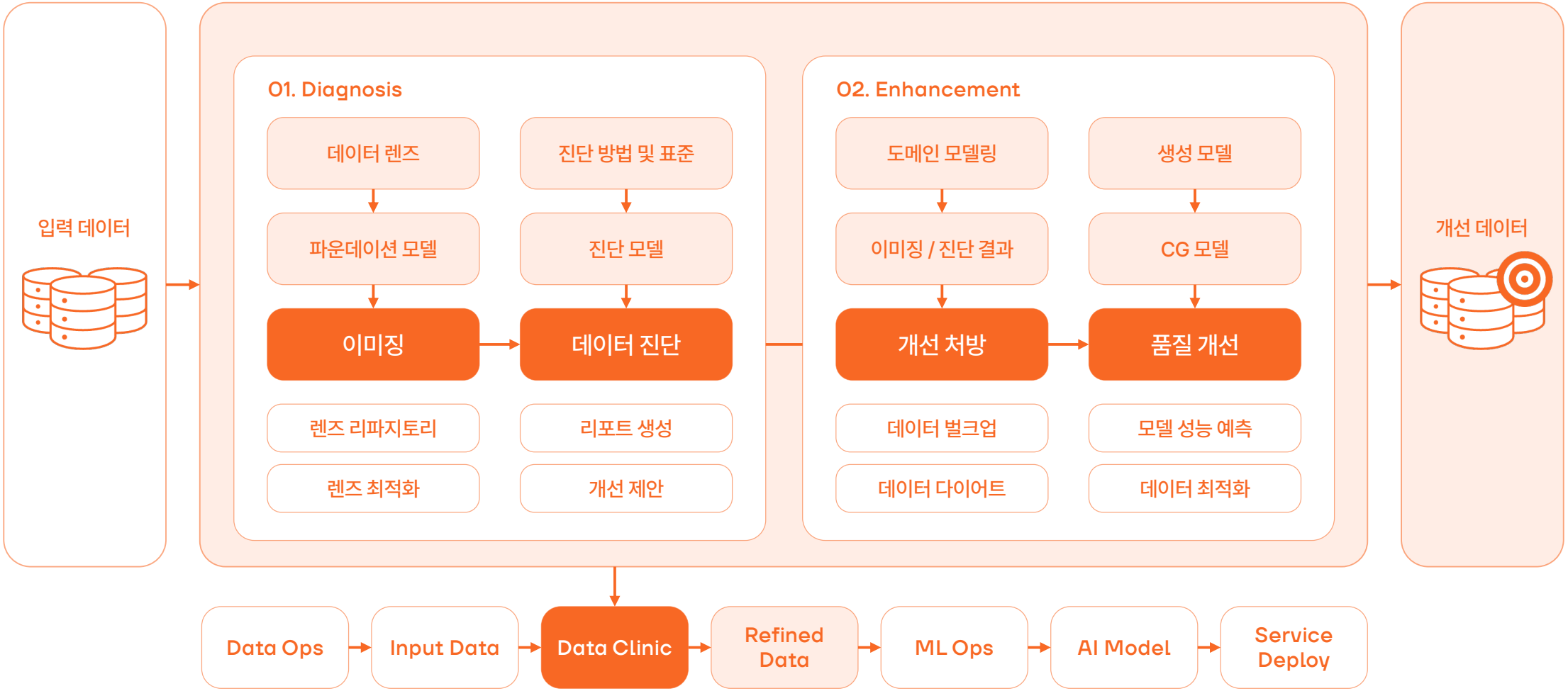
Community 4-4 Sample data (n=15)

Community 4-0 Sample data (n=15)

E-mail | contact@pebblous.ai Tel | 044-589-3826 Office | South Korea Sejong Seoul Daejeon Pohang Web | pebblous.ai

Pebblous

데이터 품질 진단 및 개선 솔루션



Pebblous

데이터 품질 개선 솔루션

비어있는 데이터 영역 · 중복 데이터 영역 · 실제 데이터 · 합성데이터

데이터 레플리카

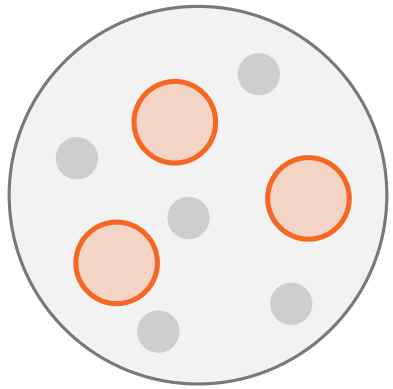
원본의 내용을 보호하면서도
통계적·분포적 특성이 유사한 가상데이터를 생성

