

# Smart Labeler

데이터 레이블링 자동화 서비스

---

창업 캡스톤 디자인 TEAM\_4

김경민, 김인제, 박재현, 채현욱

2020. 06. 25

# Index

1. Introduction
2. Problem & Solution
3. Market & Customer
4. Business model
5. Future Work



# Smart Labeler

데이터 라벨링 노가다 는 이제 그만!

Smart Labeler 는 인공지능 학습에 필요한 데이터를 수집부터 라벨링 까지의 모든 과정을 도와주는 서비스



# Problem







Google

Baidu 百度



Microsoft



facebook



인공지능 분야에서, 대기업과 스타트업이 해결하려는 문제의 성격이 다르다



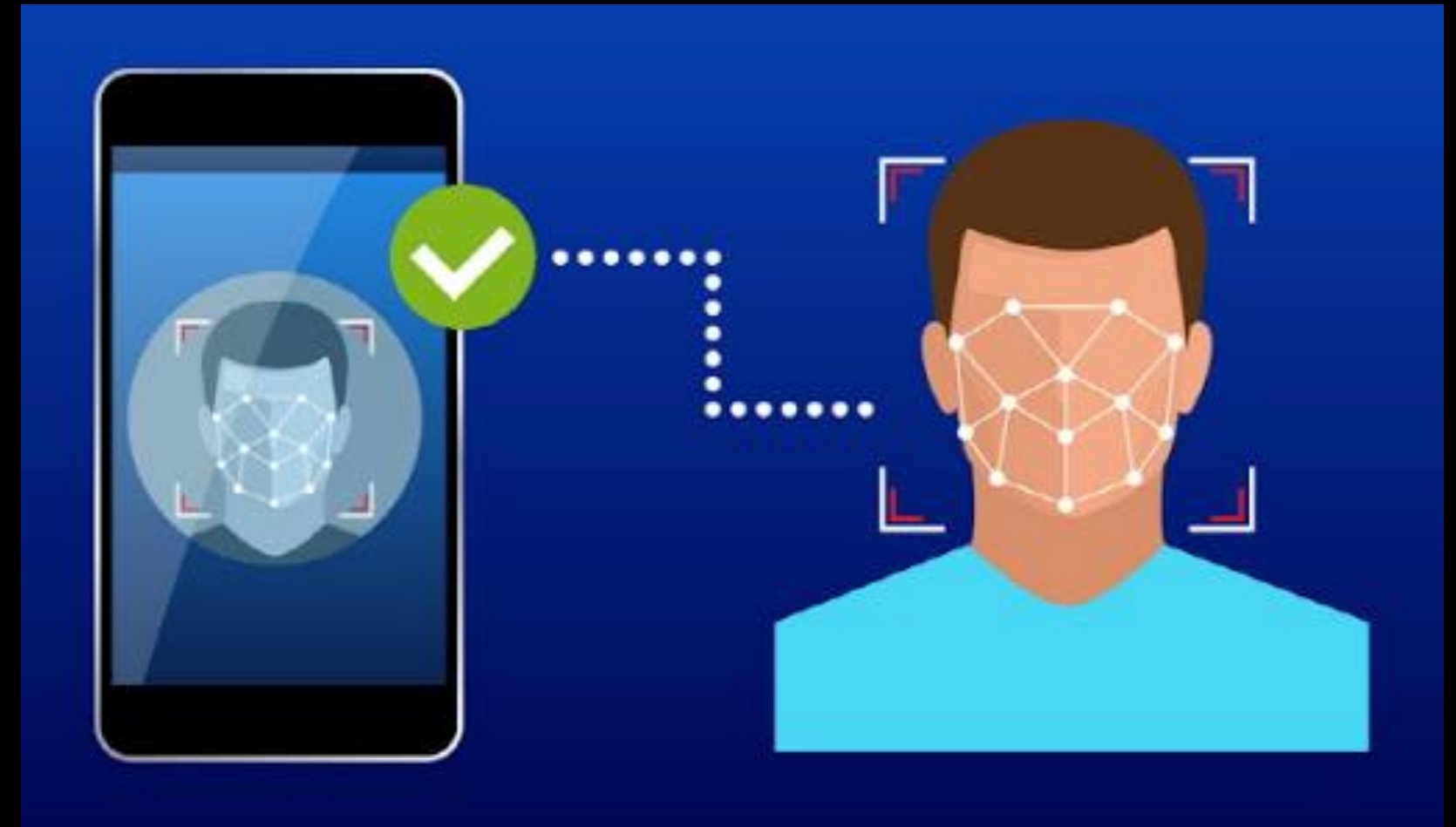
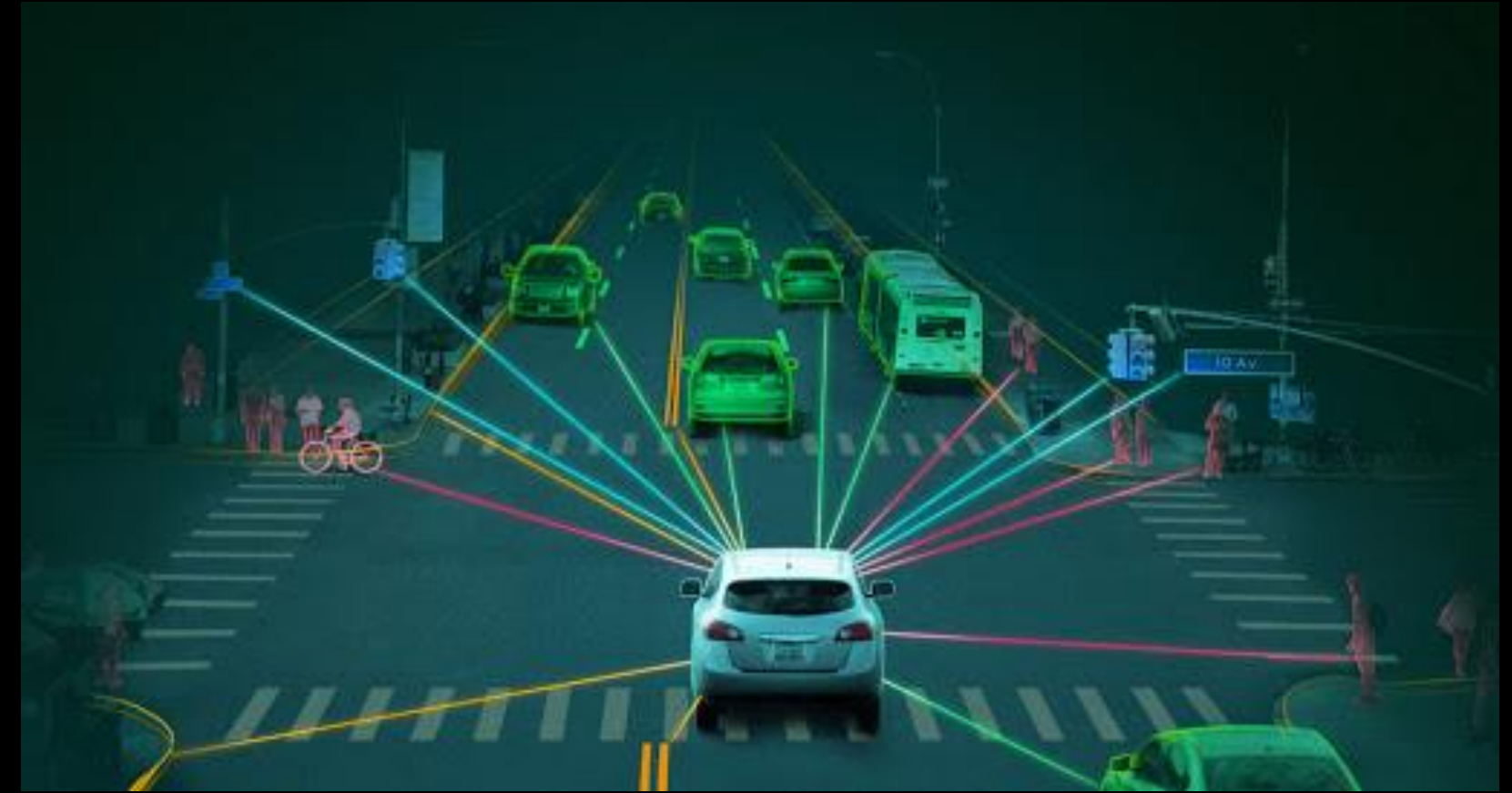
Google

Baidu 百度

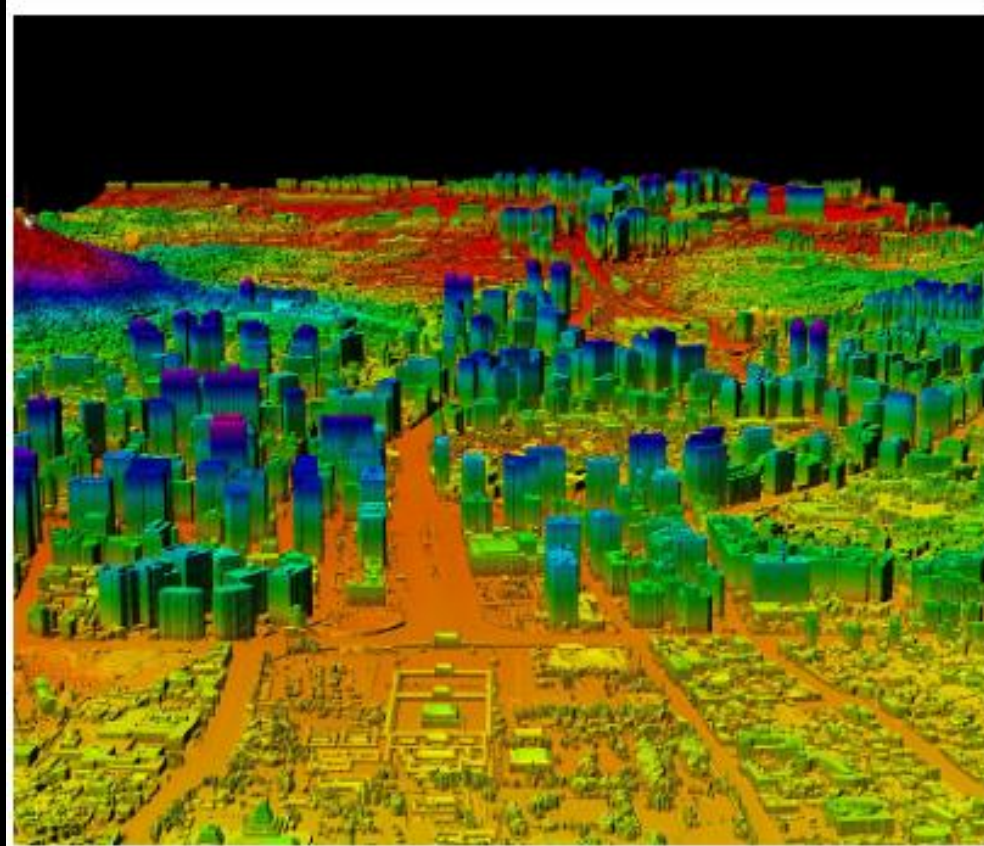
Microsoft



facebook







AI PLATFORM	Skelter Labs	persona	Mondrian AI	Superb AI	WIGO
	ImCloud	SOLIDWARE	acryl	ELGEN	bpu
FUTURE CAR				MINDs Lab	Agile SoDA
DRONE		MOBILTECH			
INTELLIGENT ROBOT					
NEW MATERIAL					
BIO	PDXen	noul	KaiPharm	Infoboss	3billion
MEDICAL	MediRita	NEUROPHET	JLK INSPECTION	ELYSIUM	AI TRICS
HEALTH & WELFARE	BUZZPOLE	Medi Whale	VUNO	baikal.ai	FineSight
HOME APPLIANCE	ICON.AI	SCATTER LAB	42MARU	ThinkWave	cochlear.ai
SMART MANUFACTURING	RAINBOW BRAIN	4IND	LAON PEOPLE	PEC	Wulala Lab
SMART CITY	NETON	CONTINUUM	ScalaWox	Alchera	BLUE SIGNAL
SMART ENERGY	ninewatt	paran	MAKINA ROCKS		
SMART ANIMAL HUSBANDRY & AGRICULTURE					
FIN-TECH	QRAFT	OezSoft	bro	Big tree	QARA
EDU-TECH	Riidi!	Starcell	SimSimi	crevate	JS
REAL ESTATE	TANKER	빅헬류			
LAW/PATENT	LAW TALK	인텔리콘	SYSTRAN	파이특허법률사무소	WERT

