

AI for AI. Data Centric을 위한 AI 기술 적용 사례

Corporate Relations & Co-founder

Hyundong Lee (hdlee@superb-ai.com / 010-3500-1188)

2022-08-24



Contents

- 01. 회사 소개
- 02. ML Ops
- 03. AI 기술 사례
- 04. Use-Case
- 05. Value

1. Superb AI 회사 소개

Superb AI는 누구나 쉽게 AI를 개발할 수 있는 “AI의 민주화”를 지향합니다.

“Democratize AI to empower humans everywhere in the world”

모든 단계의 AI 개발 조직을 지원해 AI 기술이 널리 보급될 수 있도록 기여하는 것

특장점

● SaaS 플랫폼

- AI 기반 자동화를 통해 더 빠른 데이터셋 구축 가능
- 비개발자 및 전문 개발자를 모두 고려한 UI 및 쉬운 사용성
- 기존 AI 개발용 서비스들과 쉽게 연동 및 확장 가능한 오픈 플랫폼

● 데이터 컨설팅/구축 서비스

- 고객 니즈별 다양한 형태의 서비스를 유연하게 제공 가능
- 고품질 데이터 구축을 위한 컨설팅 및 전담 인력 배치 가능
- SaaS 플랫폼 활용으로 데이터 보안 및 엄격한 프로젝트 관리 가능

글로벌 일류 기업 및 혁신 기술 기업은 **Superb AI**와 함께 성장하고 있습니다.

설립 첫 해부터 올해까지 **매년 2배 이상 성장 중** (*매출기준, YoY)

글로벌 일류 기업 및 혁신 기술 기업들이 **Superb AI**의 플랫폼과 서비스를 통해 **개발 비용/시간 절감 중**

“Suite를 사용한 후 **데이터 구축 비용이 56%이상** 줄었습니다.”
미국 AR/VR 기업 A, CEO

“Suite의 자동 라벨링 AI는 객체 탐지 정확도와 속도가 탁월합니다. **이미지 당 라벨링 소요 시간이 2초도 안걸려** 데이터 구축 시간을 대폭 감소시킬 수 있었습니다.”
일본 철강 생산 기업 B, ML 엔지니어

데이터 구축 품질이 다른 기업의 결과물에 비해 월등히 높은 수치입니다.
99.9%, ... 전체 중, 최상위 레벨로 생각합니다.

국내 데이터 품질 담당 기관 C, 수석

“Suite 덕분에 불필요한 개발 업무가 대폭 줄어들어, **2개월 만에 PoC 단계의 제품을 완성**할 수 있었어요.”

미국 자율주행 기업 D, ML 엔지니어

이미지 양품/불량품 소량으로 AI를 만들었을 때, 98.6% 성능이 나왔습니다. 자체적으로 **training data**를 좀 더 잘 만들면 개선될 것 같습니다.

스마트팩토리 도입기업 E, 총괄 부장

2. ML Ops

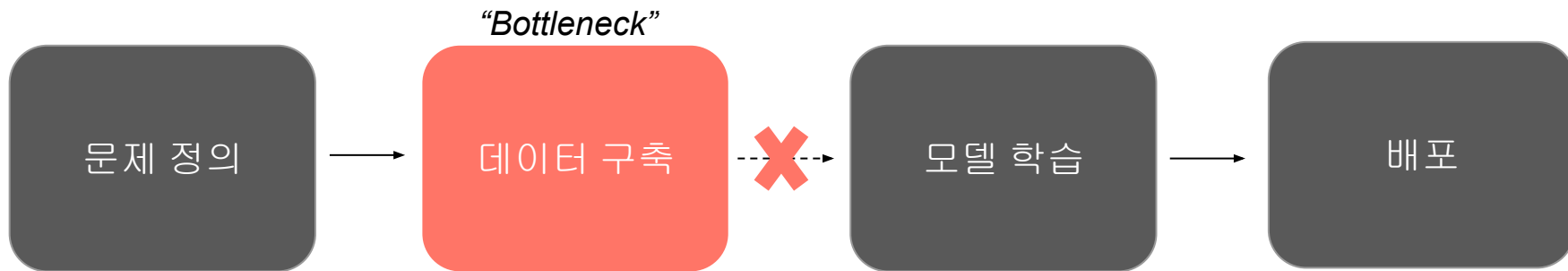
Automate Data Preparation. Deliver Better Datasets.