→ 개인정보 유출 리스크 예방

활용 예) 개인정보 보호를 위한 데이터 익명화

데이터 벌크업

데이터가 부족한 부분에 데이터셋에 적합한
정밀 타게팅 합성데이터를 생성하고 제공

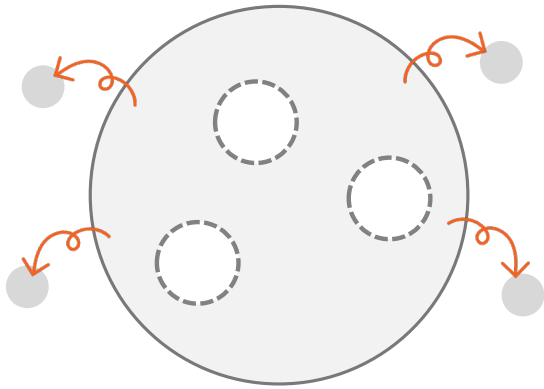


→ 데이터 수집 비용 절감, AI 성능 향상

활용 예) 예외 상황에 대한 합성데이터 생성

데이터 다이어트

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

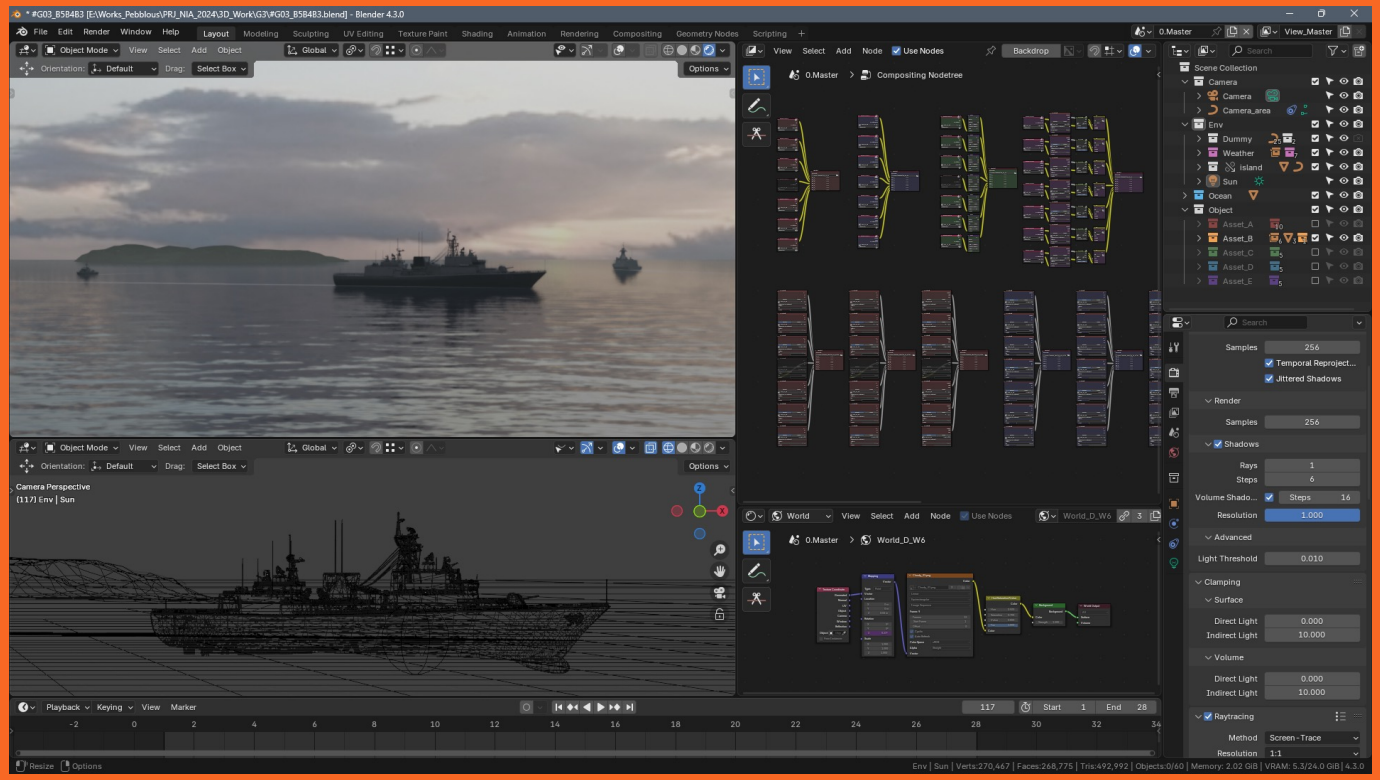


→ AI 학습 비용 절감, 개발 효율 제고

활용 예) 학습 과정에서의 데이터 경량화

합성데이터 (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.



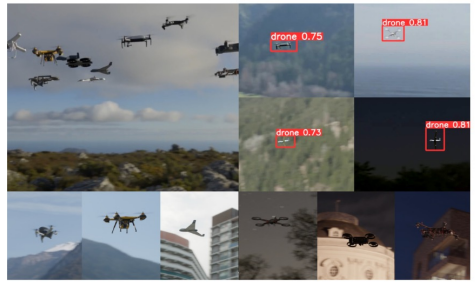
Pebblous

포트폴리오 웹페이지 링크

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

Single image 3D creation Download sample images

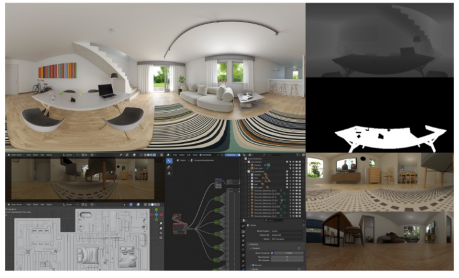
Drone Detection Download sample images



Characters and poses in special environments Download sample images

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

Robot autonomous driving field Download sample images

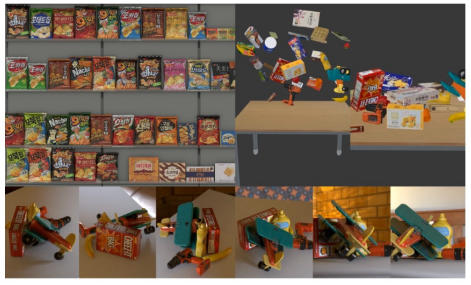


This is a dataset required for robots such as vacuum cleaners, and AI detects objects and plans optimal driving paths.