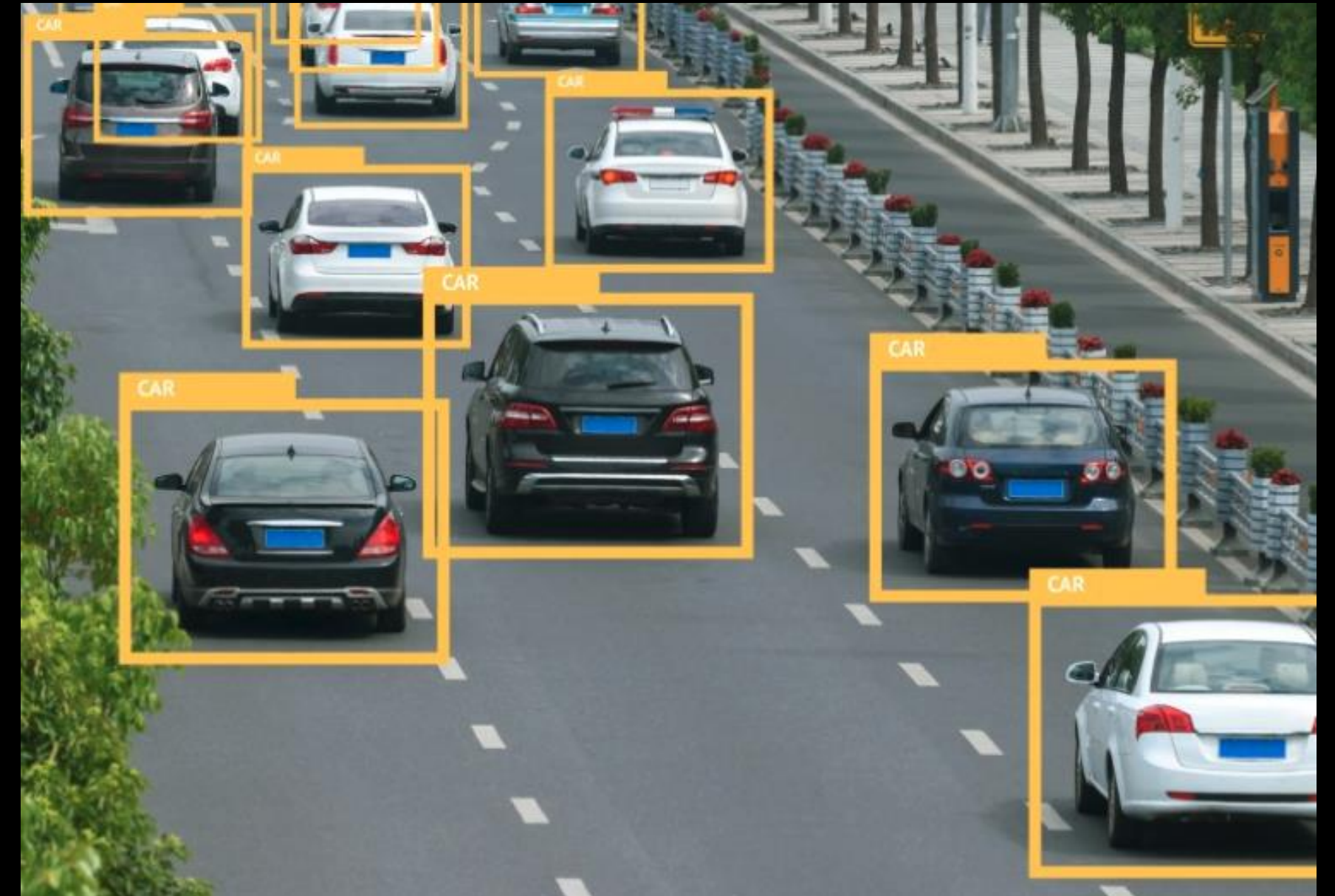
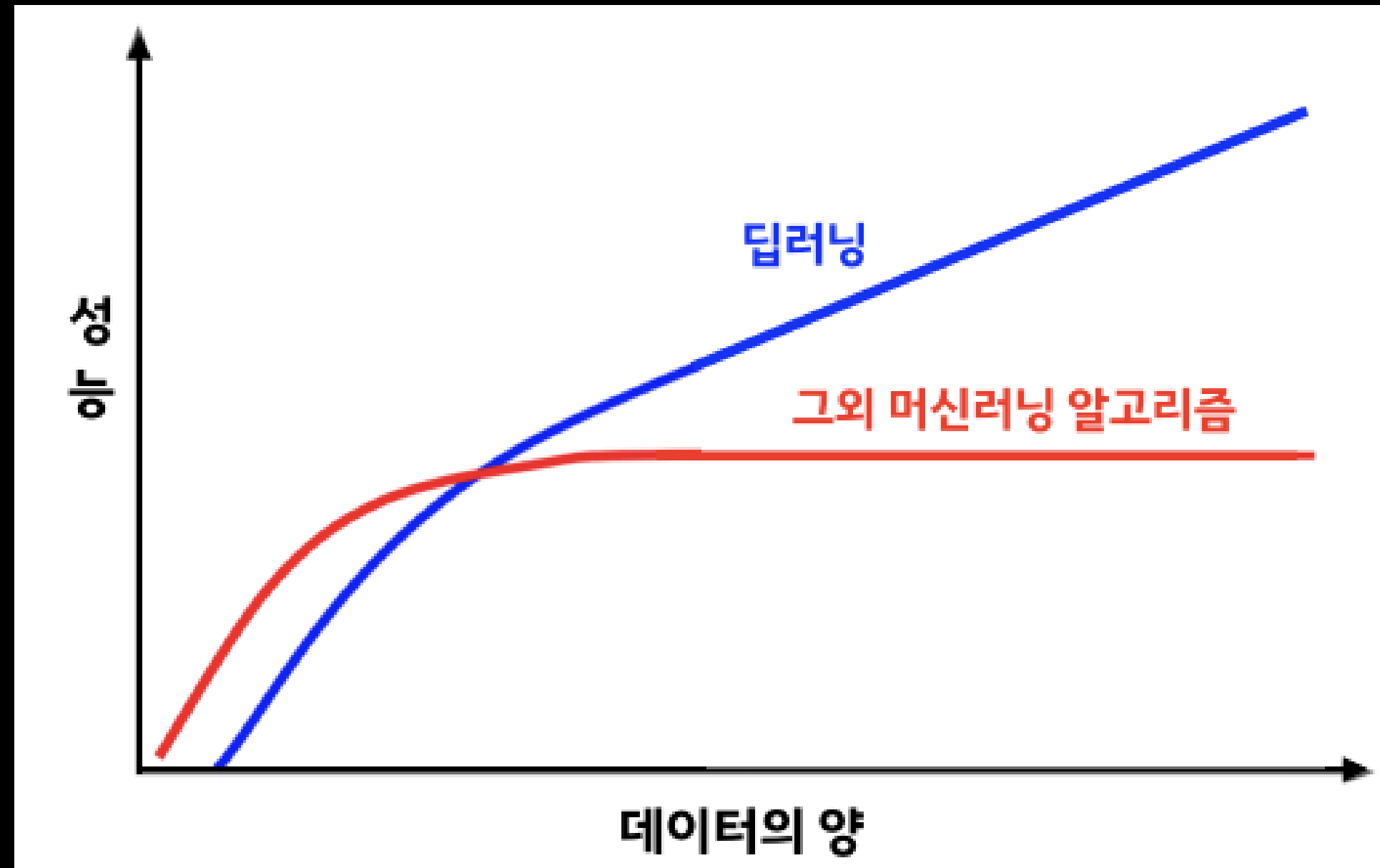




인공지능 기반 소규모 스타트업에서는 어떤 어려움을 겪고 있을까?