Superb AI는 컴퓨터 비전 학습 데이터 세트의
저장, 라벨링, 관리/분석, 그리고 큐레이션을 손쉽게 할 수 있도록 지원하여
빠르고 일관된 데이터셋 준비 작업을 현실로 만들어주고 있습니다.

데이터를 기반으로 빠르게 실험하고
AI 서비스를 빠르게 성장/ 확장 시키는 것

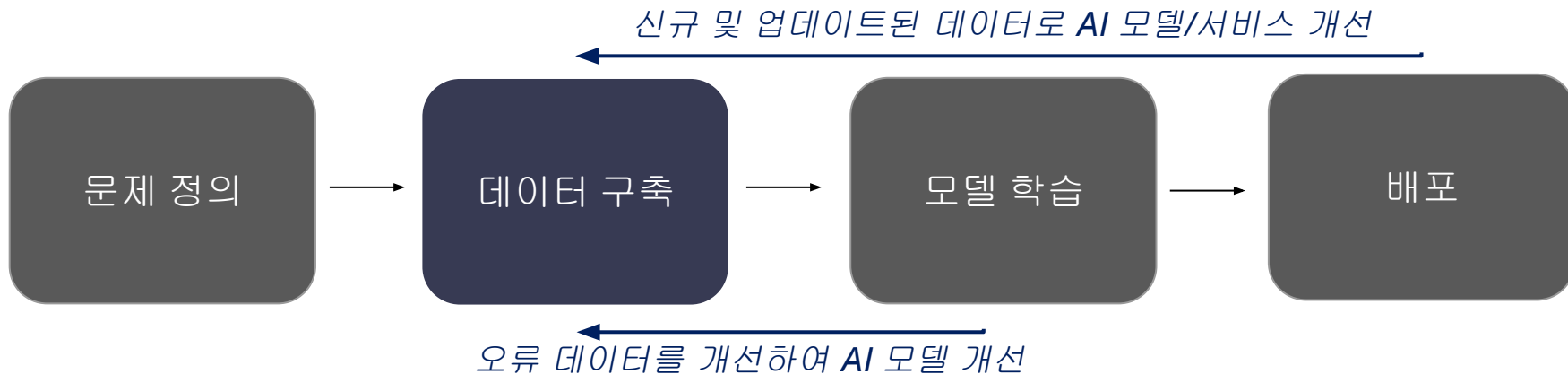
Problem in AI Development Cycle

- AI 개발 프로젝트의 **80%** 이상이 배포까지 가는 데 실패합니다.
수 많은 팀들이 데이터를 구축하는 단계에서 어마어마한 시간과 비용을 쓰게 됩니다.