Diet monitoring Download sample images

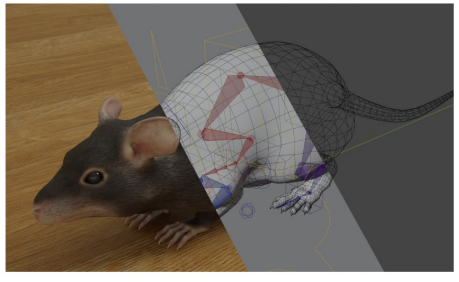
Synthetic data for meals for monitoring. Use hybrids of 3D and generated models

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.

animal behavior Download sample images

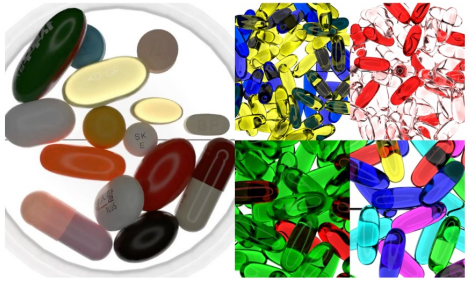


Synthesis data for AI learning to classify mouse behavior and recognize body parts

Livestock sector Download sample images

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

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

Waste Plastic Recycling Classification AI Synthetic Data Download sample images

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

Pebblous

해병대 연평도/백령도 경계 AI 학습용 합성데이터: 결과물(주간)

Pebblous

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 구분 어선 군함 상선 고정익 유인항공기 회전익 유인항공기 무인항공기 새 삐라 오물폭탄	[그림1] 참조	-
다양성 (요건)	시간대별 분포	구성비 중첨률	구성비 중첨률 50% 이상 목표 구성비: 주간 50%, 야간 50%	구성비 중첨률 96.36% 구분 / 비율(%) / 수량: 주간 51.86 77,496, 야간 48.14 71,950	달성
	계절별 분포	구성비 중첨률	구성비 중첨률 50% 이상 목표 구성비: 하절기 75%, 동절기 25%	구성비 중첨률 94.35% 구분 / 비율(%) / 수량: 하절기 72.09 107,739, 동절기 27.91 41,707	달성
	기상별 분포	구성비 중첨률	구성비 중첨률 50% 이상 목표 구성비: 맑음 14%, 흐림 14%, 비 14%, 눈 14%, 해무 42%	구성비 중첨률 80.46% 구분 / 비율(%) / 수량: 맑음 15.97 23,872, 흐림 15.98 23,887, 비 15.98 23,875, 눈 4.28 6,393, 해무 47.79 71,419	달성
	황천 급별 분포	구성비 중첨률	구성비 중첨률 50% 이상 목표 구성비: 황천 1급 14%, 황천 2급 14%, 황천 3급 14%, 황천 4급 14%, 황천 5급 14%, 황천 6급 15%, 황천 7급 15%	구성비 중첨률 98.00% 구분 / 비율(%) / 수량: 황천 1급 14.20 21,227, 황천 2급 14.29 21,355, 황천 3급 14.29 21,363, 황천 4급 14.31 21,391, 황천 5급 14.33 21,422, 황천 6급 14.28 21,338, 황천 7급 14.29 21,350	달성

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%	달성
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

Pebblous

NIA 초거대AI 합성데이터: '우수' 평가

최종평가 결과 및 보완 요구사항

과제명		32. 군 경계 작전 환경 데이터	**사업(협약)기간**	2024. 07. 01~2024. 12. 31
데이터셋명		51. 군 경계 작전 환경 내 인식 데이터		
		52. 군 경계 작전 환경 합성 데이터		
수행 기관	**주관**	흥일기업(주)	**총괄책임자**	박준영
	참여	인피닉, 페블러스, 한밭대학교 산학협력단		

종합 평가등급	매우우수 (90점이상)	우수 (89~80점)	보통 (79~70점)	미흡 (69~60점)	매우미흡 (60점미만)
		○			

평가의견 (위원1)	- 본 과제는 주간 데이터로 최적화되어 있으므로 야간 데이터의 보완이 필요함.
평가의견 (위원2)	- 결과 발표 준비가 우수하고, 구축 데이터의 품질이 양호함.
평가의견 (위원3)	- 데이터 구축 결과물의 내용을 쉽게 파악할 수 있도록 선박/사람 등 개체별, 기상현 황 (해무 짙은 정도) 별 데이터 갯수 요약표를 데이터 활용 관점에서 정리 필요.
평가의견 (위원4)	- 해병대 데이터 제공에 어려움이 있었으나 기간 내 수행이 적절히 진행됨
평가의견 (위원5)	- 군으로부터 수집된 영상에 대한 상게 분류체계의 제시가 필요합니다. - 데이터의 분류체계는 32-51, 32-52의 내용에 대해 일괄성 유지가 필요합니다.

합성데이터 포트폴리오

국방AI 합성데이터

합성데이터 포트폴리오

국방AI 합성데이터 (우천 상황)

합성데이터 포트폴리오

국방AI 합성데이터 (눈 상황)

Pebblous

국방AI 합성데이터 (포지셔닝)

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 데이터 클리닉 서비스를 제공하여, 다양한 컴퓨터 비전 학습 응용을 위한 합성데이터를 생성할 수 있도록 정밀한 품질 진단을 수행합니다.

