Pebblous Makes Data Tangible

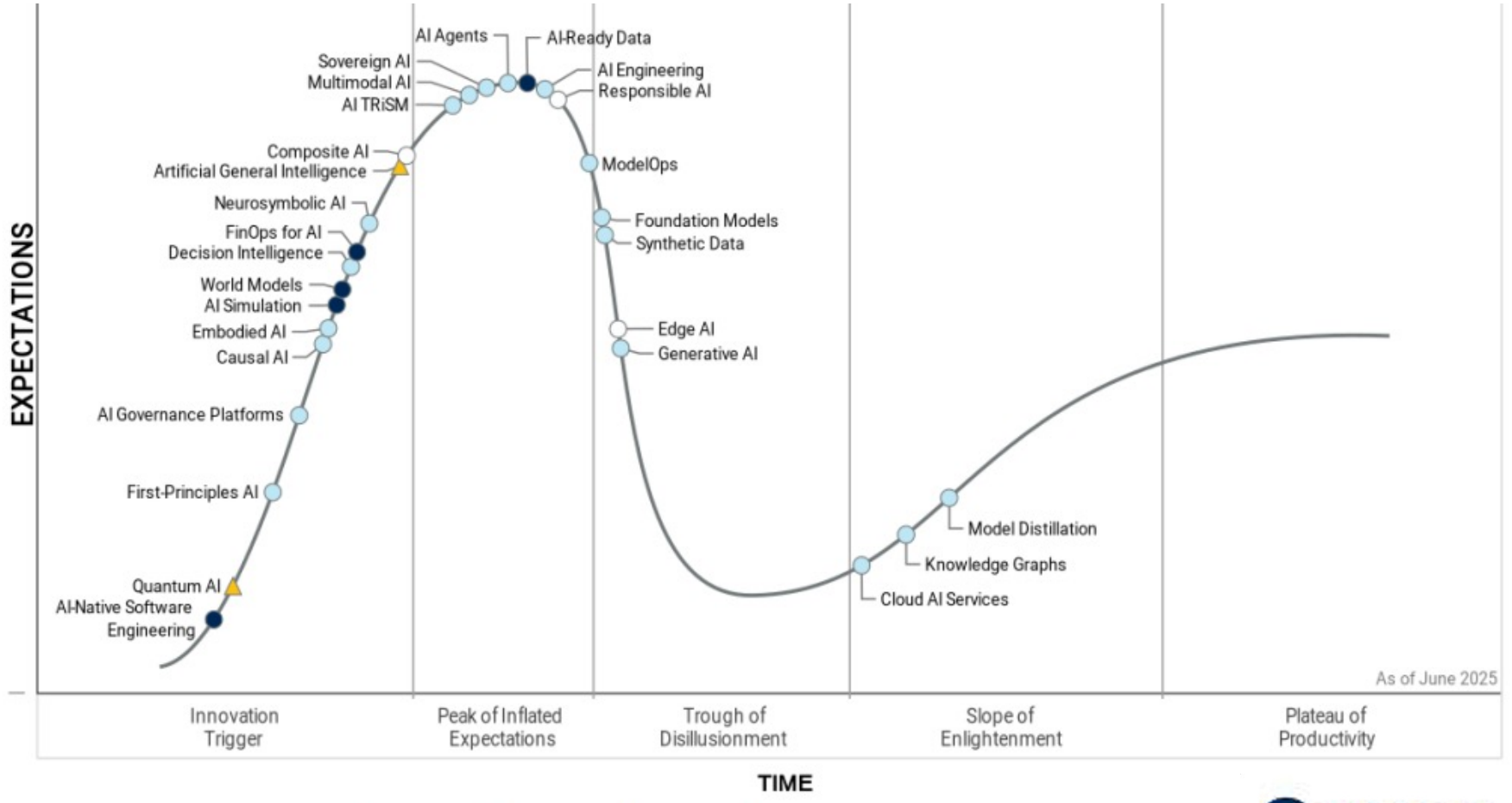
Pebblous.ai

Pebblous

Better Data Makes Better AI

Pebblous

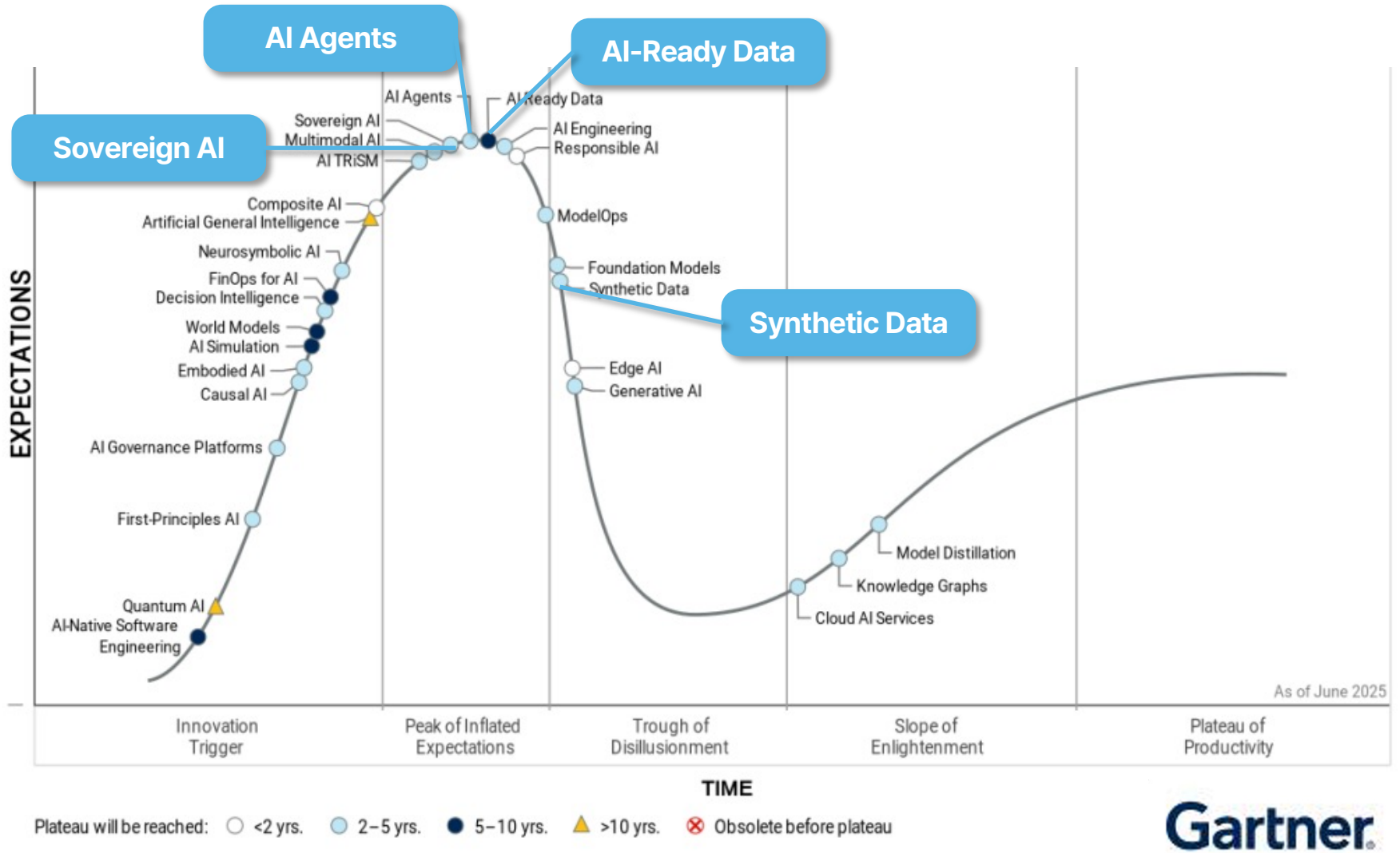
Hype Cycle for AI, 2025



Gartner

Pebblous

Hype Cycle for AI, 2025



The era of Sovereign AI calls for a complete reconfiguration of data evaluation.