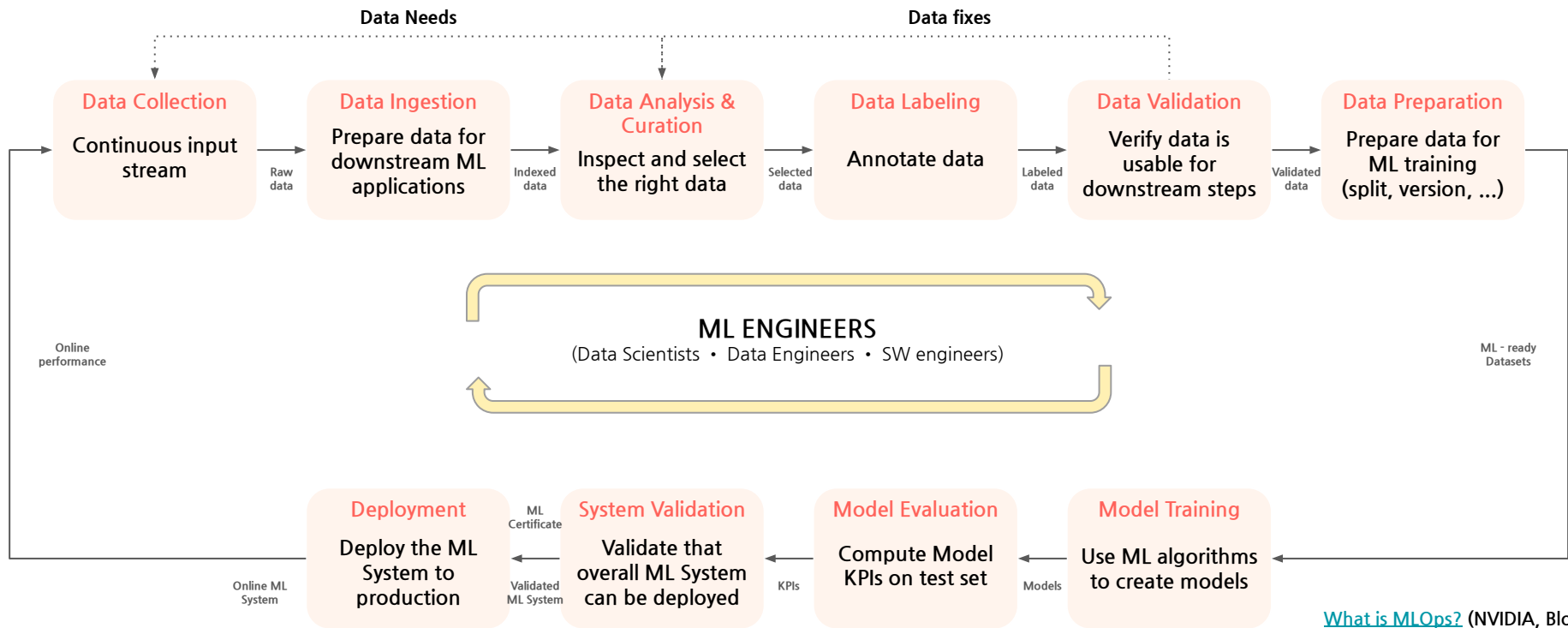
Our Solution

- 우리는 고객들이 컴퓨터 비전 어플리케이션을 지능형 자동화로 더 빠르게 구축하고 배포할 수 있도록 돕고 있습니다. 우리의 목표는 **Data Preparation** 워크플로우에서 장애물을 제거하여 데이터를 구축하고, 체계적이며 반복적으로 데이터를 개선시켜 완성도 높은 **AI** 어플리케이션 개발에 기여하는 것 입니다.



MLOps Pipeline

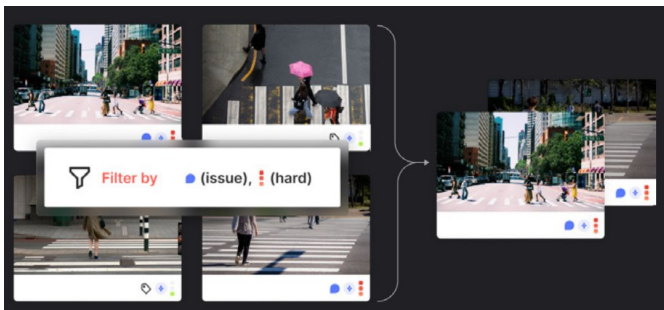
● AI서비스개발의 Lifecycle



[What is MLOps?](#) (NVIDIA, Blog)

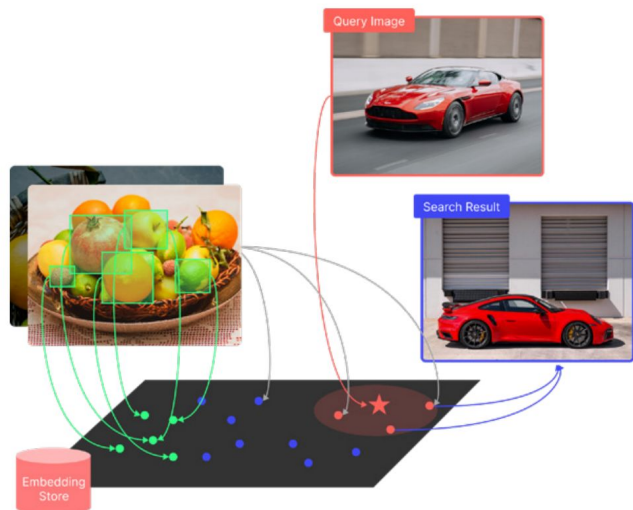
2. 세부 AI 기술사례

Curation - 원하는 데이터를 쉽고 간편하게 분류



- 원하는 데이터만 손쉽게 찾고 분류 가능
 - 필터/태그/검색 기능을 통해 손쉬운 분류/검색 가능
- 팀원들과 함께 데이터셋의 버전 관리
 - Train/Test, Batch별 데이터셋 등 원하는 단위에 따라 버전 관리 기능

Curation AI (Feature Store) - 데이터셋의 특징을 추출, 분석하여 데이터셋을 보완



- 데이터셋에서 부족한 부분을 찾기 (**Edge Curation**)