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

Electric 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

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

Traction

가트너 보고서

이미지 합성데이터 기업

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.

Gartner

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

Value proposition

Data Clinic | 가치 제안

Pebblous

01

Fast and Accurate

10만개 데이터셋 기준

1시간 쾌속 품질 평가

AI 훈련 데이터셋의 품질 평가 및 시각화

02

AI Performance

5% 합성데이터로

2% 모델 성능 향상

품질 진단 기반 정밀 타기팅 합성데이터 생성

03

Dev Efficiency

80% 데이터 경량화로

x5 GPU 효율 향상

중복 데이터를 줄여 학습 효율 최적화

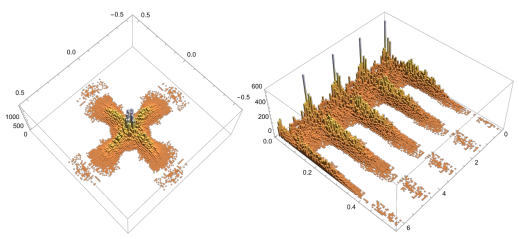
Traction

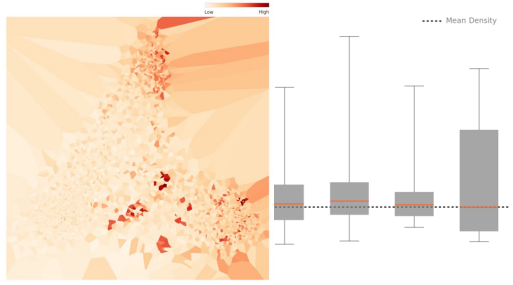
Pebblous

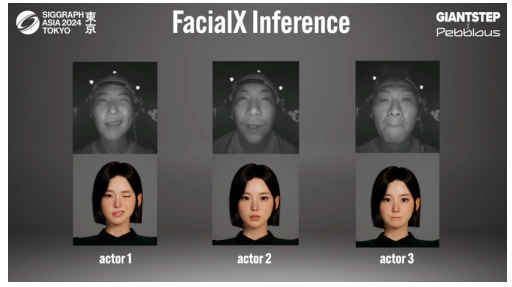
ROI 중심의 고객 사례

고객사 정보보안을 고려하여 대략의 근사치 및 내부에서 실험한 유사 사례로 설명함

	현대차	한화비전	자이언트스텝	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 페이셜 캡쳐 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 기존 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

프로젝트 명	수행 연도	적용 기술	현업의 문제 및 요구사항	해결 방법 및 결과
주행 경로 추정 모델을 위한 합성데이터	2022년	시뮬레이션 합성데이터 신경망 모델 학습	대형 주차장 내에서 GPS가 끊기면서 생기는 주행 경로 오차에 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 최소가 되도록 다수의 레인에 차량을 인입하고 인출하는 최적화 문제 해결 요청	시뮬레이터 상에서 수치해석 모델 대비 25% 성능을 개선하는 강화학습 알고리즘 개발
광명공장 조립 공정 최적화를 위한 강화학습 모델	2025년	시뮬레이터 환경에서의 강화학습	조립 공정에서 차량 모델 전환이 최소가 되도록 다수의 레인에 차량을 인입하고 인출하는 최적화 문제 해결 요청	시뮬레이터에 복잡한 조립 공정 프로세스를 구현하고, 차량 모델 전환율을 최적화 하는 강화학습 알고리즘 개발

Pebblous

페블러스 주요 고객사

KIT 국가독성과학연구소 Korea Institute of Toxicology

dip 대구디지털혁신진흥원 Daegu Digital Innovation Promotion Agency

KEPCO 전력연구원

(주)씨에이에스(C·A·S) Computer Assurance Services

SAMSUNG E&A

HYUNDAI MOTOR GROUP

Hanwha Vision

GIANTSTEP

LG전자

ETRI

Nota AI

KAIST

IGT

DDAKPLE

SIA SI ANALYTICS

TKG

AI Factory

HelloDD.com

HANKOOK driving emotion

SUNGKYUNKWAN UNIVERSITY(SKKU)