Professor Yejin Choi (Stanford University)

2025-09-25, UN Security Council

"Current AI models are heavily centered on English, with significantly lower performance in other languages. They also exhibit a narrow understanding of cultural values, leaving many communities excluded from the benefits of AI.

This issue cannot be solved in a piecemeal way — it is a fundamental problem. Therefore, we must change it from the ground up: the training data, the learning objectives, and even the evaluation methods. Only then can we build AI that is truly capable across diverse languages and contexts."

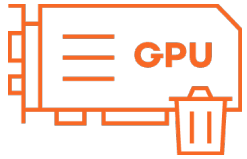
Pebblous

Pains from Data Quality

01

Poor Data Quality in AI Datasets

02



Wasted GPU Resources and Time

03

Increasing **AI Regulations**

Example Impact: EU AI Act imposes fines of 7% of global revenue or $35 million for non-compliance.

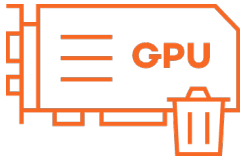
Pebblous

Continuous Quality Assessment and Improvement for **AI-Ready Data**

01

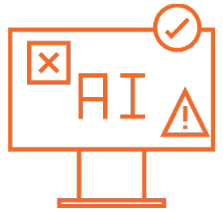
Quality Data Enhances **Model Performance**

02



Optimal Data Enhances **Dev Efficiency**

03

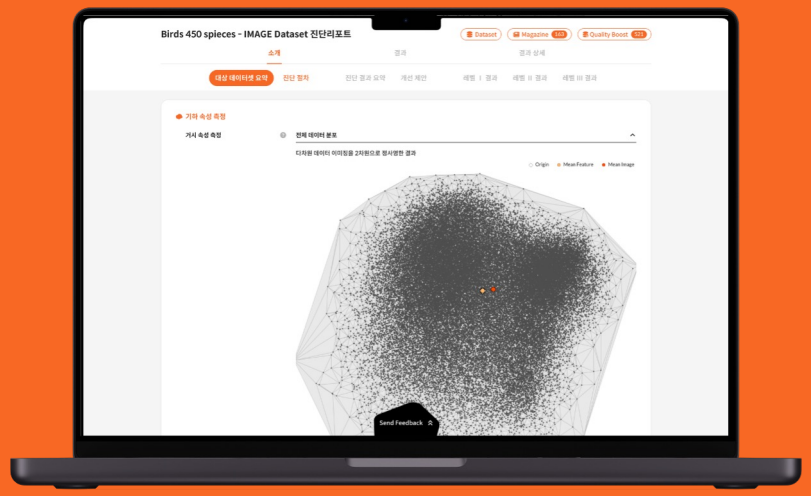


AI-Ready Data Enhances **Data Governance**

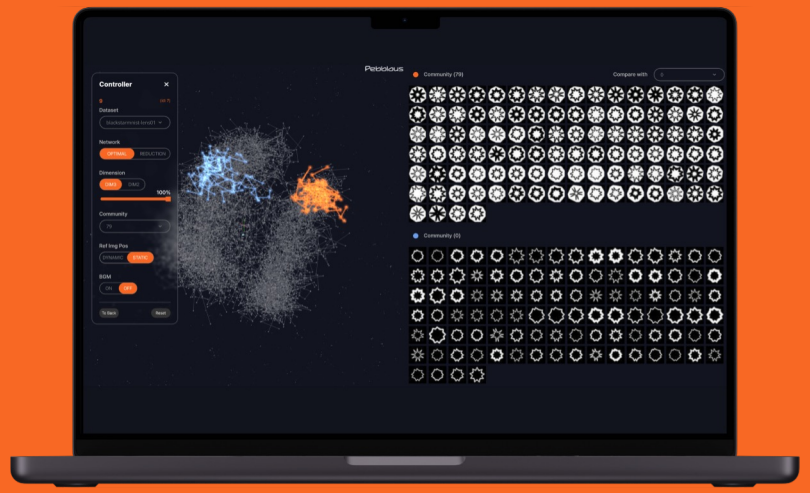
All-in-one Data Management Solution

(1) AI Ready Data,
(2) Observability,
(3) Semantics Layer,
(4) Multi-modality,
(5) SaaS/On-Prem/API