## Problem



인공지능의 성능은 라벨링된 데이터의 양에 좌우된다





Google

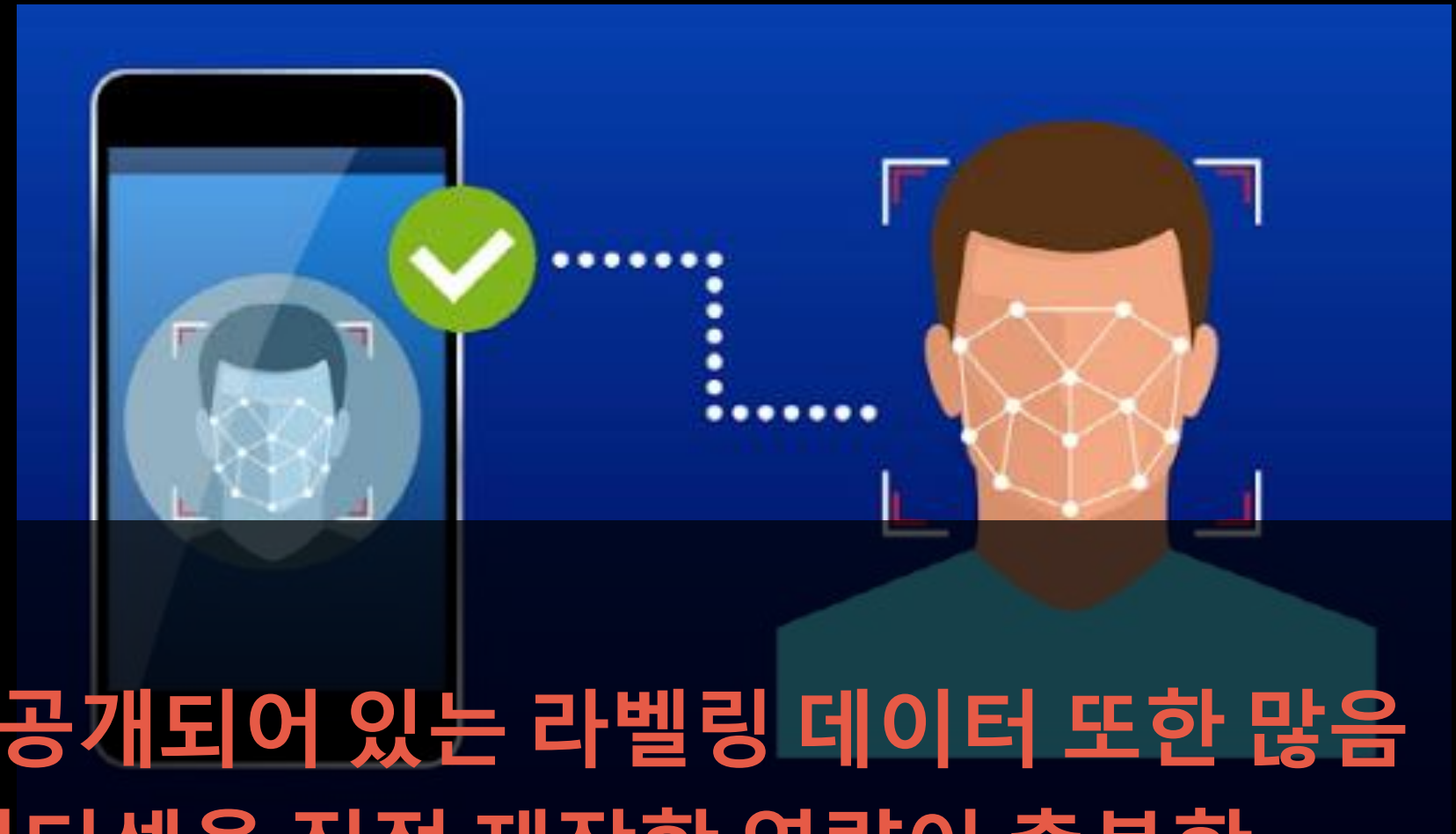
Baidu 百度



Microsoft

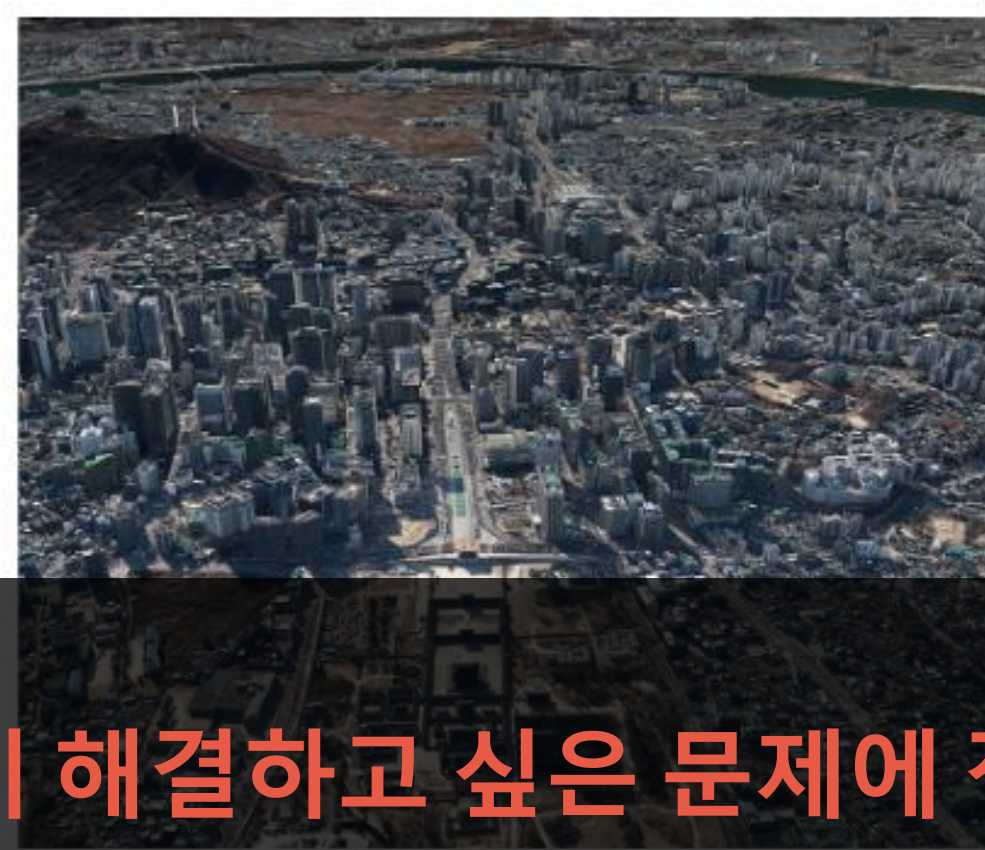
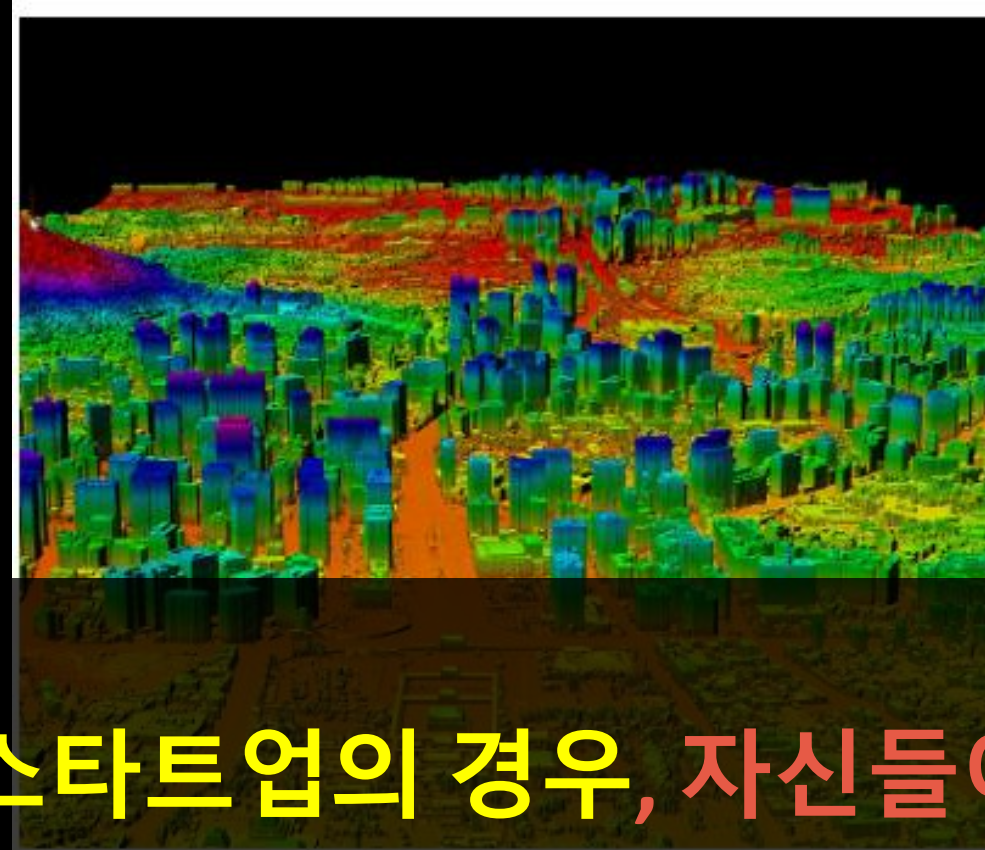


facebook



얼굴인식, 자율 주행 등 산업 규모가 큰 경우, 공개되어 있는 라벨링 데이터 또한 많음  
또한 대기업의 경우, 자본과 인력을 통해 데이터셋을 직접 제작할 역량이 충분함





스타트업의 경우, 자신들이 해결하고 싶은 문제에 집중하며,  
아무도 해결하려고 시도해보지 않았던 문제이기 때문에 데이터가 존재하지 않음



# Deep Learning and Machine Learning in the Design Engineering Workflow

Images, video, sound, vibration, etc.;  
real data and data synthesized from  
simulation

Feature engineering for machine learning;  
domain-specific signal preprocessing

Deep learning and machine learning  
algorithms; statistical methods

Learning algorithms deployed for  
inference

## Access Data



Sensors



Files



Databases

## Analyze Data



Data  
exploration



Preprocessing



Domain-specific  
algorithms

## Develop



AI model



Algorithm  
development



Modeling and  
simulation

## Deploy



Desktop apps



Enterprise  
systems



Embedded  
devices



# Deep Learning and Machine Learning in the Design Engineering Workflow

Images, video, sound, vibration, etc.;  
real data and data synthesized from  
simulation