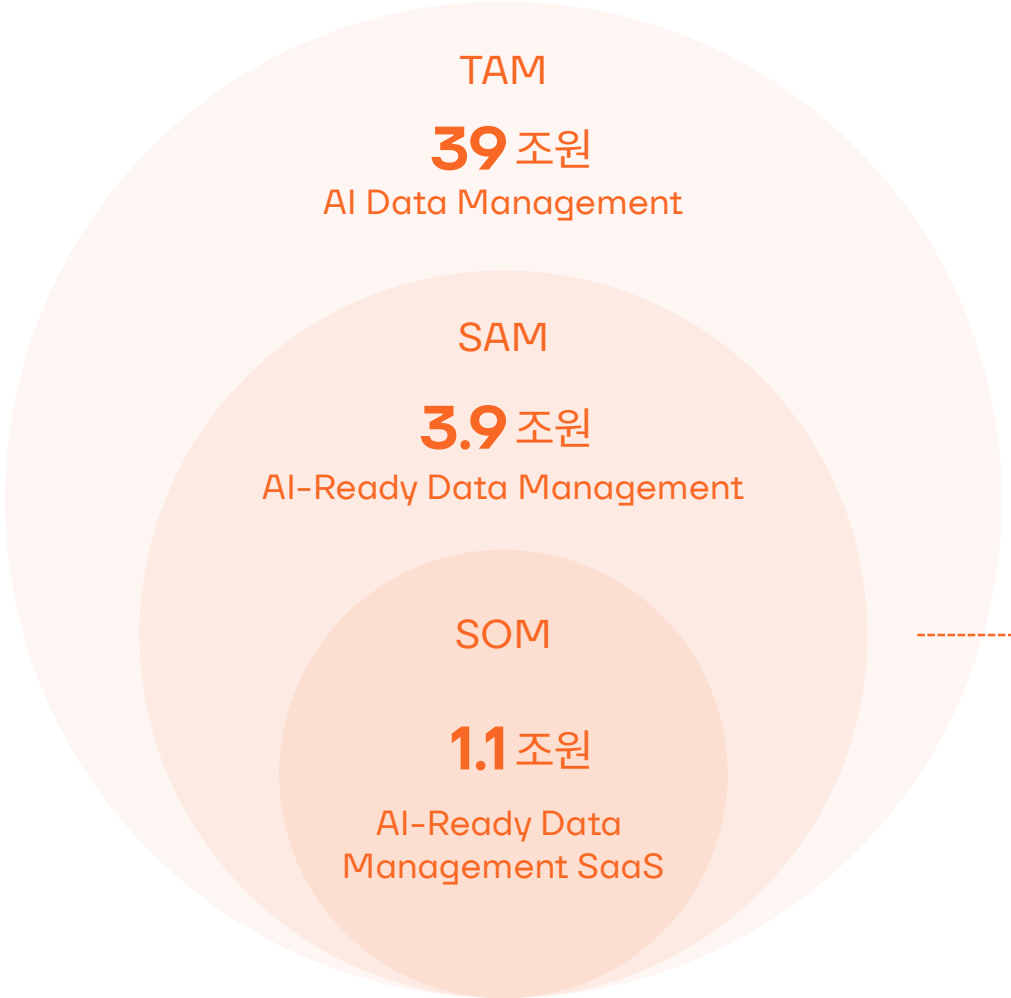
impact station

페블러스 주요 고객사

MYEONG MOON BIO & DENTAL CO., LTD.
InTheTech
Syntherixs
K1ECOTECH
금강방재주식회사
Rootlab
OceanLightAI
MJSOFT 엠제이소프트
OCEAN KOREA GLOBAL TECHNOLOGY COMPANY
한국OSG
CIPEnergy
벽진BIO텍
서주사이언티픽
KERIS 한국교육학술정보원
KIRIA 한국로봇산업진흥원 KOREA INSTITUTE FOR ROBOT INDUSTRY ADVANCEMENT
Narae CnD
AJIN INDUSTRIAL CO. LTD.
AIRS
BIG WAVE
SPHERE AX
옥타랩 OCTA LAB
우리로봇
생각의 탄생