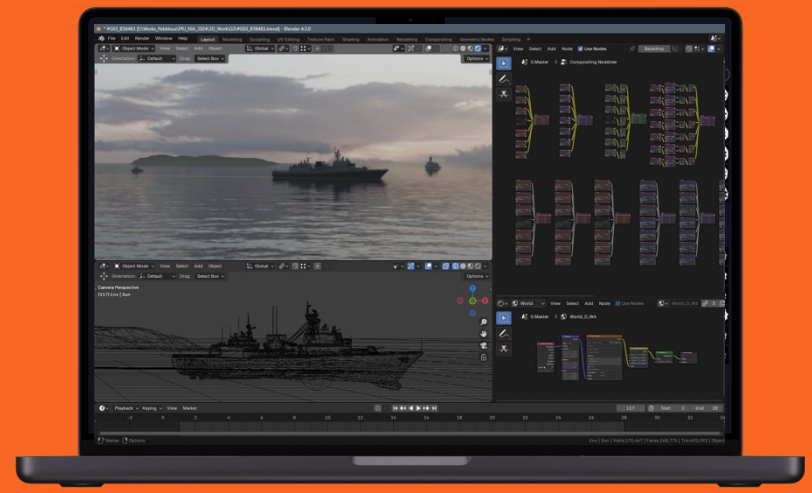
DataClinic



PebbloScope



Synthetic Data



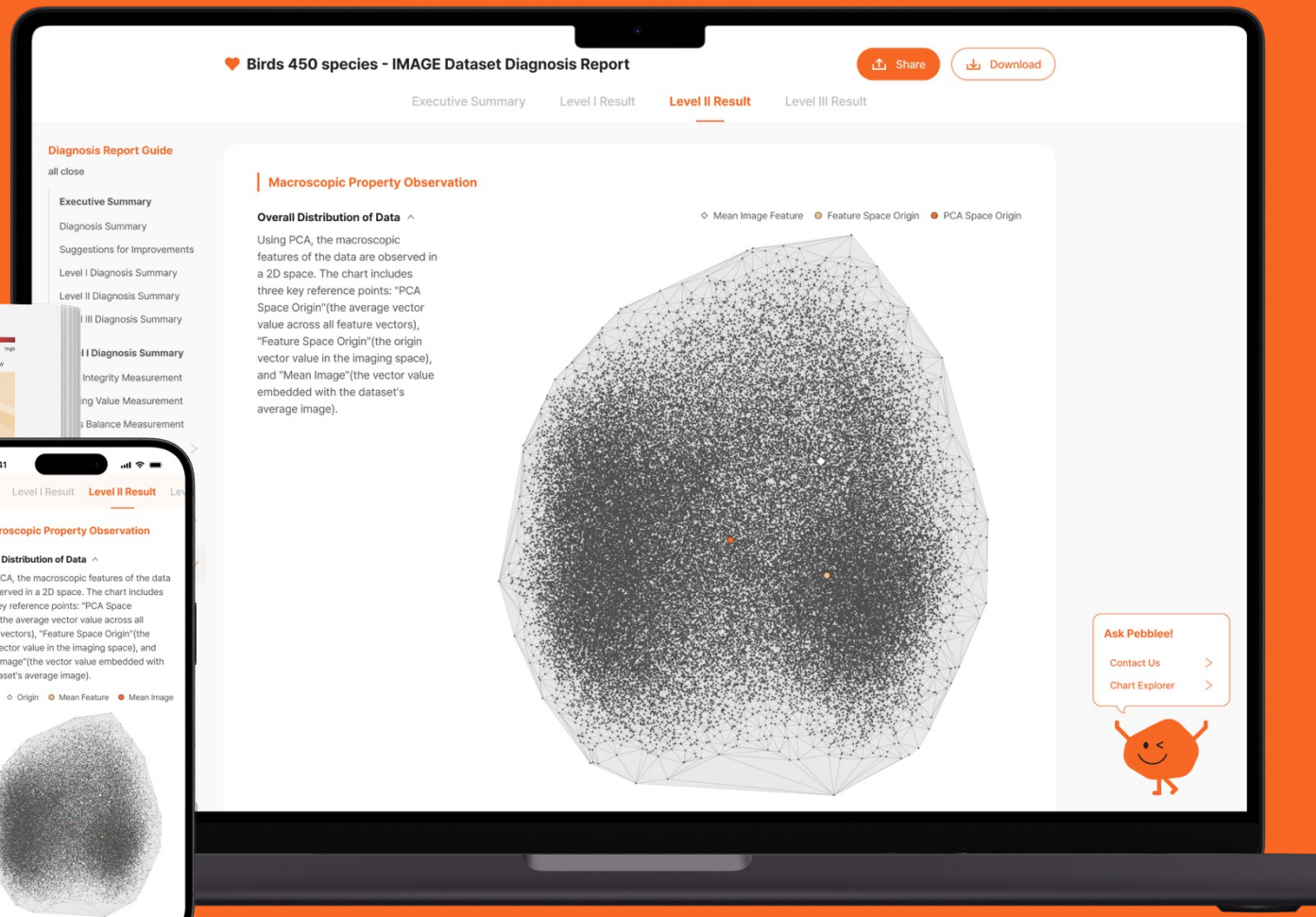
Product 01

DataClinic

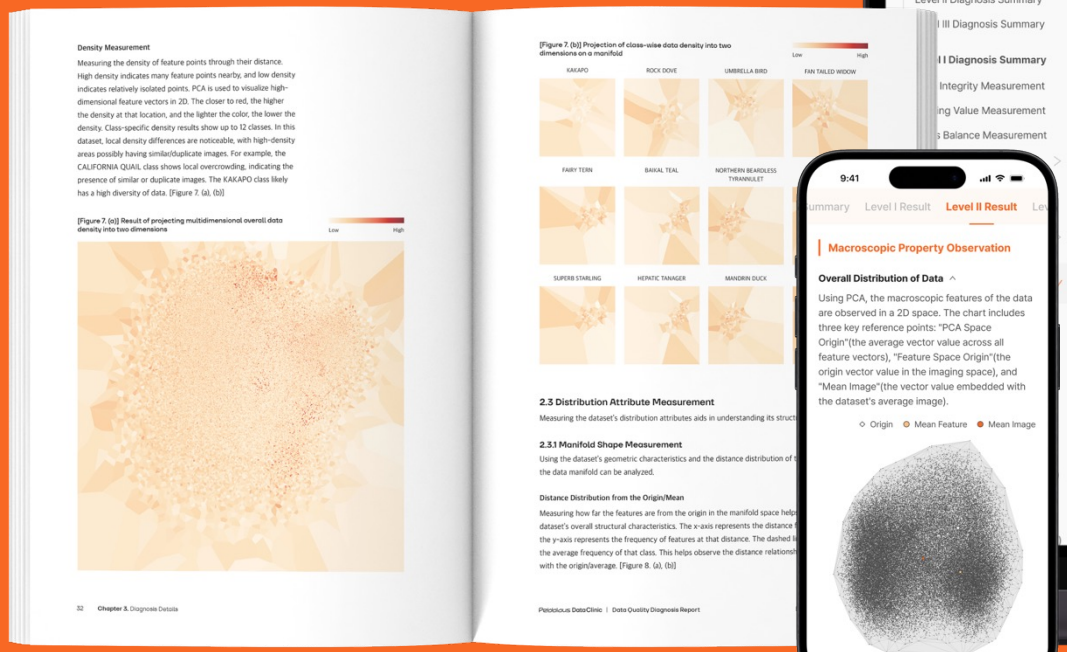
Pebblous

All-in-One SaaS Solution for **AI Data Quality Assessment and Quality Improvement**

web version



pdf version



mobile version

Product 01

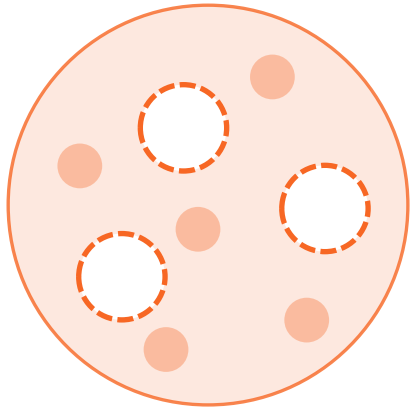
DataClinic

Pebblous

Low Density Area | Redundancy Area | Real Data | Synthetic Data

Data Replica

Creates synthetic data similar to the original while protecting content.

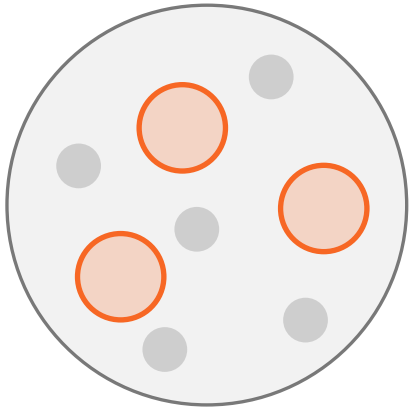


Benefit: Prevents personal information leaks.

Example: Data anonymization for privacy.

Data Bulk-Up

Adds synthetic data to fill gaps in the dataset.

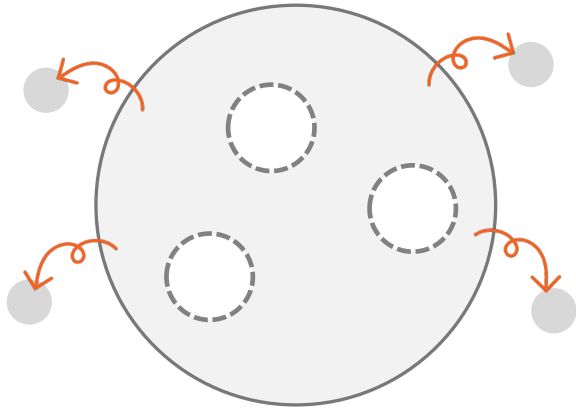


Benefit: Cuts data collection costs and boosts AI performance.

Example: Generating edge case data.

Data Diet

Removes redundant data while keeping model performance intact.



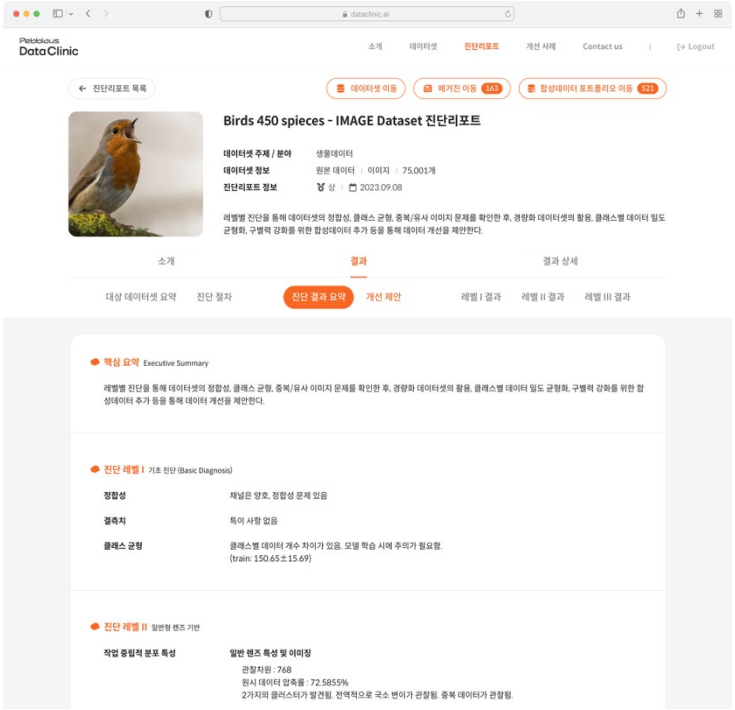
Benefit: Reduces AI costs and improves dev efficiency.