Feature engineering for machine learning;  
domain-specific signal preprocessing

## Access Data



Sensors



Files



Databases

## Analyze Data



Data  
exploration



Preprocessing



Domain-specific  
algorithms

Deep learning and machine learning  
algorithms; statistical methods

## Develop



AI model



Algorithm  
development



Modeling and  
simulation

Learning algorithms deployed for  
inference

## Deploy



Desktop apps



Enterprise  
systems



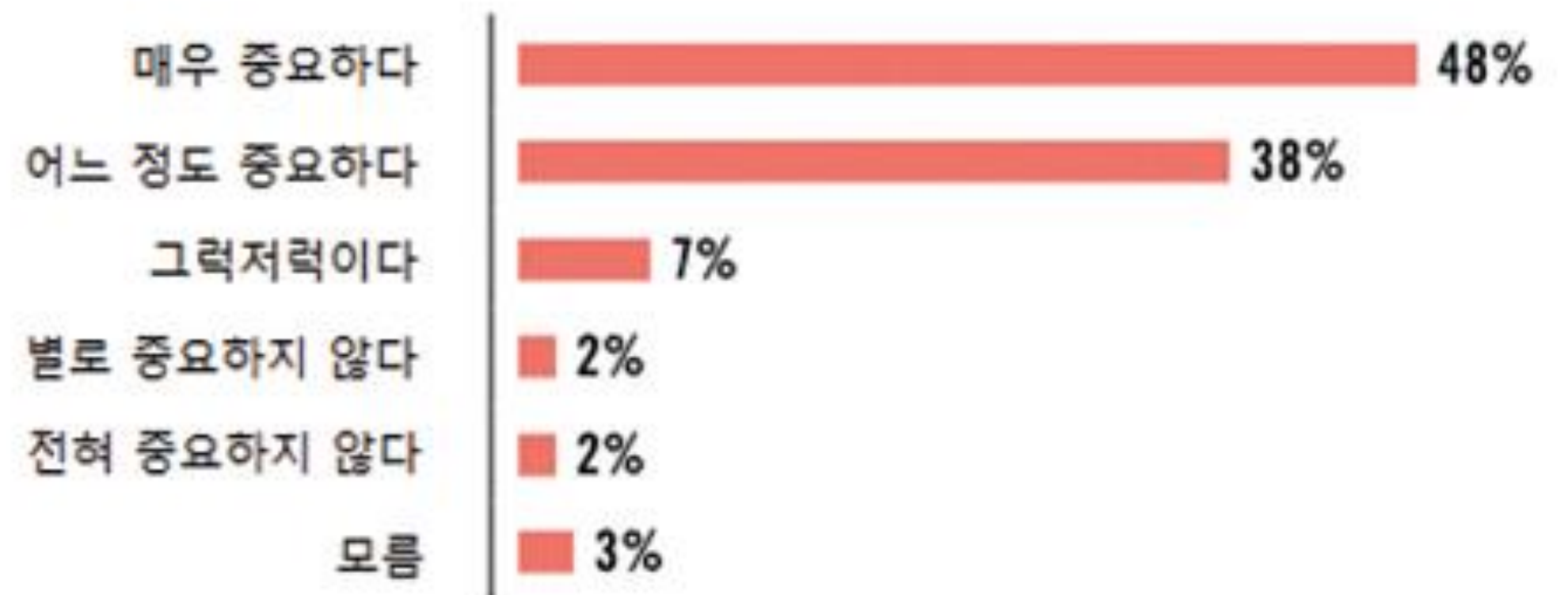
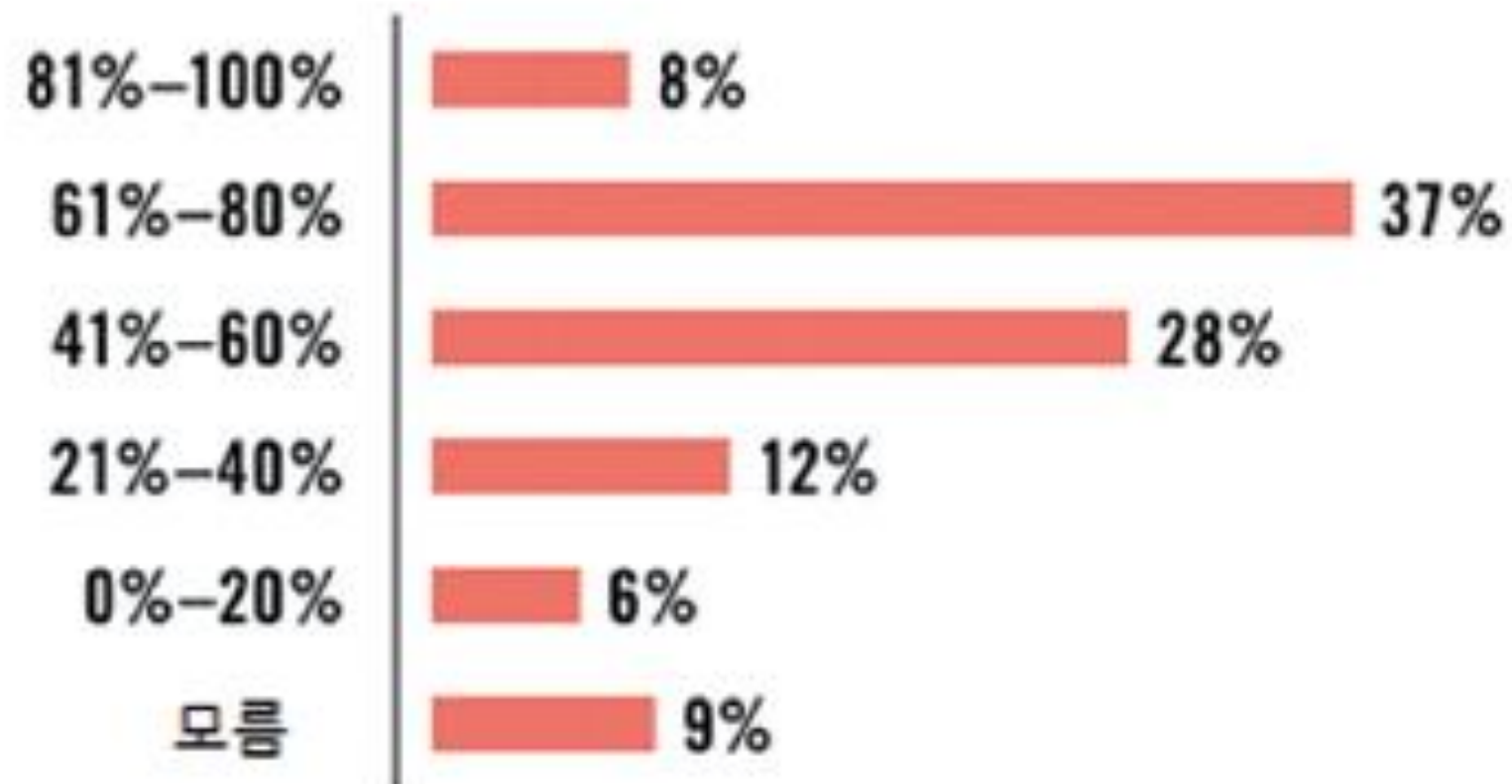
Embedded  
devices

인공지능을 통해 문제를 해결하는 과정에서, 데이터 수집 및 분석이 절반 이상을 차지함



데이터 전처리를 위해 필요한 시간은 전체 BI 및 분석 프로젝트에 걸린 시간의 어느 정도를 차지 합니까?

데이터 전처리를 과정에 소요되는 시간과 자원을 줄이는 것이 귀사에 얼마나 중요합니까?



출처: IT DAILY, 2017

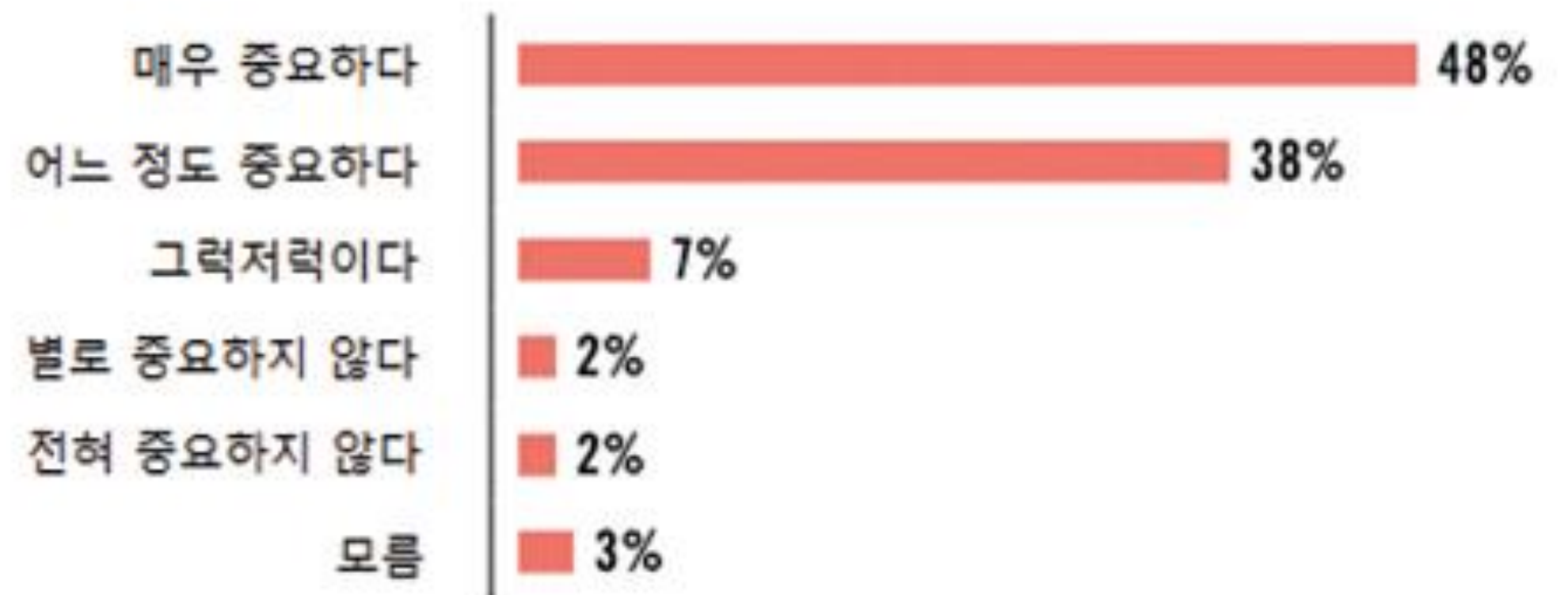
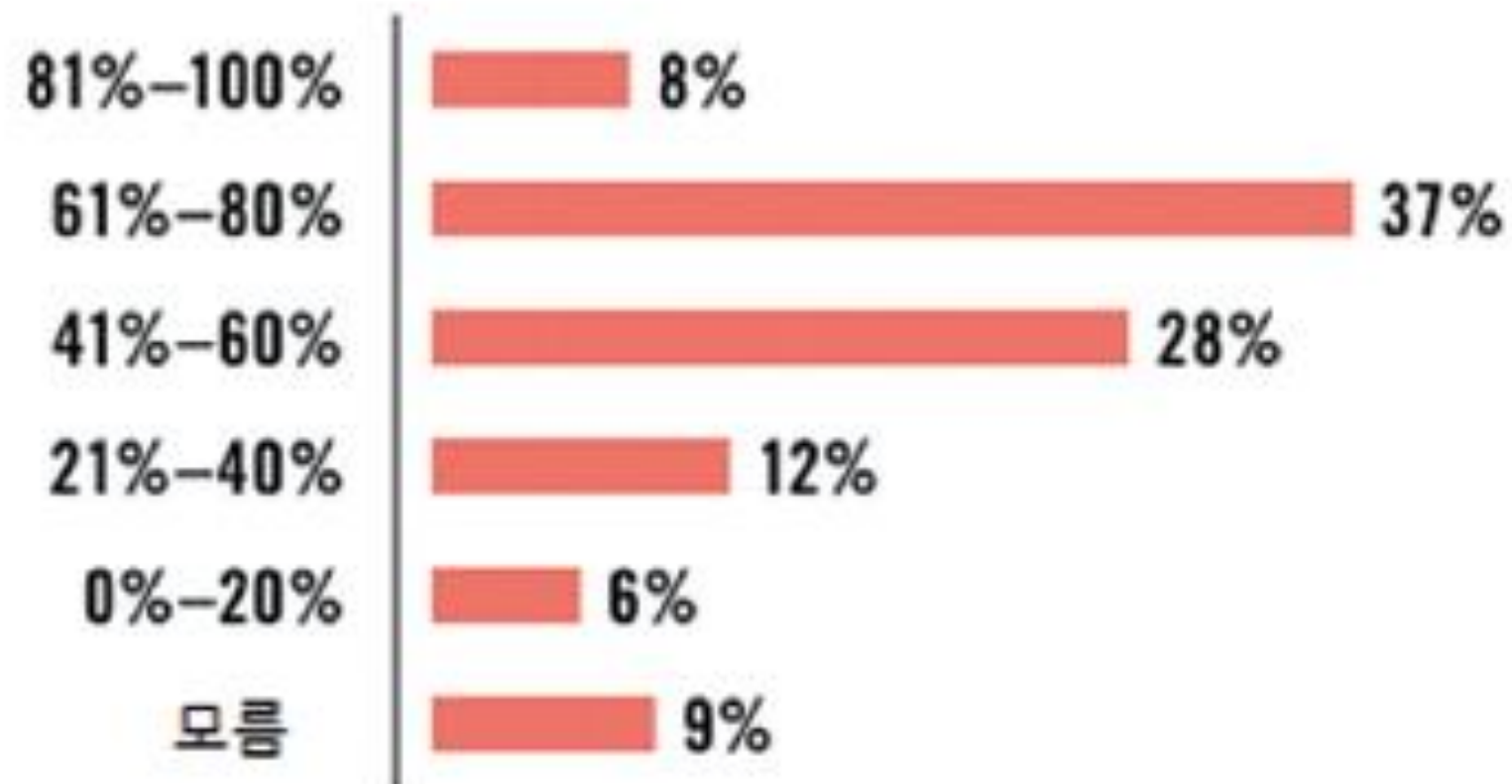
데이터 수집 및 전처리에 소요되는 시간은 전체 과정에서 **50% 이상**

소규모, 고급인력 위주로 구성된 인공지능 기반 스타트업에선 데이터 수집 및 전처리는 **골칫덩어리**



데이터 전처리를 위해 필요한 시간은 전체 BI 및 분석 프로젝트에 걸린 시간의 어느 정도를 차지 합니까?