- 데이터셋의 특성을 분석하여, **Edge case**를 자동으로 찾아내고 이를 보완

- **Test Set Curation**

- 데이터들의 대표적인 이미지들을 자동으로 분석하여 **Test Set** 제안

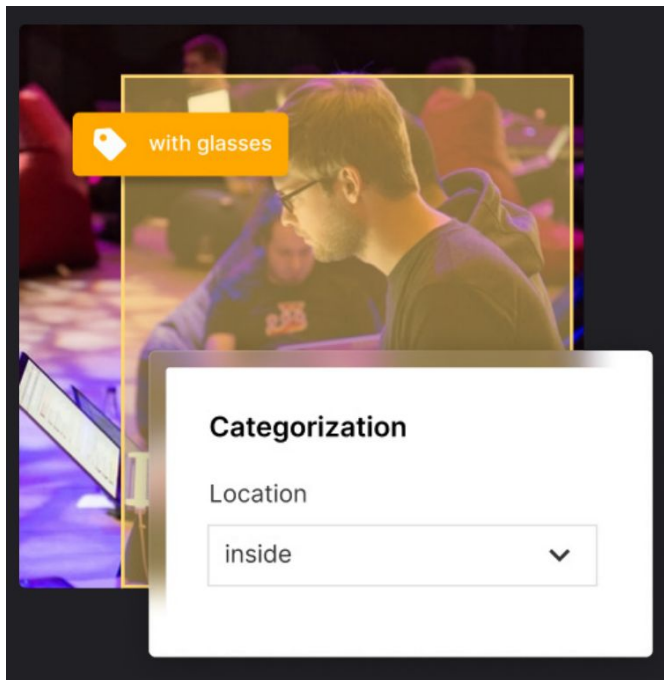
이미지 검색 기능 - 특정 데이터의 **Feature**와 유사한 특성의 데이터 검색 (**Natural Language / Image Query**)

The screenshot displays the Superb AI web interface. On the left is a dark sidebar with navigation links: Tenant, Projects, Data, Users, Integrations, Find Similar (highlighted), and Settings. The main content area is titled 'Result' and shows an 'Applied Query' section with 'Text query' containing tags 'Sunny' and 'Ship'. Below this, an 'Image query (2/10)' section displays two images of sailboats. At the bottom, a grid of 'Similar images from AAA Project (123 items)' is shown, with the second image in the first row highlighted with a red border and a checkmark. On the right side of the interface, a floating menu is visible with options: '1 Selected', 'Edit Tags', 'Send to Project', 'Select All', 'Deselect All', and a list of projects: 'Cat vs Dog', 'Transportation V1', 'Transportation V2', and 'Transportation V3'.

- 유사한 데이터 검색

- 유사한 특성을 가진 데이터를 검색
- **Text** 검색 / **Image** 검색 제공

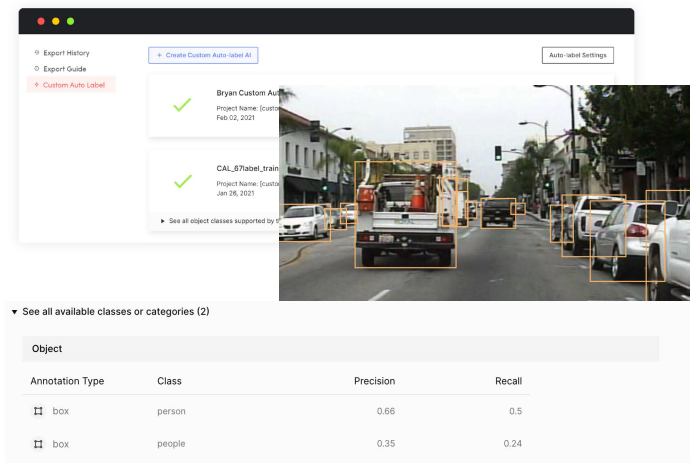
Labeling - 필요할 때마다, 언제든지 가능한 데이터셋 가공



- 커스터마이징이 필요없는 강력한 라벨링 도구 지원
 - Image/Video/3D Point Cloud 데이터 지원
 - Bounding Box, Segmentation, Keypoint, Polyline, Classification 등 다양한 Annotation 도구 지원
- 라벨링 필요 시 라벨링/검수 전담 인력 제공
 - 인력이 필요한 프로젝트/데이터에 한하여 고객 전담 Workforce 제공
 - 훈련된 전문 인력이 직접 프로젝트 컨설팅 및 라벨링/검수 대행