Example: Streamlining training datasets.

Product 01

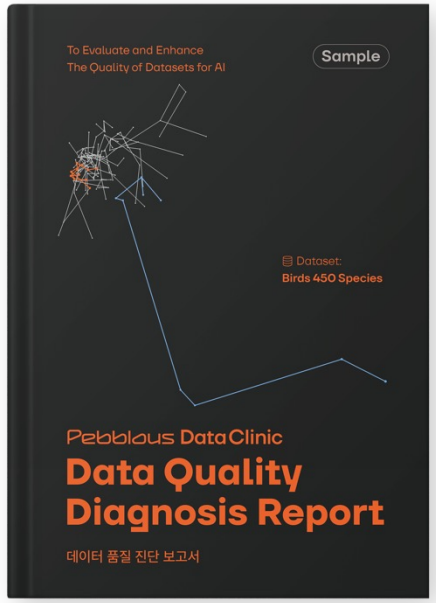
Data Qyality Report

Pebblous



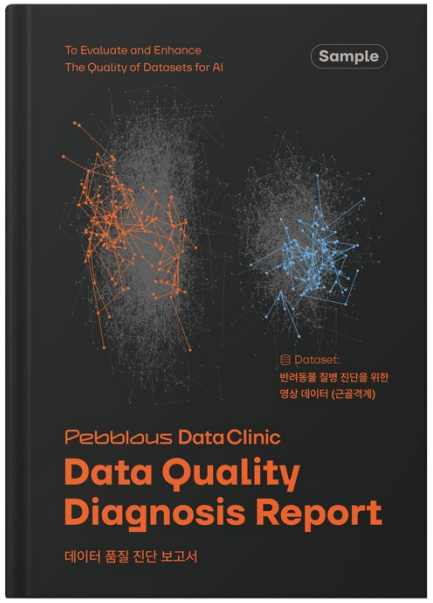
Birds-450 Dataset

Bird 450 Species Dataset (Kaggle)



Pet-bone Dataset

Pet Musculoskeletal X-ray Images (AI Hub)



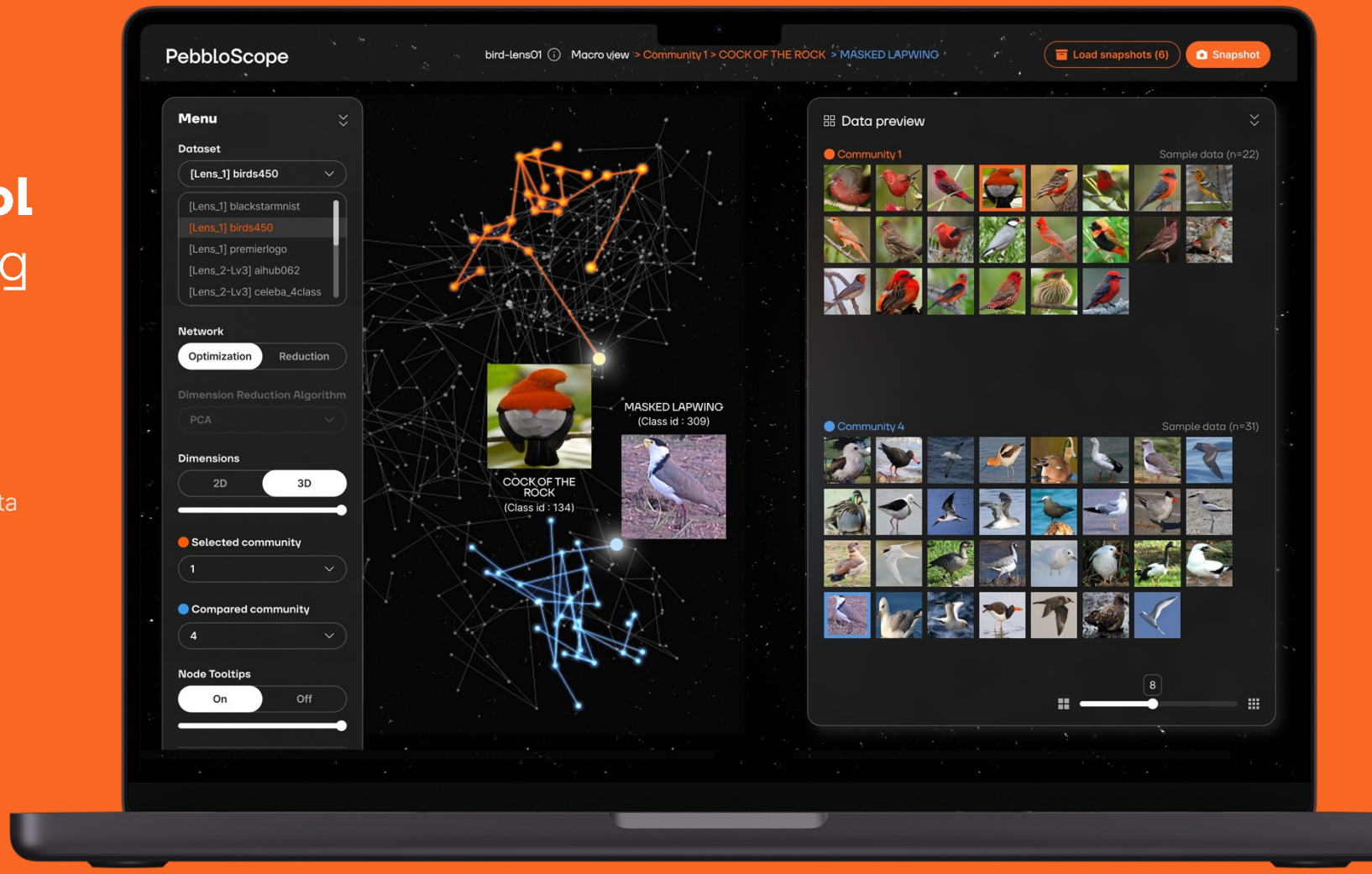
Sample Download
https://bit.ly/진단보고서

Pebblous

PebbloScope

Interactive 3D **Data Communication Tool** for visualization and sharing actionable insights

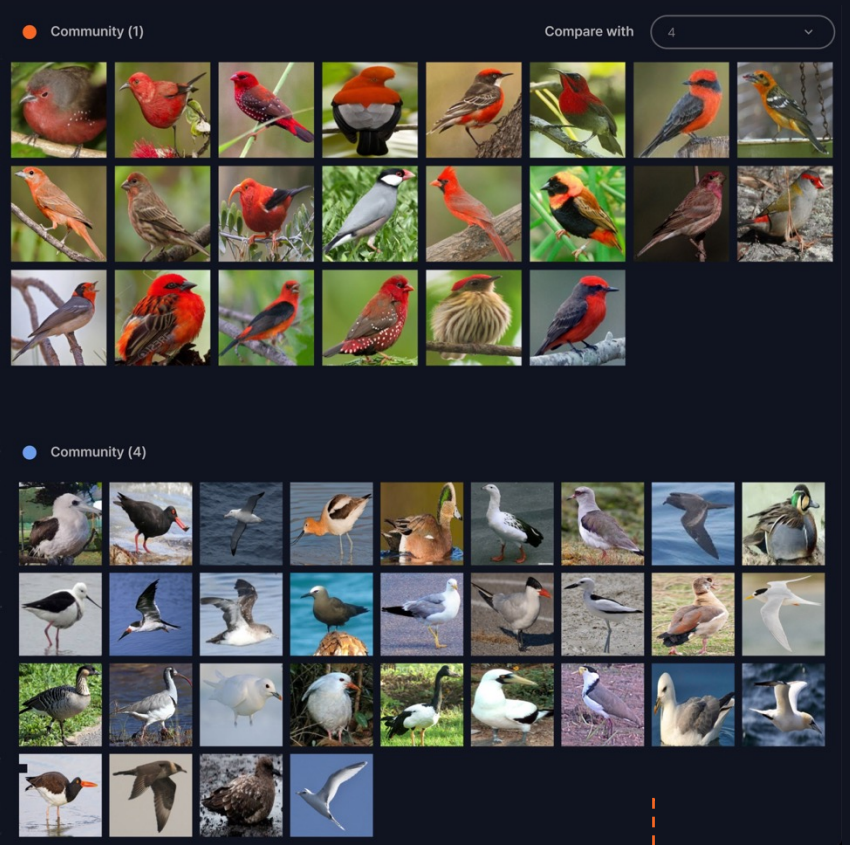
A data communication tool that transforms high-dimensional data into a three-dimensional space, allowing you to interactively explore different attributes and gain insights for data analysis.