데이터 전처리를 과정에 소요되는 시간과 자원을 줄이는 것이 귀사에 얼마나 중요합니까?



출처: IT DAILY, 2017

하지만, 데이터 수집 및 라벨링은 기업의 자산이자 경쟁력이며, 기술개발 전에 선행되어야 한다



# Smart Labeler



## Smart Labeler

인공지능 학습을 제외한  
데이터 관련 모든 과정에  
대한 원스톱 서비스 제공



Raw data 수집



데이터 분류



데이터 라벨링



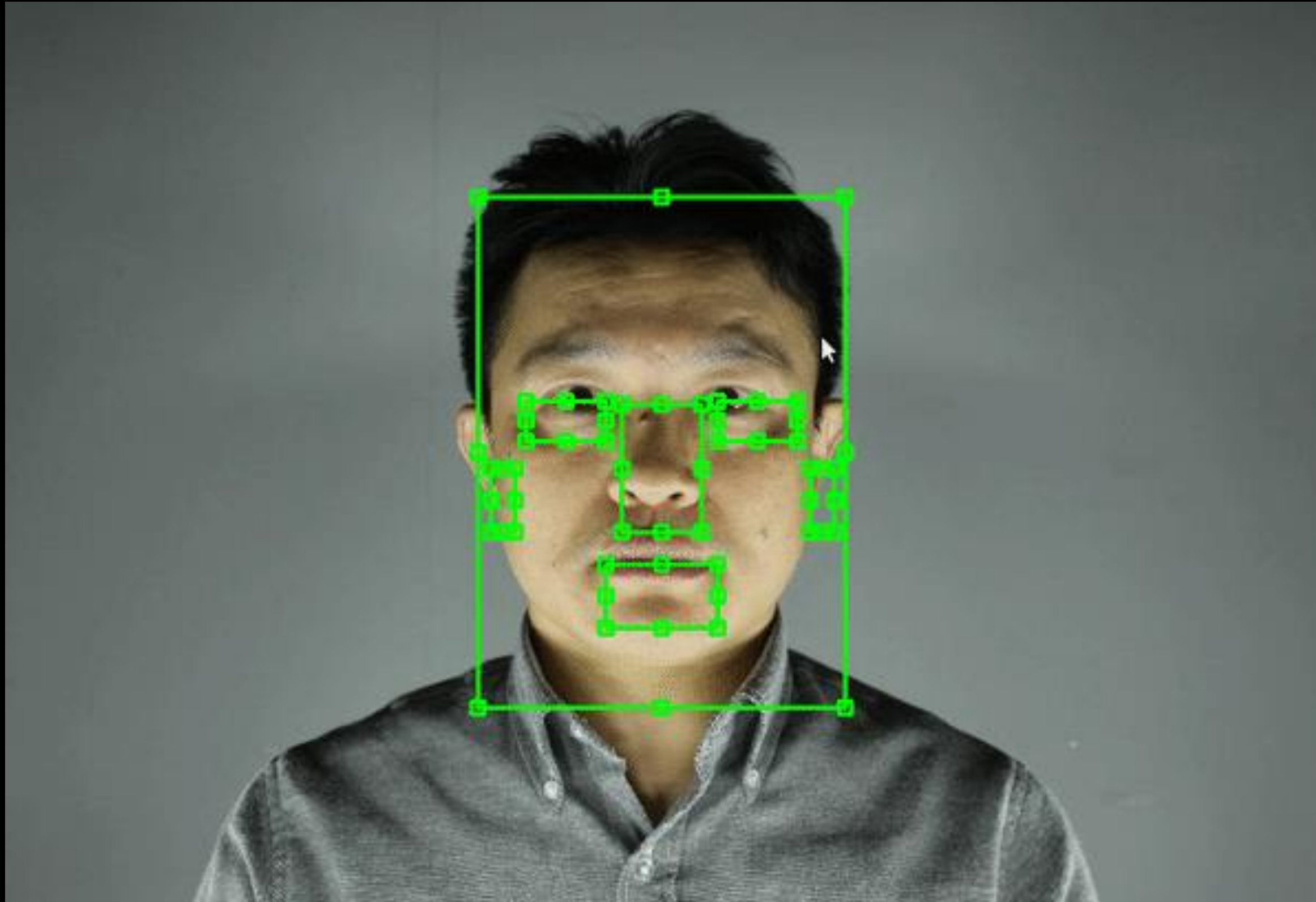
데이터 분석



인공지능 서비스 개발

인공지능 기반 소규모 스타트업에게 '골칫덩어리'인 데이터 수집부터 라벨링까지의 전 단계에 대한 서비스 제공  
스타트업은 인공지능 개발에만 모든 인력과 자원을 집중 가능

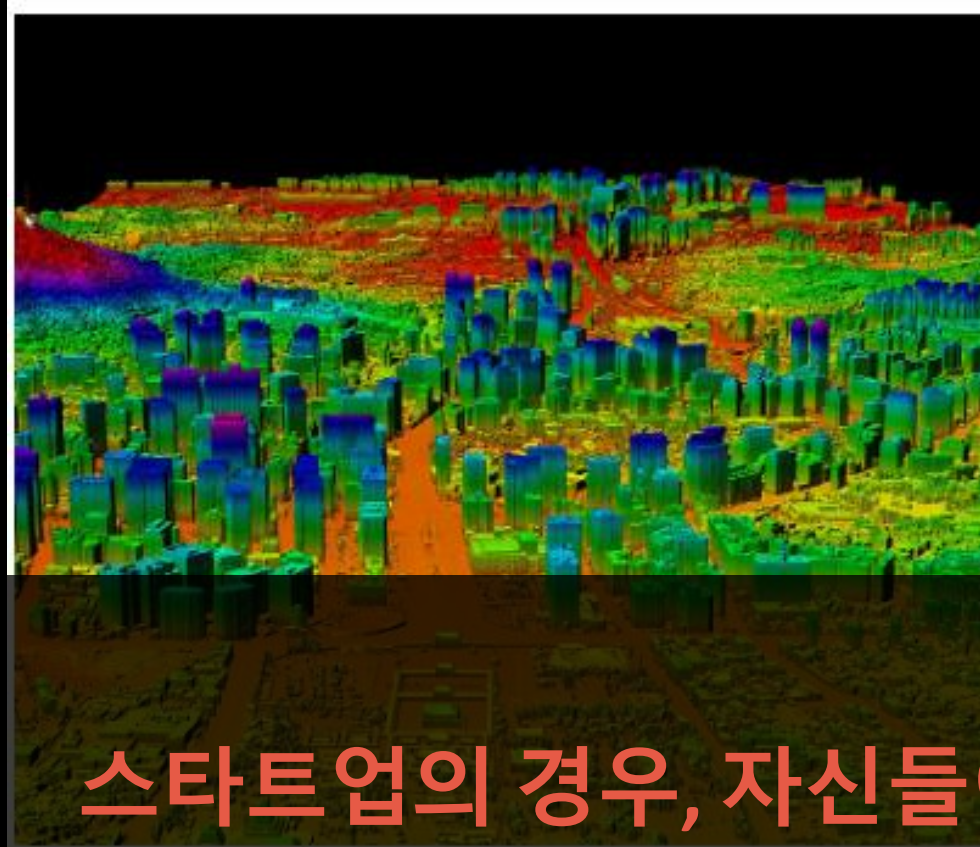




AHAM 사의 한국인 얼굴 인식 데이터셋 라벨링 과제를 해결하기 위해  
라벨링에 사용될 크기 조정과 드래그를 통한 이동이 쉽게 가능한 라벨링 툴 직접 개발