Pebblous

AI-Ready Data Management 시장 타기팅



TAM의 목표 점유율
1%

반영한 수익 시장
3,900억원

Pebblous

데이터 클리닉 적용 도메인

- 모빌리티
- 스포츠 산업
- 메타버스
- 자원 재활용
- 금융
- 패션
- 제조물류
- 국방
- 제약/의료

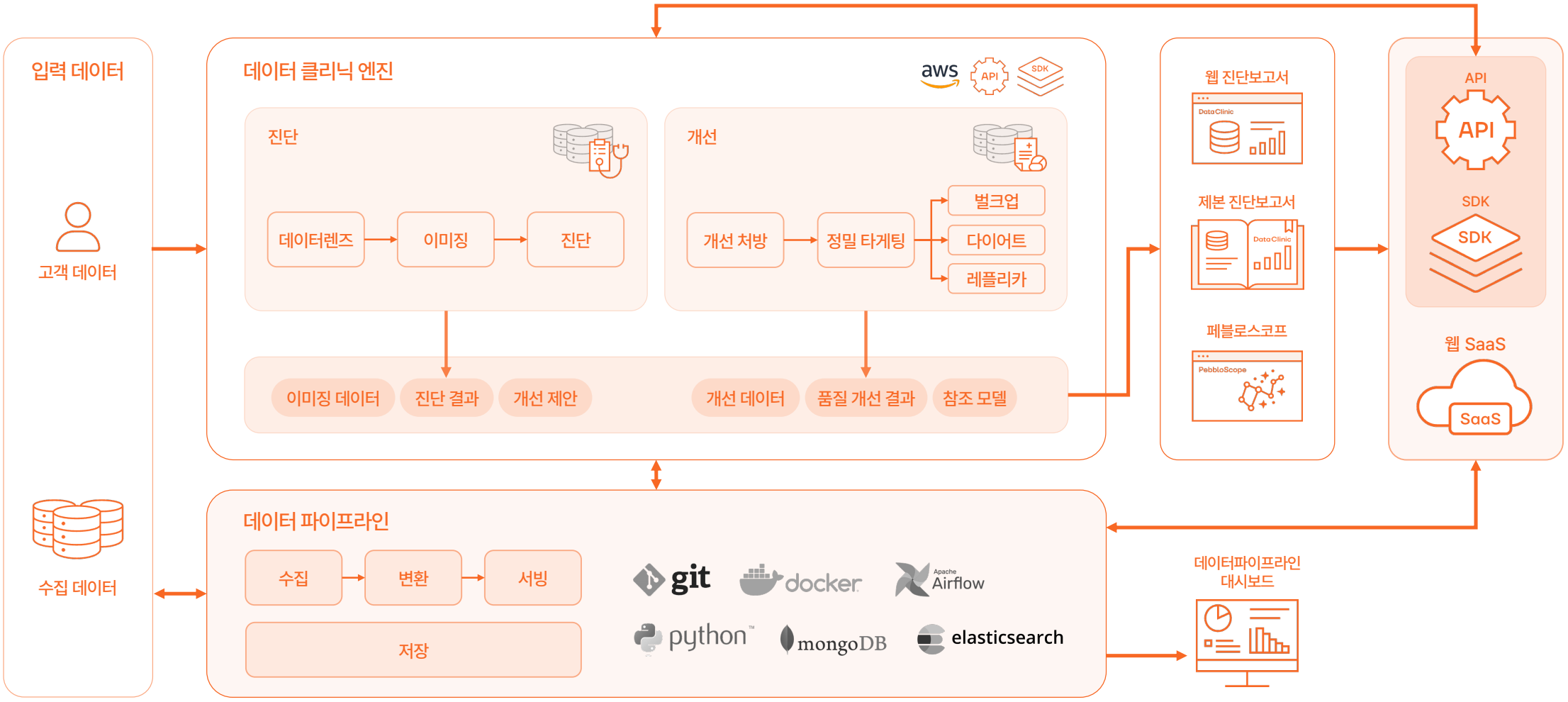
인공지능 데이터 클리닉

DataClinic

- 데이터 이미징
- 데이터 품질 진단
- 개선 처방
- 품질 개선

Pebblous

데이터 클리닉 SaaS 모델



데이터 클리닉 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
데이터 클리닉 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 진단보고서	✓ 웹 진단보고서	✓ 웹 진단보고서
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 1,000만원
데이터 벌크업				✓ 커스텀 프로젝트로 진행, 최대 10K 이미지, 주석포함 기준 6,000만원
온프레미스 설치				✓ 고객사 내부 사용 기준. 5억원/년. HW 별도

Competitor

Pebblous

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	**다양항 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.

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

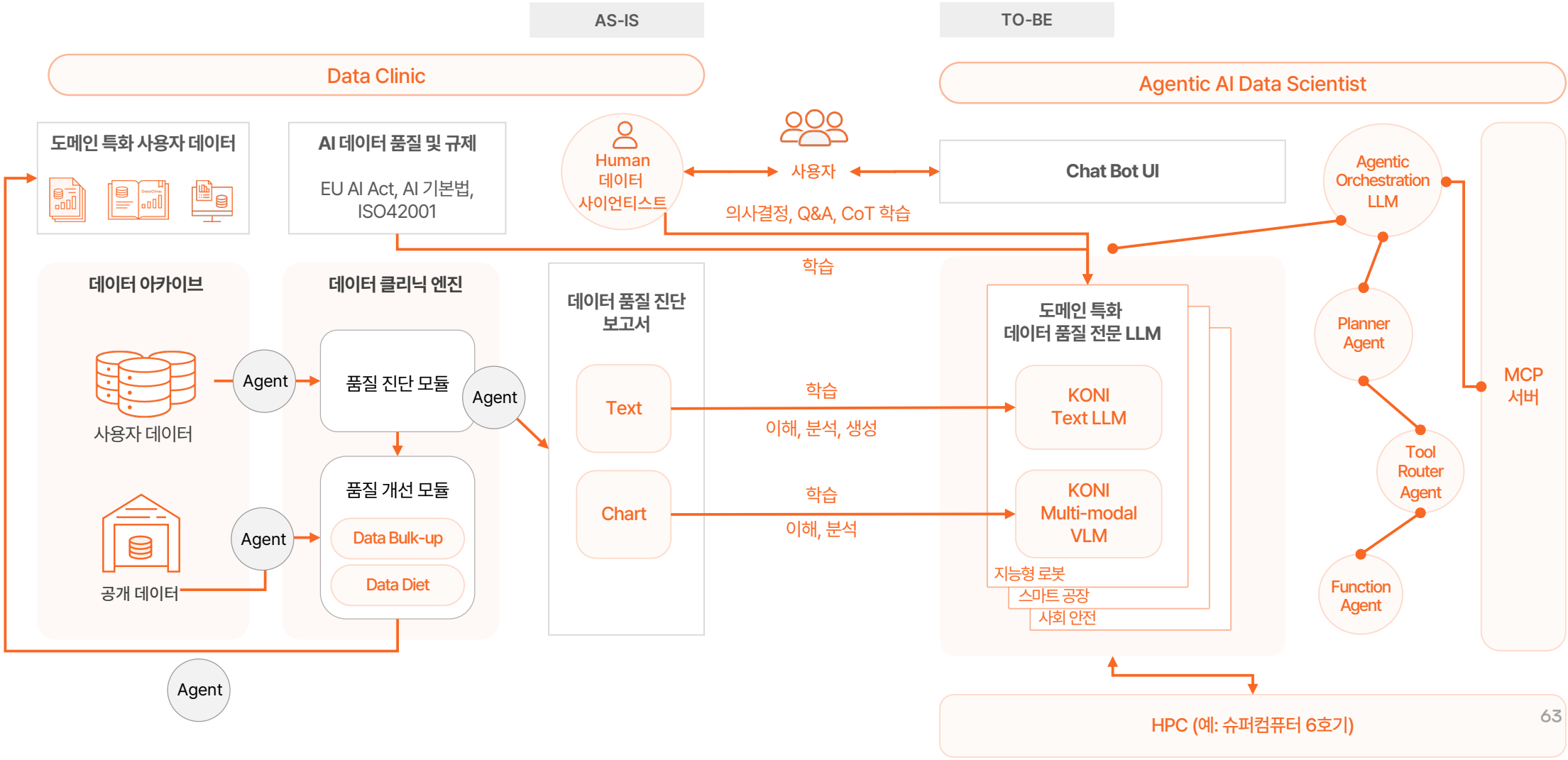
01 데이터의 AI 적합성을 지속적으로 평가하고 개선하는
차세대 거버넌스를 위한 **데이터 매니지먼트 솔루션**