Product 02

PebbloScope

Pebblous



Visualization

- Explore 3D data communities optimized through Pebblous Data Lens.
- Identify clusters of similar or redundant data and analyze distribution.

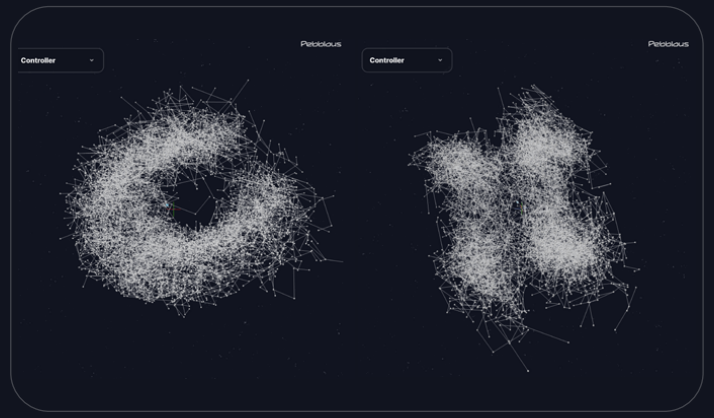
Data Samples

- View and compare original images within data communities to analyze key characteristics.

Demo Video

Pebblous

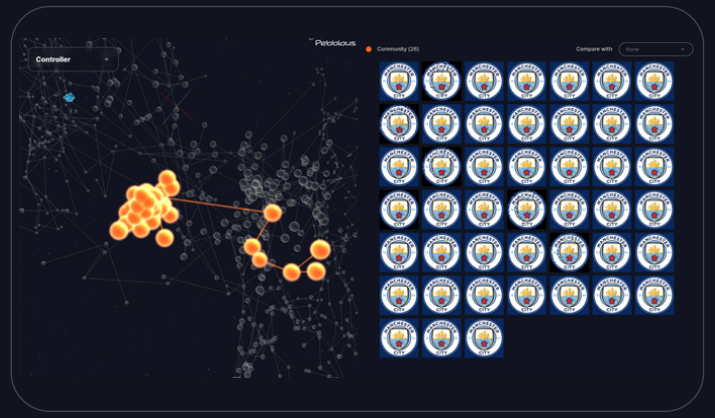
PebbloScope — Key Features



Optimized 3D Data Visualization

- AI training datasets processed through Pebblous DataLens are visualized in 3D.
- Enables observation of unique and distinct distribution characteristics of the data.

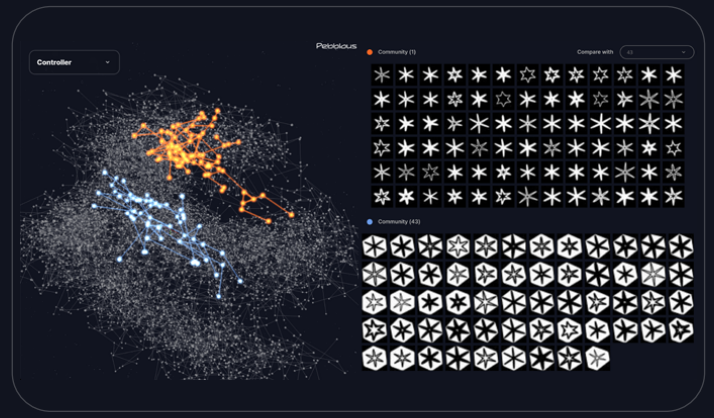
Example: Detecting cluster shapes from different perspectives.



Exploring Data Communities to Identify Issues

- Data communities, based on optimal dimensional distances, allow identification of data redundancy.
- Useful for spotting problematic or overpopulated clusters.

Example: Identifying overpopulated clusters in datasets like logo collections.



Comparison to Analyze Data Characteristics

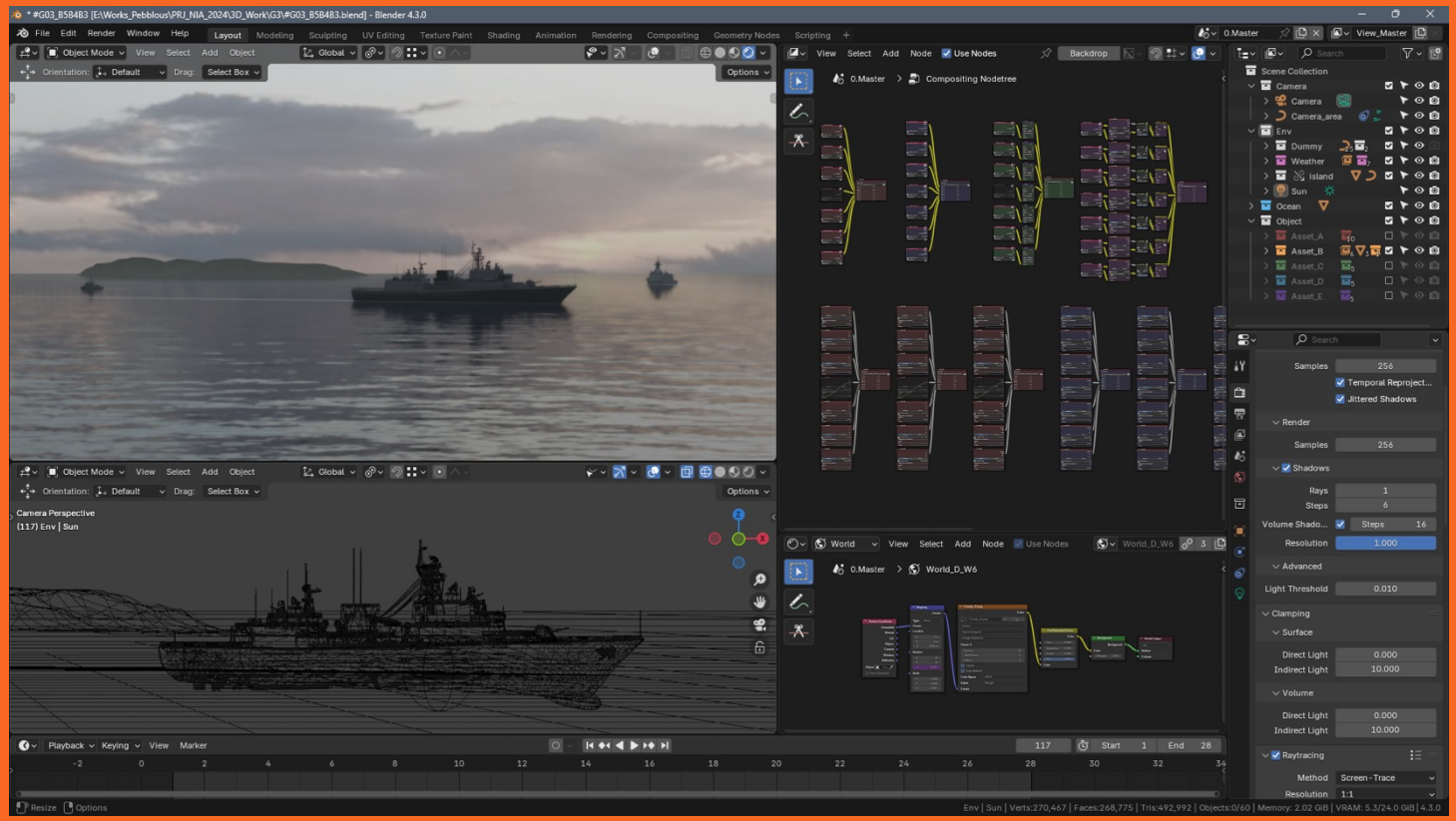
- Compare data communities to highlight distinctive cluster features.
- Provides deeper insights for unique data distributions.