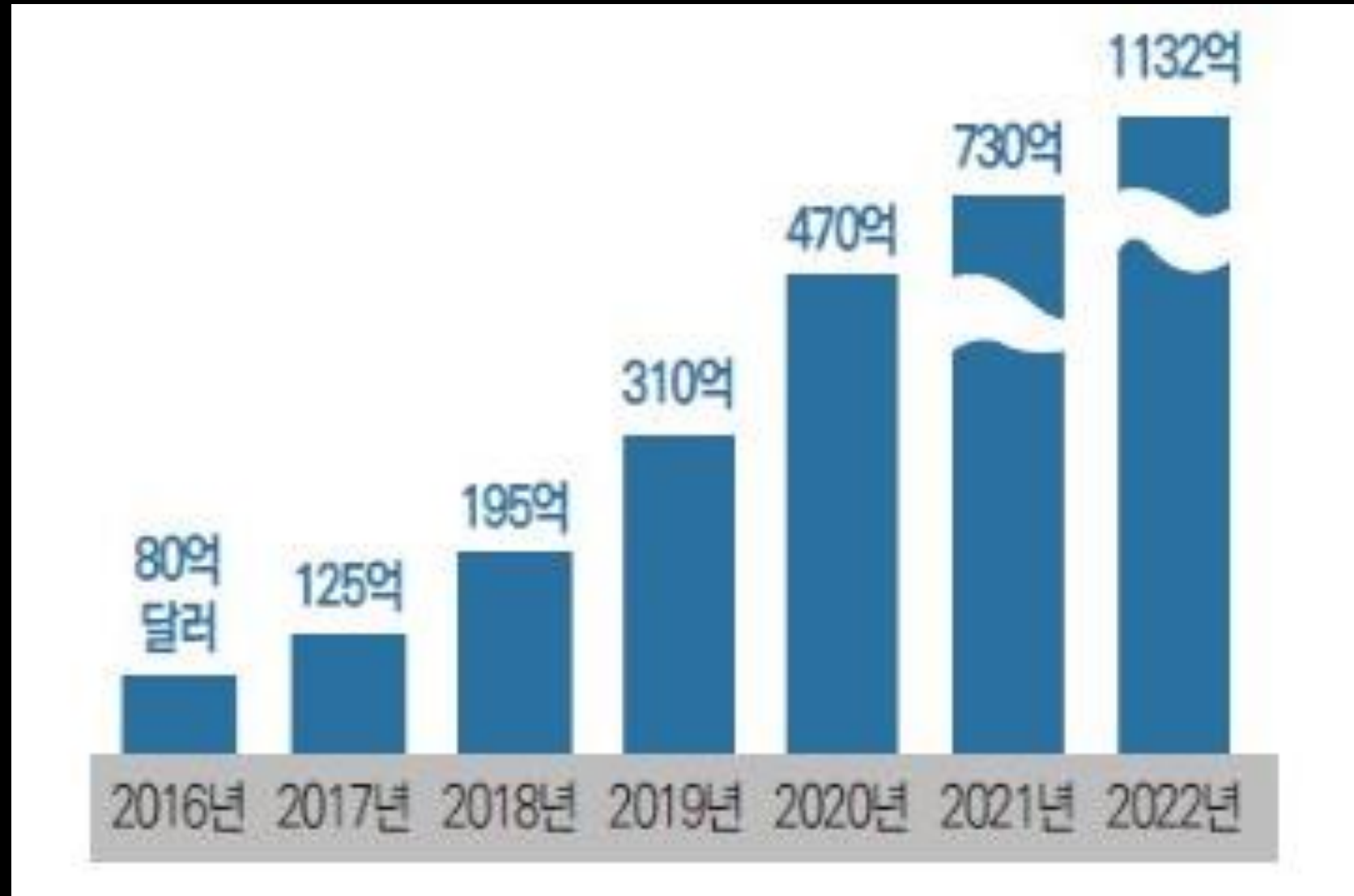
## 잠재적인 고객 설정



스타트업의 경우, 자신들이 해결하고 싶은 문제에 집중하며,  
아무도 해결하려고 시도해보지 않았던 문제이기 때문에 데이터가 존재하지 않음

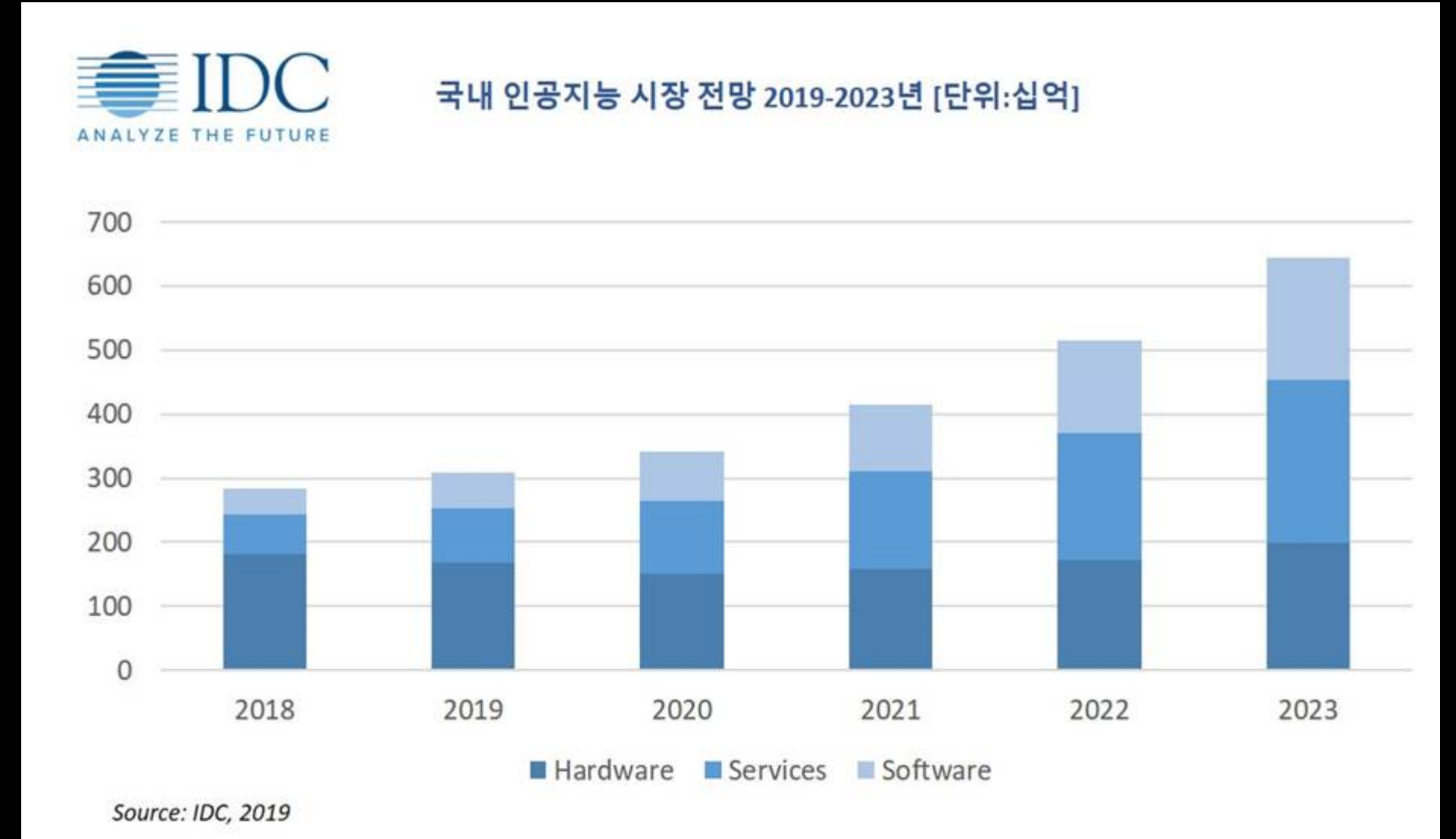


# 시장 규모



출처: 융합연구정책센터

글로벌 AI 시장 연평균 성장률 19.7%, 시장규모 100조원



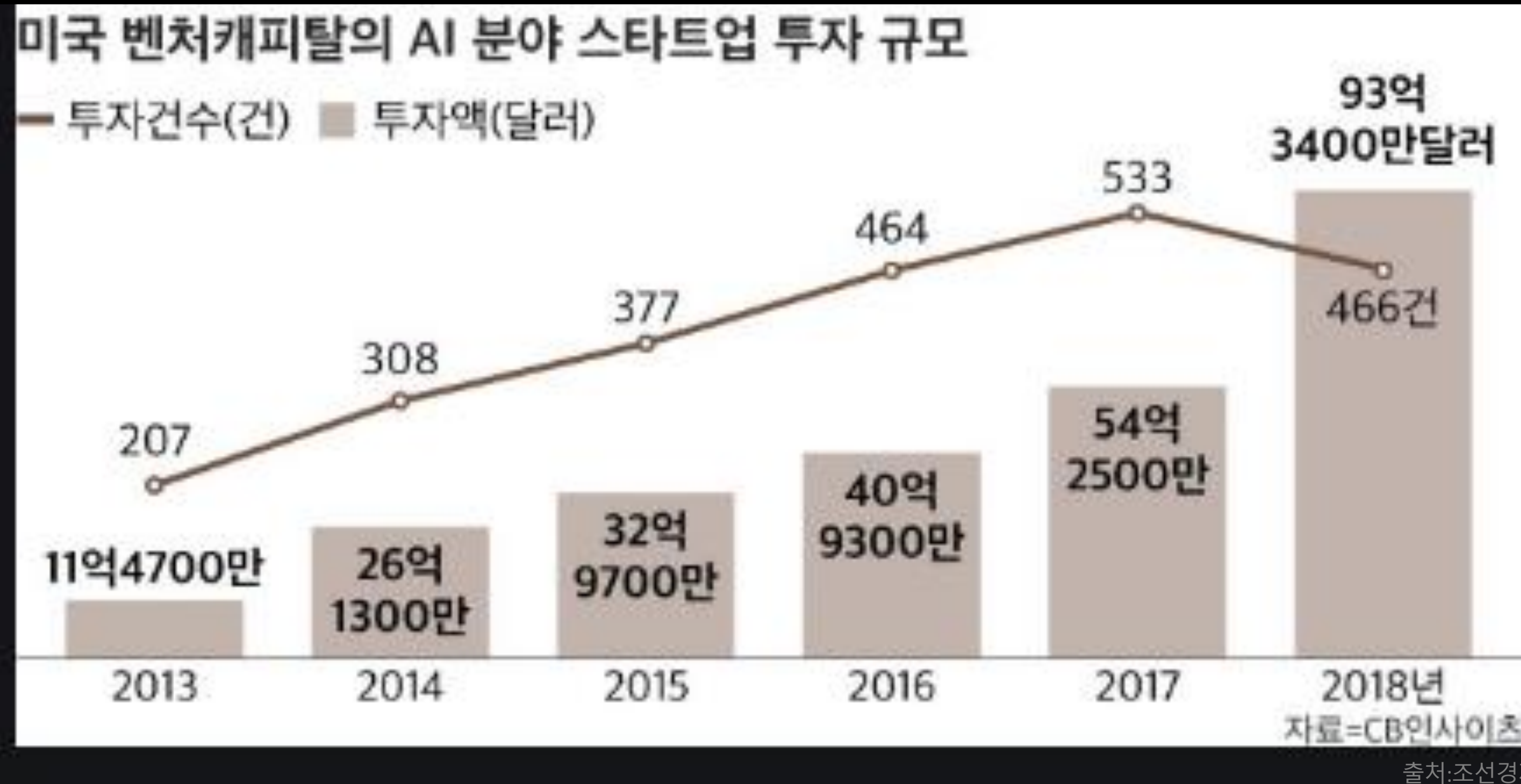
Source: IDC, 2019

출처: 정보통신신문

국내 AI 시장 2023년 까지 연평균 17.8% 성장, 시장규모 6조원



## 타겟 시장 규모



아무도 시도해보지 않았던 문제를 해결하려는 AI 스타트업을 타겟으로 설정하여 고객 및 시장 규모 구체화



## 경쟁사 분석



□ Superb AI





# 경쟁력





# 경쟁력



## Smart Labeler

인공지능 학습을 제외한  
데이터 관련 모든 과정에  
대한 원스톱 서비스 제공



Raw data 수집



데이터 분류



데이터 라벨링



데이터 분석



인공지능 서비스 개발



# 경쟁력



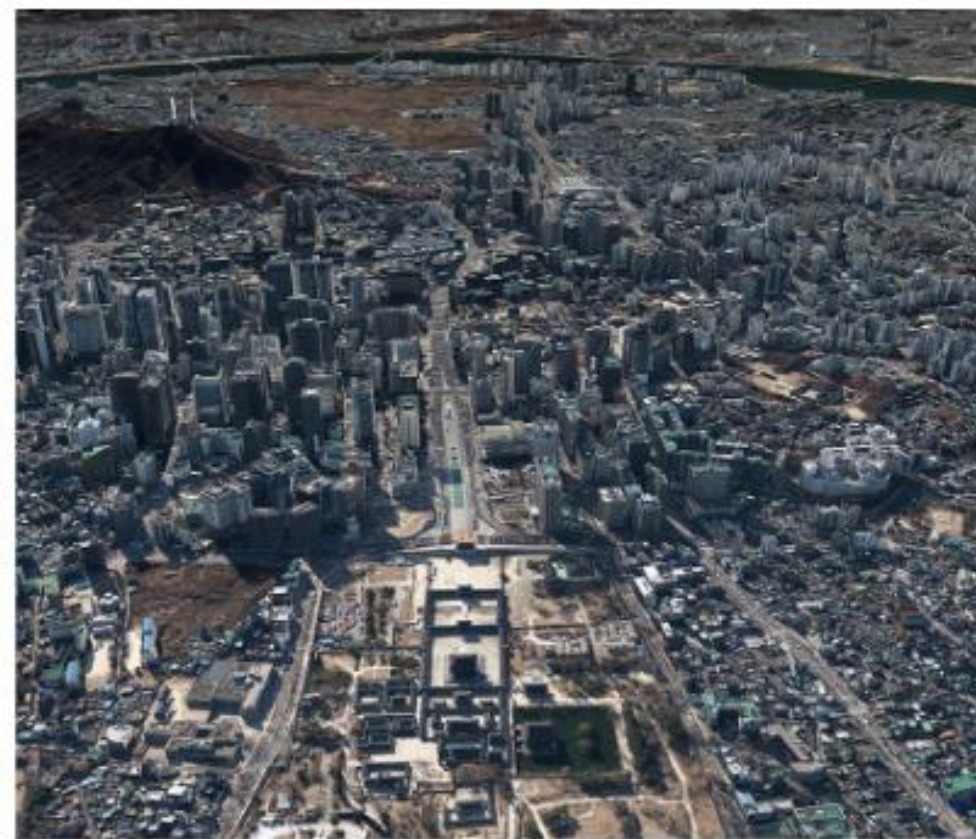
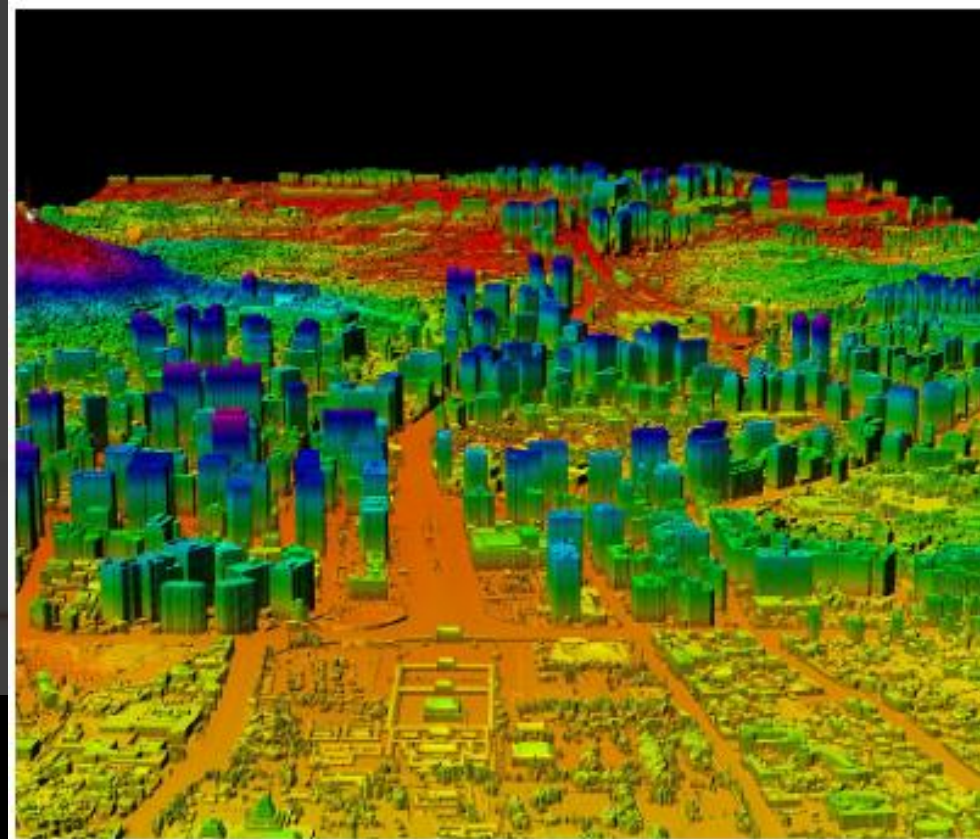
빅데이터로  
물고기  
키운다

바다 위  
산업혁명  
아쿠아팜

CARD  
NEWS  
전자신문