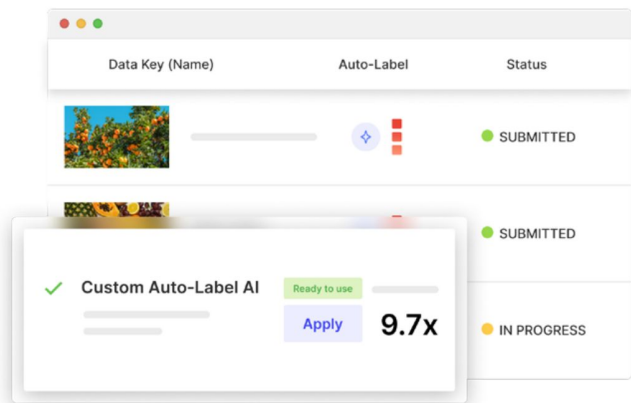
Auto Label - Superb AI 가 제공하는 110종의 기본 AI 기능

- Superb AI 자체 모델 활용 (*관련 미국 특허 11종 등록)
- Pre-trained model (Auto Labeling) : 사람, 차량 등 common object 110여 종에 대한 추가 학습 없는 자동 라벨링 가능



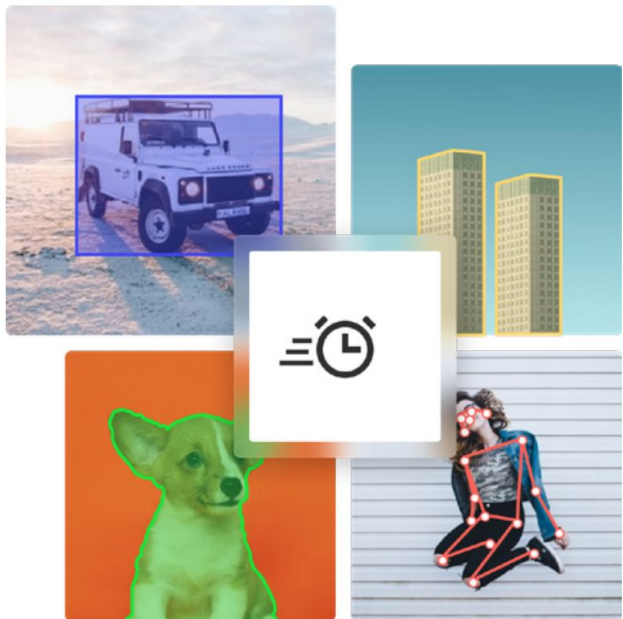
[Blog 링크](#)

Custom Auto Label - 데이터셋에 특화된 모델로, 라벨링 속도 향상



- 코딩 없이 오토라벨용 모델을 손쉽게 제작 가능
 - 코딩 없이 몇 번의 클릭만으로 지정된 GT 모델 학습 및 생성, 적용 가능
- 특수/도메인 데이터셋도 적용 가능
 - Bounding Box, Polygon Segmentation, Keypoint, Classification 등 다양한 모델 지원
- **Transfer Learning/ Few shot Learning** 기반으로 소규모 데이터셋으로 빠르게 고성능 AI 생성
 - 500 Label 기준, 30분 이내 학습으로 고성능 AI 확보 가능

Model Assisted Labeling 지원 - 자사의 고도화된 모델을 라벨링에 쉽게 활용

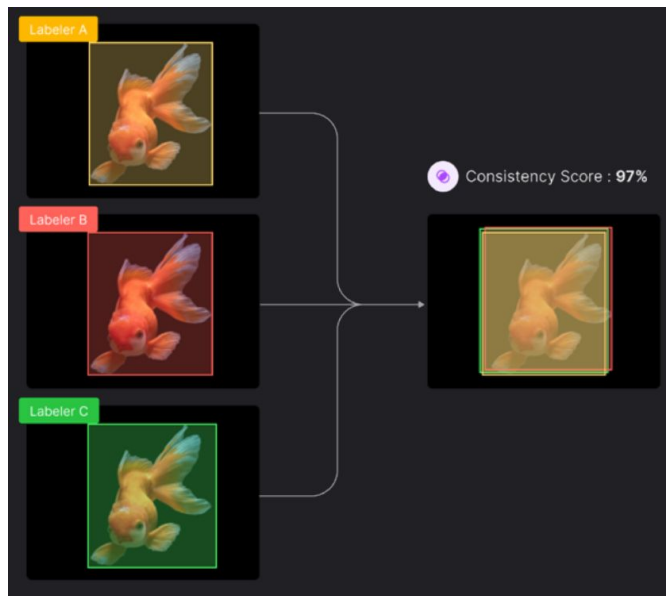


- **고객의 모델 결과값을 쉽게 연동 가능**
 - SDK/CLI를 통해 모델 결과값 혹은 기존 보유한 데이터셋을 업로드 가능
- **라벨 시각화를 통해 즉시 개선**
 - 수정이 필요한 라벨을 **Annotation Tool**에서 시각화하여 빠르게 확인하고 개선
- **Class의 불균형을 빠르게 확인 가능**
 - 데이터 통계 기능을 통한 **class**의 불균형 확인