Example: Comparing similar data clusters within the same class.

Pebblous

Synthetic Data

For AI training, synthetic data is generated when:
(1) Data quantity is insufficient,
(2) Real data cannot be obtained, or
(3) Data from diverse environments is required.



Product 03

Pebblous

Synthetic Data Portfolio

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

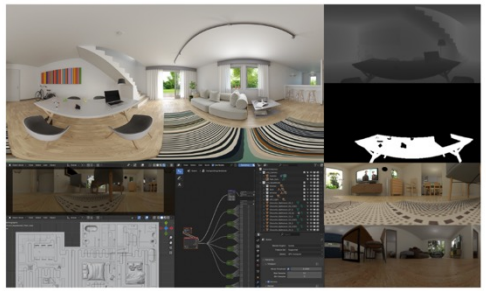
Single image 3D creation Download sample images

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.

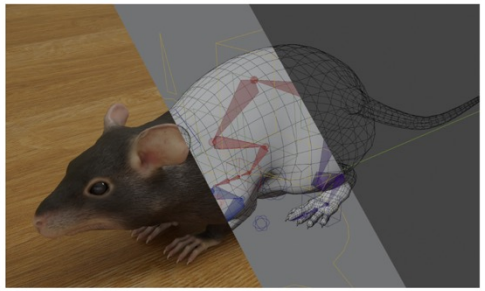
Diet monitoring Download sample images

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

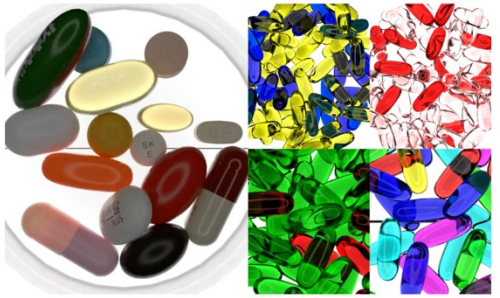


Synthesis data for AI learning to classify mouse behavior and recognize body parts

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

1. **Beyond Data Quality:** From Actionable Insights to AI-Readiness

2. **Immediate ROI:** AI Model Performance and Development Efficiency

01

Rapid Diagnostics

1 Hour Diagnosis

100K Data Samples

For 256 pixel image dataset

02

AI Performance

5% Synthesis

2% Model Boost

Data Bulk-up: Precision-targeting synthetic data

03

Operation Efficiency

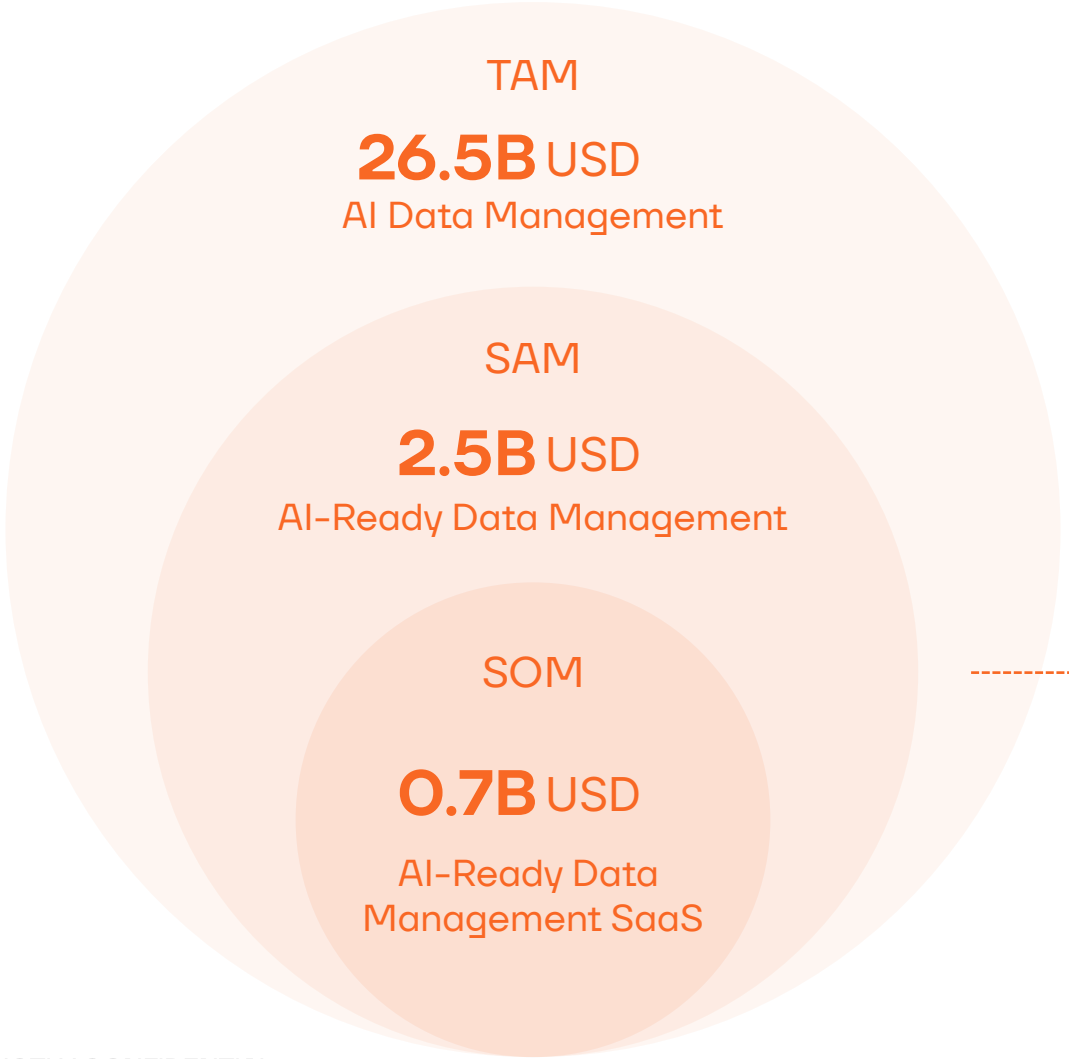
90% Data Reduction

5x GPU Utilization

Data Diet for optimizing volume by removing redundancy.

Pebblous

Data Management Market for **AI-Ready Data**



Market Share of TAM
1%

Estimated Revenue
0.2B USD

Pebblous

Application **Domains**

- Mobility
- Sports
- Metaverse
- Recycling
- Finance
- Fashion
- Logistics
- Military
- Pharma / Medical

AI Data Doctor

DataClinic

- Data Imaging
- Quality Diagnosis
- Prescription
- Improvement

Pebblous

Competitor Comparison — data management, synthetic data, and visualization

Company	Location	Fund Raising	Revenue '23	Key Functions	Key Differentiator
Anomalo	US	88M USD (Series B, '24.11)	.	Recognized as a company specializing in data quality diagnostics in the 2025 Gartner Report.	Pebblous provides data quality diagnostics for unstructured data, complemented by PebbloScope visualization.
Shelf.io	US	60.7M USD (Series B, '21.8)	.	Recognized as a company specializing in data quality diagnostics in the 2025 Gartner Report.	Pebblous generates synthetic data based on quality diagnostics and provides unique visualization support through PebbloScope.
SuperbAI	KR	35M USD (Series C, '24.9)	.	A SaaS-based DataOps platform, recently launched a synthetic data service.	Unlike manual sourcing, SelectStar focuses on automated DataOps, recently targeting AI-Ready data evaluation as its core business.
Tonic.ai	US	46.2M USD (Series B, '21.11)	11.5M USD	Synthetic data generation for personal information protection	Pebblous provides synthetic data generation for unstructured data based on quality diagnostics.
Virtualitics	US	30.8M USD (Series C, '23.8)	18.2M USD	Data analysis and visualization powered by AI	Pebblous offers detailed quality diagnostics for AI training datasets, with both installed and web-based versions of PebbloScope.