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

Pebblous

솔루션 확장: 자율형 데이터 클리닉 Agentic Data Clinic



Pebblous

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

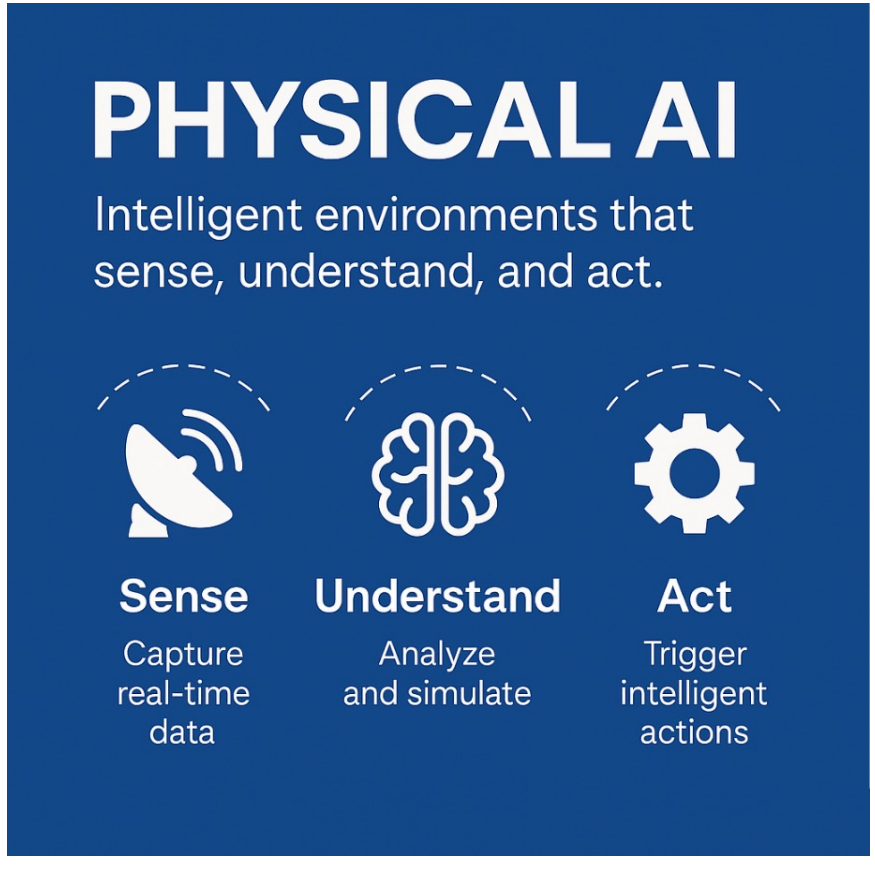
Initiatives:Emerging Technologies and Trends Impact on Products and Services; Increase Product Traction

> 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



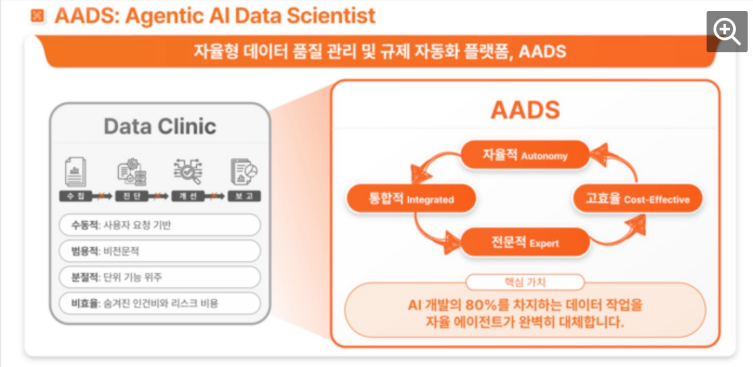
2025년 8월, 글로벌 빅테크 육성사업 선정

페블러스 주관으로 KISTI와 컨소시엄을 이루어 대형과제를 수주하여, AI 에이전트 개발 진행중

전자신문 | etnews

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

발행일 : 2025-08-21 09:18



페블러스 데이터 클리닉에서 진화한 AADS

AADS는 데이터 관리 모든 단계 엔드투엔드 자동화된 시스템이다.