Manual Review

	Labeling Guideline	Consensus Labeling	Tag / Issue Tracking	Review Mode
사용 시기	라벨링 시작 전 라벨링 진행 중	라벨링 시작 전 라벨링 진행 중	라벨링 진행 중	라벨링 완료 후
목적	라벨링 가이드 제공	라벨링 가이드의 불명확성 확인	데이터 단위의 문제점을 기록/공유	라벨링 된 데이터를 승인/거절 (재작업)
주요 내용	현재 버전의 통일된 가이드라인을 라벨링 프로젝트 메인 화면에 제공 (pdf 업로드 가능)	1개의 데이터를 여러 명의 작업자에게 배정하여 라벨링 진행 후, 작업자 간 라벨링 결과를 비교하여 일치도 점수 (Consistency score) 표시	검수 모드에서 데이터의 문제 (Issue)나 논의가 필요한 내용을 기록하여 이해관계자 간 소통	라벨링 완료된 데이터에 한하여 검수 진행 - Approved/ Rejected 중 택 1 반려된 데이터는 기존 작업자에게 재할당
비고	작업에 참여하는 모든 인원이 동일한 가이드라인을 보며 작업 가능	가이드 개선 작업 또는 라벨러 선정/평가 시 활용 가능	모든 이해관계자가 실시간으로 소통 가능	Manager / Reviewer가 검수 시 사용하는 기능 다수의 Review Round 활용 가능

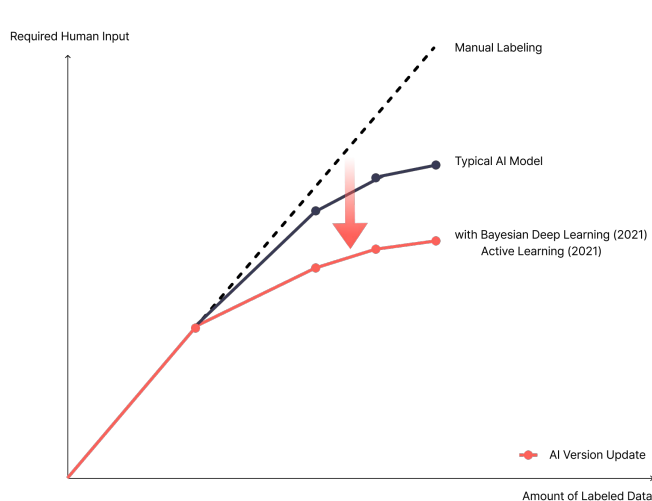
Advanced QA - 실시간 소통을 통한 데이터 품질 개선



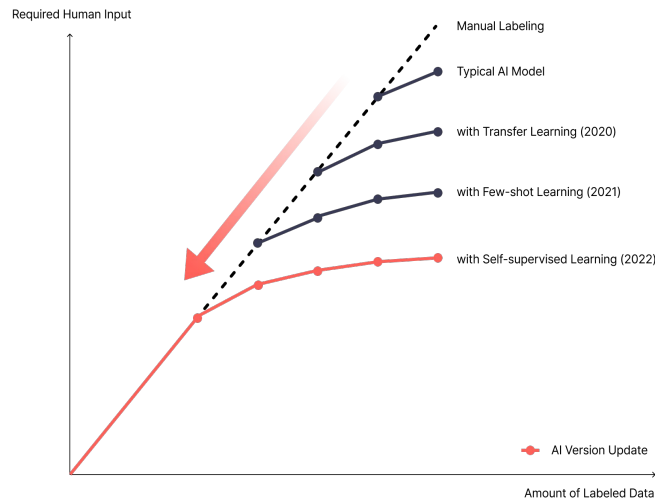
- 컨센서스 라벨링 도구로 라벨의 협업 및 오류를 파악 가능
 - 컨센서스(Consensus) 라벨링의 Consistency Score 계산을 통한 라벨링 오류 사전 파악
- 사용자 간 실시간 소통을 통해 품질을 조율
 - 라벨러/PM/엔지니어 간 실시간 피드백 기능을 통한 데이터 품질 조율
- 전수 검사 프로세스로 라벨링 정확도 증가
 - 라벨링 완료 후 리뷰 프로세스를 통한 품질 검증