Traction

Gartner Report

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

The shortage of GenAI-ready data is cited as one of the main reasons behind the failure of generative AI deployments.
Product managers of data management software vendors must enhance their capabilities to handle unstructured data—through in-house development, integration, or partnerships—to effectively support customers in implementing generative AI.

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

Gartner Report

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

Pebblous provides Data Clinic services alongside its synthetic data solutions, performing precise quality diagnostics to generate synthetic data tailored for a wide range of computer vision learning applications.

The Korean startup Pebblous extends such simulations to encompass human and animal behavior, offering a unique perspective on interaction by generating precision-targeted synthetic data grounded in data quality diagnostics.

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

Gartner Report

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

Synthetic data is emerging as a core infrastructure that enables privacy protection and regulatory compliance, while facilitating AI training and collaboration without relying on real data.

It is rapidly spreading across sectors such as public services and healthcare, and Gartner projects that synthetic data will replace real data by 2030. CIOs must focus on quality validation and strategic technology partnerships to foster an ecosystem of innovation.

Pebblous

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

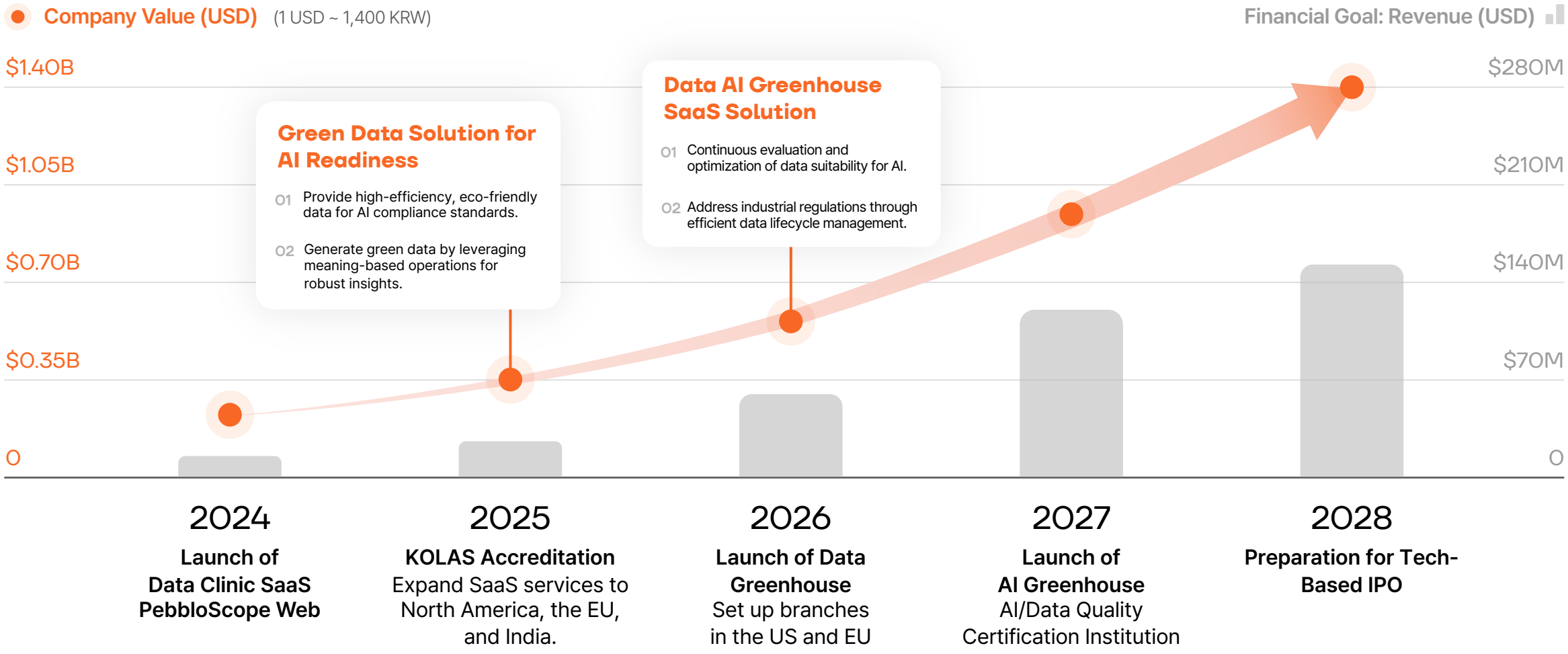
Gartner

Pricing Policy

	Free Plan	Basic Plan	Pro Plan	Enterprise Plan
	Free trial for basic features of Data Clinic and PebbloScope.	Access to Data Clinic results on various public datasets. PebbloScope snapshots	Level II Diagnosis and PebbloScope Creation on Customer's Private Data.	Pro Plan + Level III Diagnosis, Data Diet and Bulk-Up Custom Projects.
Price	**Free**	**$10 / Month ($100 / Year)**	**$500 / Month ($5,000 / Year)**	**$5,000 / Month ($50,000 / Year)**
Data Clinic Web	Public Data	Public Data	+ Private Data	+ Private Data
PebbloScope Web	Public Data	Public Data	+ Private Data	+ Private Data
Max Data Volume		Up to 100K images	Up to 1M images	Up to 1M images
Interactive Visualization		Available	Available	+ Advanced Features
PebbloScope Snapshots		Available	Available	Available
Synthetic Data Testing		Available	Available	Available
Result Download			Available	Available
PebbloScope Creation			Available	Available
Level II Diagnostics			Available	Available
Level III Diagnostics				Available
Data Diet / Bulk-up				Available
On-Prem				Available
Custom Projects				Consulting / Custom Dev

Pebblous

SaaS based Revenue and Beyond



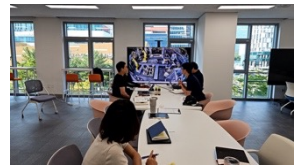
Hyundai Motor Company Project

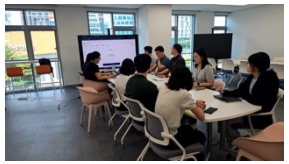
Project Title	Year	Technology	Solution and Outcomes
Synthetic Data for Driving Path Estimation Models	2022	Synthetic Data Deep Learning	Since collecting actual driving paths was extremely challenging, synthetic data was generated to train the model and develop a deployable algorithm.
Tire Wear Prediction Model	2022	Experimental Design & Model Training	Extracted key features from time-series vehicle sensor data and developed a tire wear prediction algorithm.
Synthetic Data for PBV Child Transport Vehicles	2023	CG & GenAI Synthetic Data	Generated synthetic data using both computer graphics and generative AI to simulate various postures and appearances of children captured by wide-angle cameras installed inside vehicles.
Paint Process Optimization	2025	Reinforcement Learning	Developed a reinforcement learning algorithm that achieved a 25% performance improvement over the numerical simulation model within the simulator environment.
Assembly Process Optimization	2025	Reinforcement Learning	Implemented a complex assembly process within the simulator and developed a reinforcement learning algorithm to optimize vehicle model changeover rates.