## Smart Labeler

인공지능 학습을 제외한  
데이터 관련 모든 과정에  
대한 원스톱 서비스 제공



가장 첫 단계인, 데이터를 어디서 어떤 방법으로 수집할 것인가? 에 대한 솔루션 제공



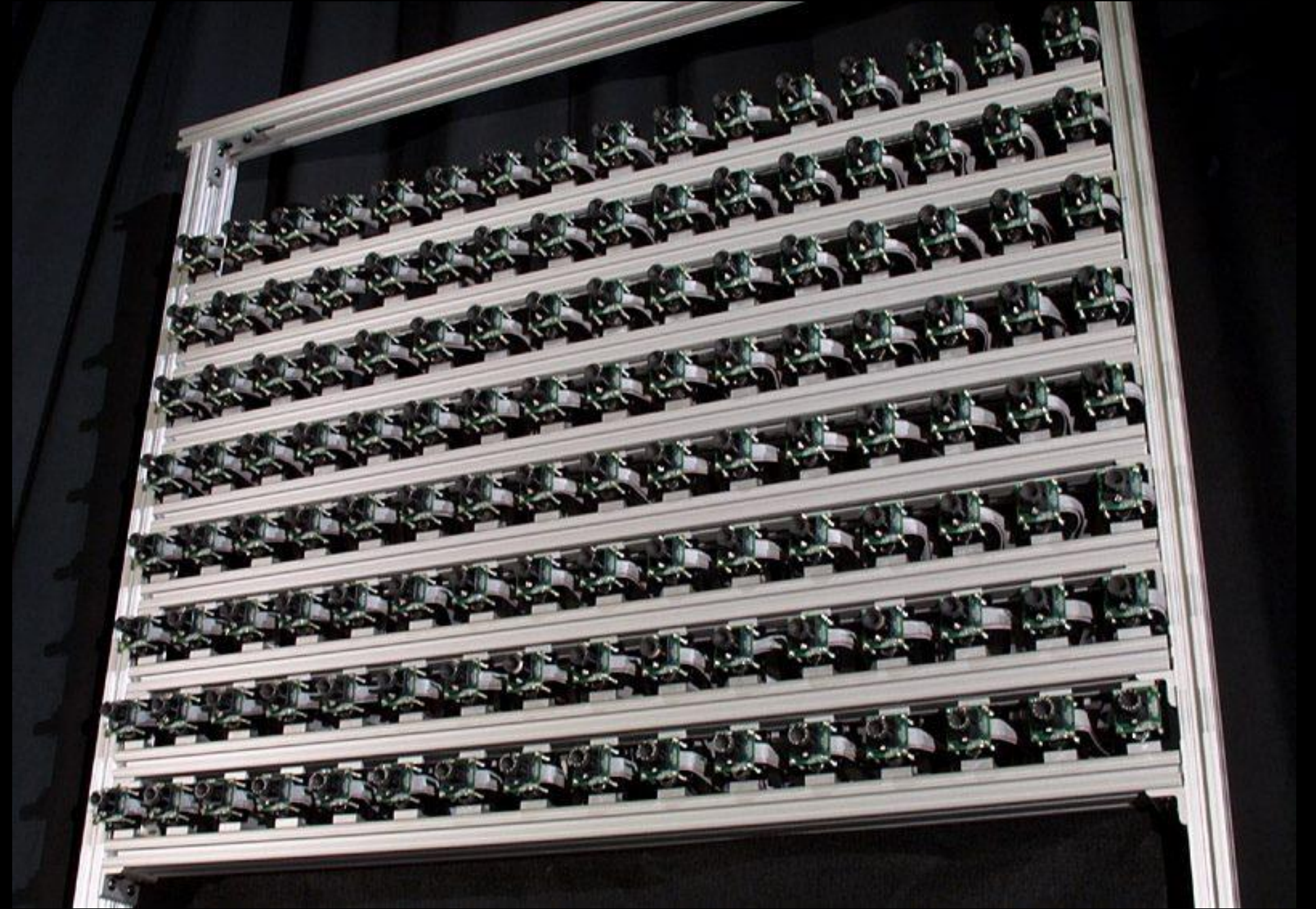


NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis



인공지능 기반으로 1장의 사진 만으로 3D 이미지를 생성하는 연구, 2020. 05





어떤 문제를 인공지능으로 해결하려는 스타트업이나, 연구자 들 에게는 데이터가 존재하지 않아서  
직접 데이터를 수집후 라벨링 해야하는 경우가 대부분 임



# 경쟁력



해결하려는 문제에 맞는 데이터를 어디서, 어떻게 수집할 것인지에 대한 고민을 스타트업이나 연구자가 꼭 해야 할까?  
문제 해결과 인공지능 개발에 집중!



## 비즈니스 모델

데이터 유형	목표	단위	등급 1	등급 2
이미지	분류	이미지	\$35	\$25
	경계 상자	경계 상자	\$63	\$49
	세분화	세그먼트	\$870	\$850
	회전 상자	경계 상자	\$86	\$60
	다각형/폴리라인	다각형/폴리라인	\$257	\$180
동영상	분류	5초 동영상	\$86	\$60
	객체 감지	경계 상자	\$86	\$60
	객체 추적	30초 동영상의 객체	\$686	\$480
	이벤트	30초 동영상의 이벤트	\$214	\$150
텍스트	분류	50단어	\$129	\$90
	감정	50단어	\$200	\$140
	항목 추출	항목	\$86	\$60