Uncertainty Estimation - Auto Label 사용 결과를 Validation AI 가 평가하여 재학습에 필요한 Data 선정

- 라벨링 난이도를 AI가 판별해주어 필요한 라벨만 검수 가능
 - Uncertainty Estimation 기반 라벨 난이도를 판별하여, 검수가 필요한 라벨만 효과적으로 분류

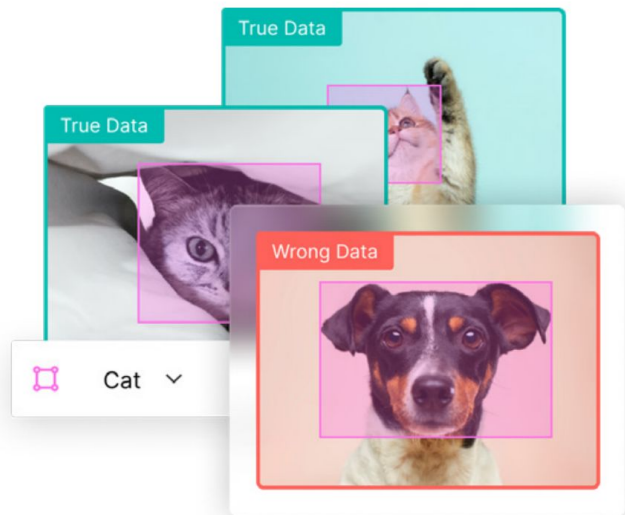


더 적은 데이터로 더 나은 모델을 학습하고..



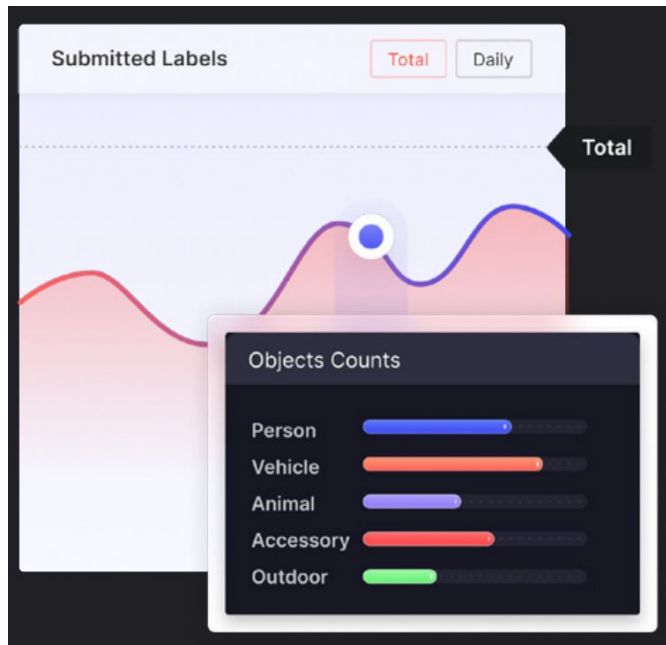
... AI 모델을 더 자주 업그레이드 합니다

라벨 오류 탐지 기능 - AI가 자동으로 라벨링 오류를 분류



- 검수가 완료된 데이터의 품질도 한번 더 체크
 - 작업 완료된 데이터셋 내 품질 불량 데이터를 자동으로 탐지
- **Human Error** 를 잡기위해 구축 **GT**를 기반으로 **AI**를 만들어 데이터를 검수
 - AI를 활용한 데이터셋 검증을 통해 보다 데이터셋 품질을 확보함

Management - 데이터 현황을 한 눈에, 관리는 유연하게



- 데이터 프로젝트 현황을 실시간으로 모니터링 가능
 - 프로젝트 대시보드/통계를 통해 전체 프로젝트 기간 예측 가능
- 데이터 분포도를 통한 편향성을 미리 파악 가능
 - 객체별 분포도를 통한 데이터셋 편향 및 오류 확인
- 사용자 생산성을 추적하고 관리 가능
 - 라벨러/검수자 리포트를 통한 작업 생산성 관리

Custom AI Training / Automatic performance evaluation - 지정된 GT에 특화된 AI 생성 및 성능 평가

- **Transfer Learning/ Few shot Learning** 기반으로 소규모 데이터셋으로 빠르게 고성능 AI 생성
 - 500 Label 기준, 30분 이내 학습으로 고성능 AI 확보 가능