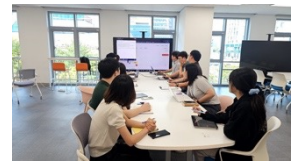
Pebblous

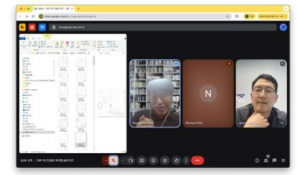
June 2025 — Data Quality Assessment and Improvement Project

Pebblous carried out **data quality improvement projects** for 30 companies









Conducted data quality diagnostics and improvement for 30 companies in the Daegu region, analyzing datasets across diverse domains including manufacturing, healthcare, AI development, and robotics.

Pebblous

August 2025 — **Selected for the Global Big Tech Development Program**

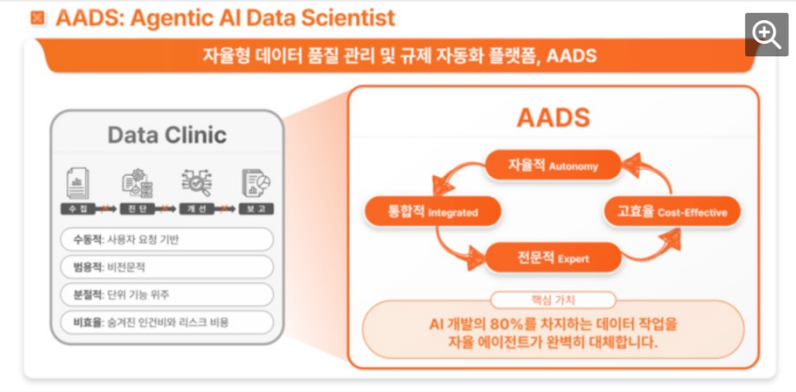
전자신문 etnews

In collaboration with **KISTI**,

developing the **Agentic AI Data Scientist platform**,

a project with a **total budget of 6.1B KRW**,

aimed at **data quality compliance** and **global market expansion**.



페블러스 데이터 클리닉에서 진화한 AADS

AADS는 데이터 관리 모든 단계 엔드투엔드 자동화된 시스템이다.

Led by Pebblous in consortium with KISTI, the company secured a large-scale national project and is currently developing an AI agent system.

Pebblous

Key Customer Cases

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

SUNG KYUN KWAN UNIVERSITY(SKKU)

impact station

Pebblous

Key Customer Cases

MYEONG MOON BIO & DENTAL CO., LTD.

InTheTech

Syntherixs

KIECOTECH

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

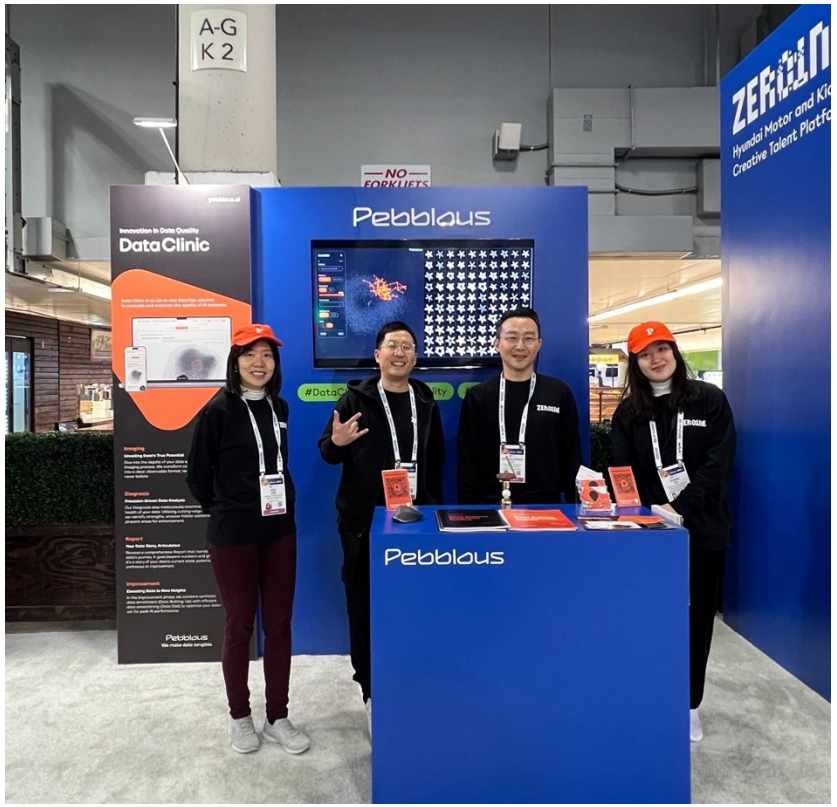
생각의 탄생

Pebblous

CES 2024

Eureka Pavilion Booth

Exhibition — Data Clinic



Pebblous

CES 2025

Eureka Pavilion Booth

Exhibition — PebbloScope

MWC 2025

Data Clinic Booth Exhibition

Pitching as a Leading Korean AI Startup

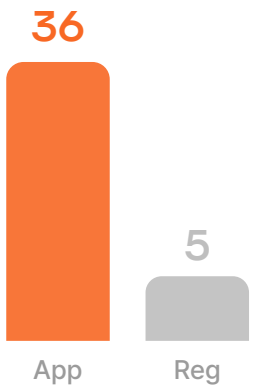
Pebblous

Global IP Strategy for Business Expansion

Patents

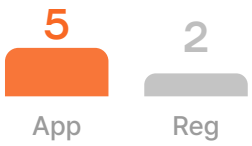
Domestic

Applications: 36
Registrations: 5



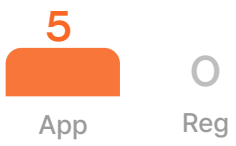
US

Applications: 5
Registrations: 2



PCT

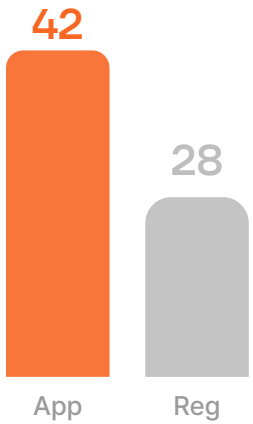
Applications: 5
Registrations: 0



Trademarks

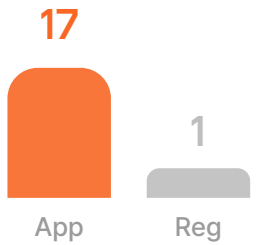
Domestic

Applications: 42
Registrations: 28



US

Applications: 17
Registrations: 1



Team

Total 50 Years of R&D Experiences
in Data and AI filed

CEO | Co-Founder

Joo-Haeng Lee, PhD

Profile

PhD, Computer Science, POSTECH

Principal Researcher in CG & AI at ETRI for 20+ years

Adjunct Prof. Taejae Univ. DS&AI

Code Painting Artist: 15+ Exhibitions

COO | Co-Founder

Jeongwon Lee, PhD

Profile

PhD, Bio and Brain Eng., KAIST

BS & MS, EE, Seoul National Univ.

Principal Researcher in AI at ETRI for 20+ years

Pebblous

Pebblous Team

BD & Marketing	2
(US) Former Google B2B Marketer (Korea) BizDev, UC Berkeley	
AI Engineer	4
POSTECH PhD in Computer Sci. SNU PhD Candidate KAIST, Industrial Engineering	
CG Engineer	1
Developer	4
UI/UX Designer	3
Back Office	3

Pebblous

Investment Attraction and Fund Utilization Plan

Total Investment

USD 2.38M

KRW 3.1B

Last Round
Oct. 2022

Additional Fundraising Plan

USD 1.54M + α

KRW 2.0B

Plan #1

Developing global business and marketing

Plan #2

Enhancing customer support and service operations

Plan #3

Expanding SaaS development and operational infrastructure

Plan #4

Data/AI quality certification services
(KOLAS certification)

Fabulous Data With

Pebblous

Better Data Makes Better AI

Pebblous.ai