KISTI와 협력해 AADS 플랫폼 공동 개발...총사업비 61억원 규모

데이터 품질·규제 대응·해외 시장 확장 목표

2025년 7월,

‘국가대표 AI’ 사업 출사표

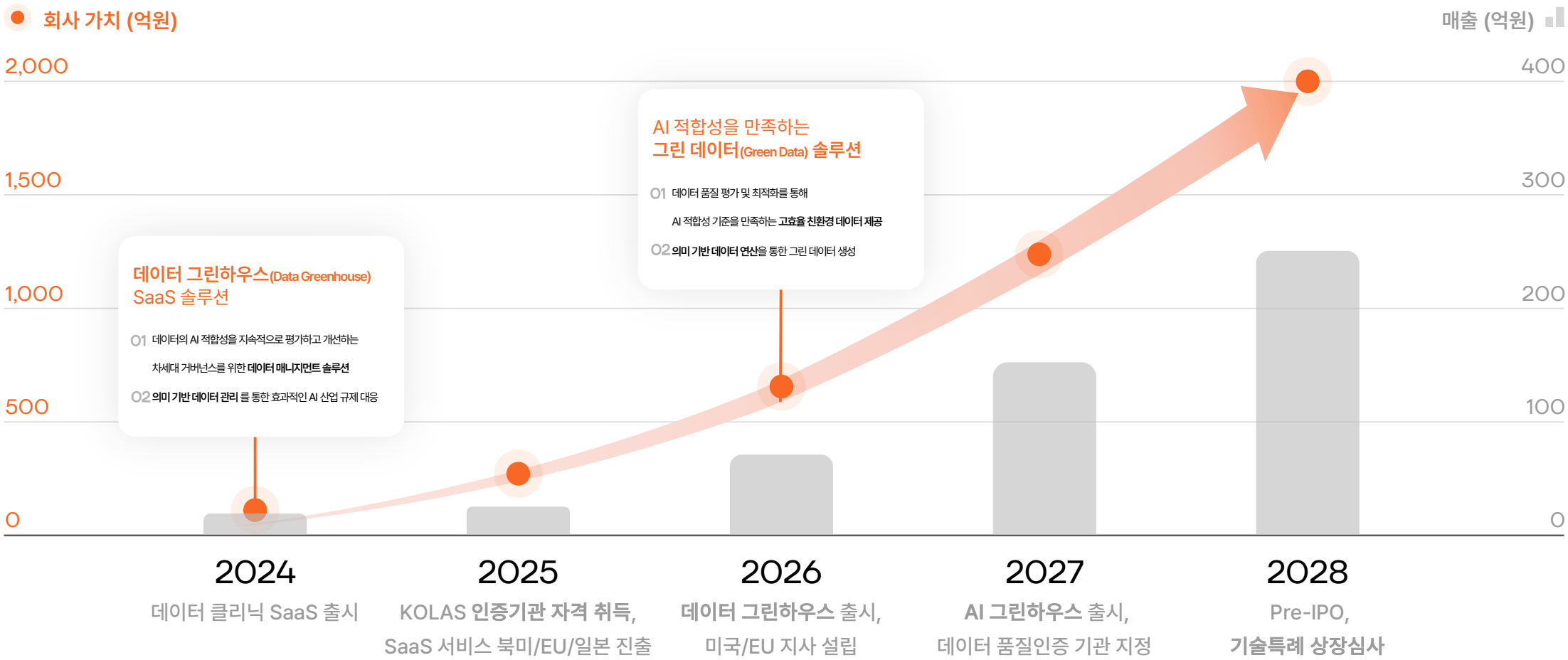
코난테크놀리지 컨소시엄에
참여하여 2차 발표평가 선정 후
최종탈락

페블러스의 역할인 “모델 취약성
분석 기반 타겟 합성데이터 생성”은
좋은 평가를 받음

평가의견: 모델 취약성 분석 기반 타겟 데이터 합성을 통한 개선 계획이 우수함

Pebblous

SaaS based Revenue



Pebblous

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
AI/Data Engineer	4
CG Engineer	1
SW Engineer	3
UI/UX Designer	1
Back Office	3

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

POSTECH 컴퓨터공학 박사
서울대 공학박사 수료
KAIST 전산학과 졸업

Company history

페블러스 연혁

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

2021

Oct 팁스(TIPS) 지원사업 선정

Nov (주)페블러스 법인 설립
Seed 투자 유치 (I)

2022

Feb 벤처기업 인증

Apr 기업부설연구소 설립
Seed 투자 유치 (II)

May 데이터바우처 사업 4건 매출 계약

Oct Pre-series A 투자

Nov '도전! K-스타트업 2022' 왕중왕전

2023

Jun 신용보증기금 퍼스트펭귄 선정

Sep 'DataClinic' 베타 오픈

Oct SWITCH 2023 부스 전시 (싱가포르)

2024

Jan CES 2024 부스 전시 (미국, 라스베가스)

Mar TIPS 성공 종료
현대자동차 ZERO1NE 크리에이터 선정

Apr IITP SW컴퓨팅산업 원천기술개발사업 선정

May NIA 초거대AI지원사업 국방 수요 합성데이터 생성

Jun 넥스트라이즈 2024 부스 전시 (코엑스)

2025

Jan CES 2025 부스 전시 (미국, 라스베가스)

Mar MWC 2025 부스 전시 (스페인, 바르셀로나)

June 대구디지털혁신진흥원 데이터 품질 진단 및 개선 사업 수주

Aug 조달청 혁신제품 지정

글로벌 빅테크 육성사업 선정: Agentic AI Data Scientist 개발

Fabulous Data With

Pebblous

Better Data Makes Better AI

Pebblous.ai