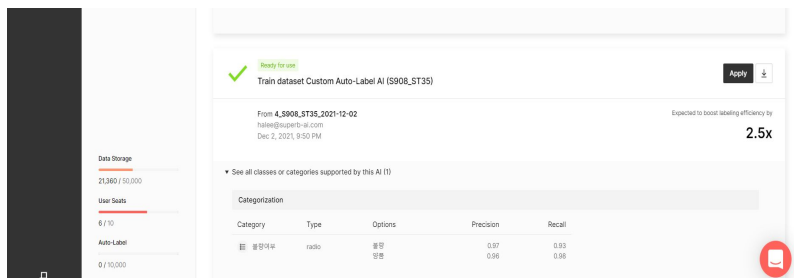
The screenshot shows the 'Custom Auto-Label' interface. It lists three training methods, each with an 'Apply' button highlighted by a red box:

- random add 1520 (490 + 349 + random add 681)**: Expected to boost labeling efficiency by **5.6x**.
- active learning 1520 (s 490 / v 349 / u 681)**: Expected to boost labeling efficiency by **6.6x**.
- base 839 (val 349)**: Expected to boost labeling efficiency by **4.7x**.

Dataset	<u>839</u> (초기GT = 490 + val 349)		<u>1520</u> (Active learning = 490+349+Uncertain 681)		% Change	
F1-score	X 4.7		X 6.6		40.4% (mAP 6.2% + mAR 12.9% ↑)	
Class	Precision	Recall	Precision	Recall	Precision	Recall
푸시업바	0.92	0.88	0.96	0.92	4%	5%
덤벨(아령)	0.88	0.88	0.92	0.92	5%	5%
원판	0.96	0.76	0.96	0.88	0%	16%
원력기	0.95	0.72	1	0.76	5%	6%
악력기	0.75	0.75	0.9	0.92	20%	23%
헬스스트랩	0.84	0.78	0.88	0.92	5%	18%
헬스장갑	0.95	0.65	0.97	0.91	2%	40%
줄넘기	0.96	1	0.96	1	0%	0%
운동밴드	0.8	0.8	0.96	0.92	20%	15%
폼롤러	0.96	0.92	0.96	1	0%	9%
요가링	0.78	0.72	0.96	0.96	23%	33%
요가패드	0.96	0.92	0.96	1	0%	9%
짐볼	0.93	1	0.96	1	3%	0%
체중계	1	0.96	1	1	0%	4%

참고 : 2021년 NIA 데이터 구축사업 - 소형객체-운동기구 Polygon Segmentation (초기GT를 활용한 - Active Learning 기법 / 1cycle)

Custom AI Download / API - 학습된 AI모델을 Suite 외부에서 사용



- **Download 기능을 활용한 Off-line에서의 사용 기능 제공**
 - 인터넷 유무와 상관없는 장소에서 사용하기 위한 AI 다운로드 기능 (스마트팩토리 분야 Reference)
- **API 기능을 활용한 On-line에서의 사용 기능 제공**
 - Suite 내부 + 외부에서도 쉽게 사용할 수 있는 AI 기능

4. Use-case

- **Computer Vision** 전 영역 고객의 문제를 함께 해결하는 파트너로 활동 중
 - Medical, AR/VR, Smart Farm, Retail, Satellite, Robotics, Etc 등
- 아래 4개 **Case**의 경우 **Suite** 를 이용한 이후,
4.32~10.21배 ML Application 개발 효율성 증가 / 공정 자동화 실현

자율주행 개발 A사

Custom AI (Segmentation) 데이터 구축
- 기존 대비 23% 수준으로 시간 절감

홈피트니스 서비스 제공 C사

Custom AI (Key-point) 데이터 구축
- 기존 대비 9% 수준으로 시간 절감

스마트 팩토리 개발 B사

Custom AI (Classification) 를 Off-line 공정에 도입
- Suite의 Auto ML 기술로 공정 자동화 실현

화학 제조 기업 D사

Curation AI / Custom AI (Segmentation) 데이터 구축
- 기존대비 20% 수준으로 시간 절감

5. Value

글로벌 일류 기업 및 혁신 기술 기업은 **Superb AI**와 함께 성장하고자 합니다

Superb AI가 고객에게 전달드리고자 하는 Value

AI Application

명확한 설계

Data Centric 기반 통계/ 분석

시간/ 비용 절감

AI 도입을 위한 장벽 해소

다양한 AI (API) 제공

AI 기술/사업간의 장벽 해소

AI 성능 향상 / 적용영역 확장

체계적이고, 정교한 데이터 관리

감사합니다.

이현동 이사
hdlee@superb-ai.com