2020  
데이터  
바우처  
지원사업

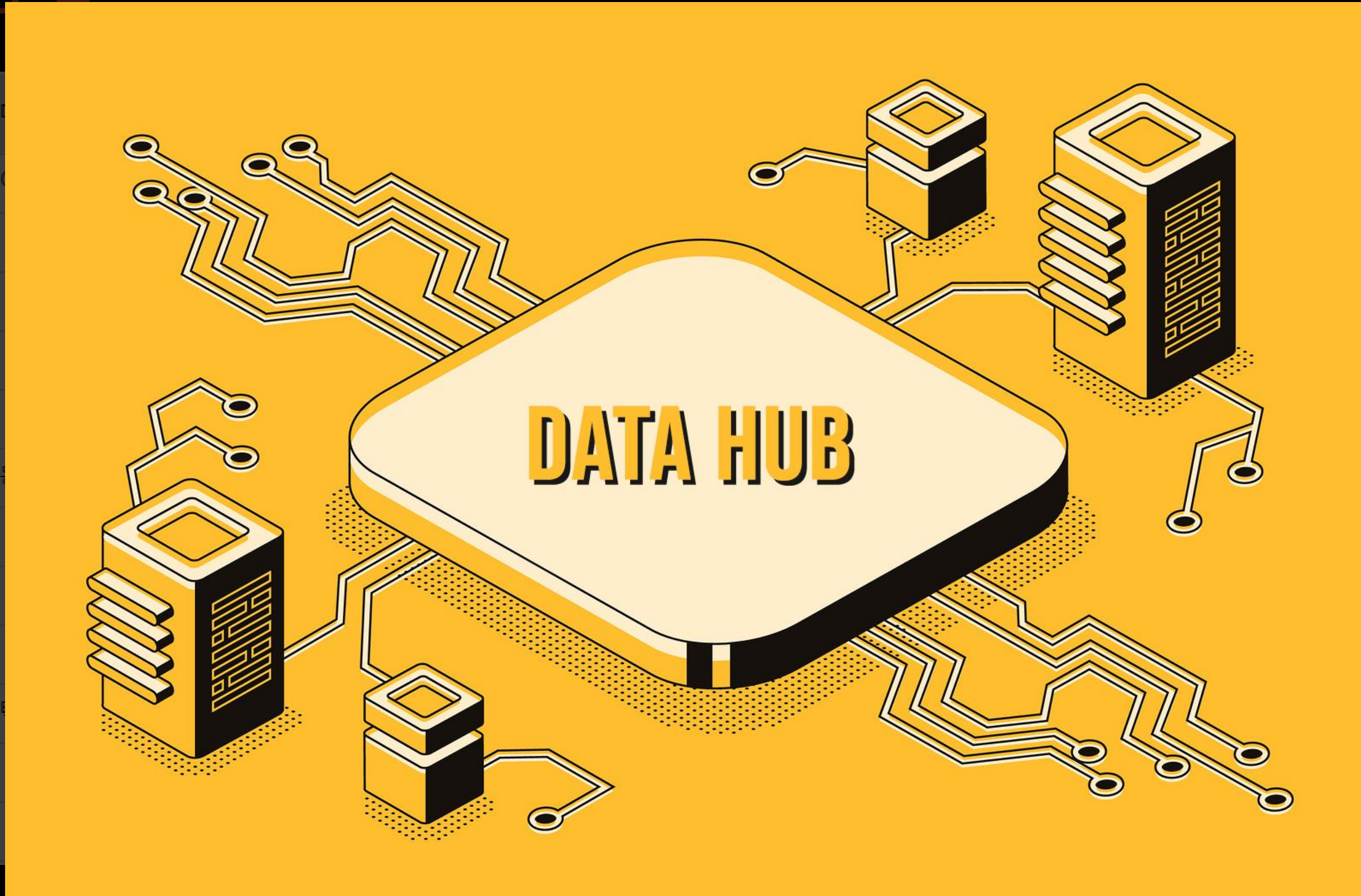
**혜택안내** 정부지원금 최대 **7천만원** + AWS 크레딧 등  
최소 **1,000만원** 어치 셀렉트스타 자체 혜택

**지원자격** 예비창업자도, 스타트업도, 중소기업도,  
다른 정부 지원을 받고 있어도 지원가능!

[datavoucher.com](https://datavoucher.com)

데이터의 저작권 관련 라이선스를 사용자가 선택할 수 있으며, 이에 따른 차등적인 수수료 부과  
오픈소스 라이선스 선택 시, 경쟁사에 비해 저렴한 가격에 솔루션 제공

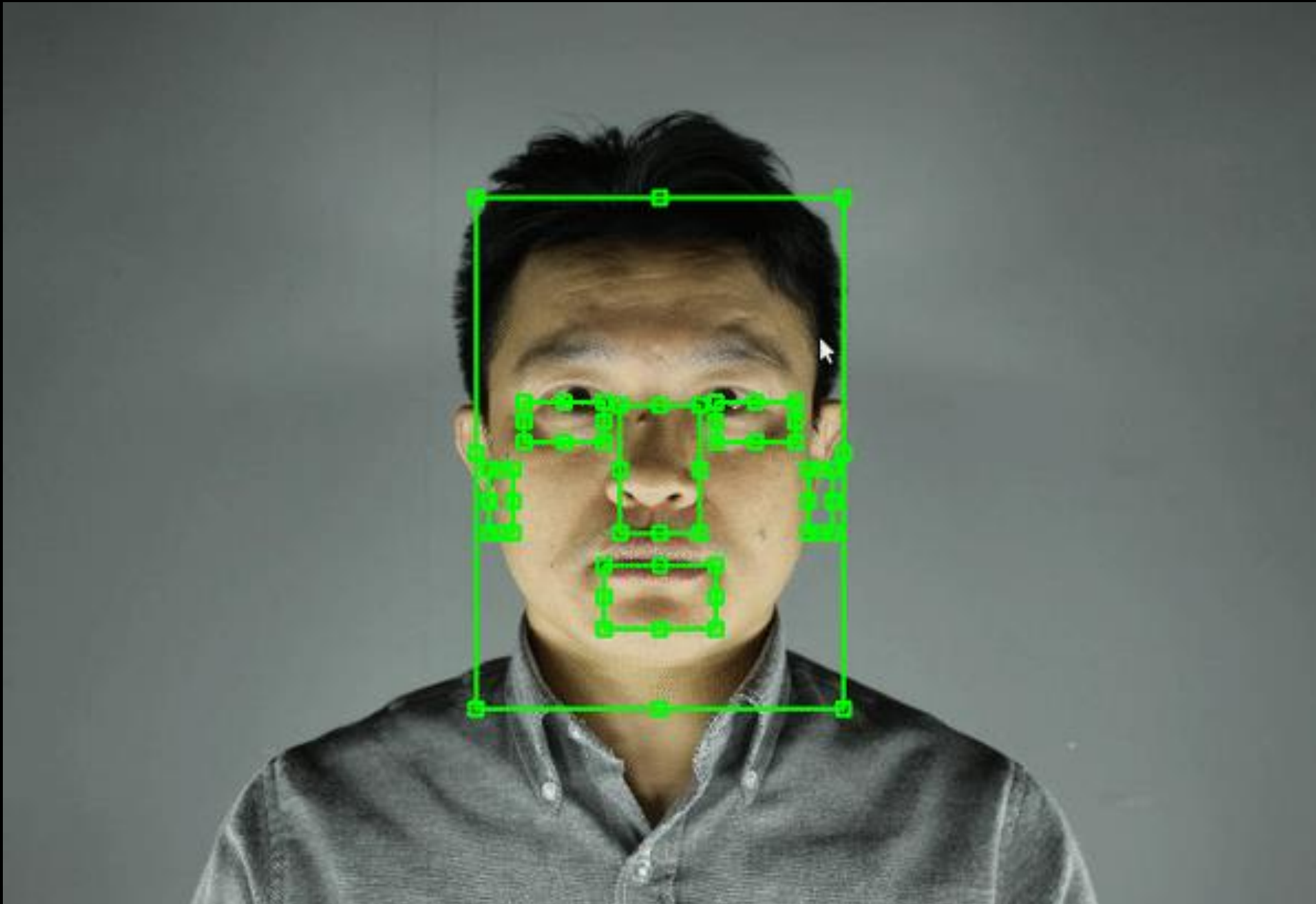




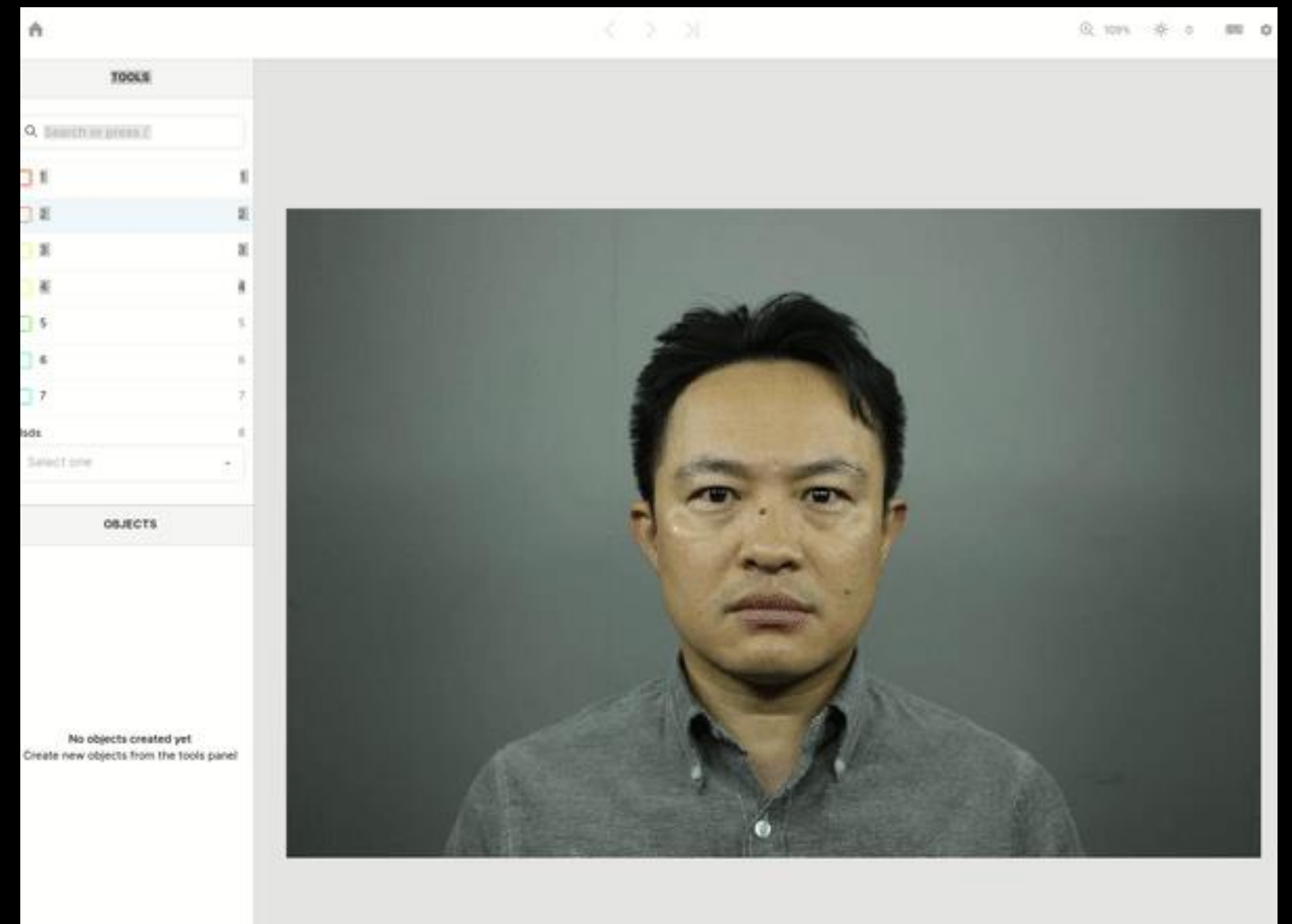
데이터의 저작권 관련 권익보호를 사용자가 선택할 수 있으며, 임대료를 차등적인 수수료 부과  
이전에는 존재하지 않던, 새로운 데이터들의 허브, 플랫폼화  
오픈소스 라이선스 선택 시, 경쟁사에 비해 저렴한 가격에 솔루션 제공



## 최종 결과물 및 성과



Smart Labeler  1장 라벨링에 10초 소요



경쟁사의 라벨링 툴, 1장 라벨링에 20~30 초 소요

AHAM 사의 한국인 얼굴 인식 데이터셋 라벨링 제작과 관련하여 경쟁사 보다 효율적인  
라벨링 툴 자체 개발, 소스코드 제공 및 오픈소스화



## 최종 결과물 및 성과



**X 3600**

**총 3600 장** 의 한국인 얼굴 인식 데이터셋 라벨링 완료 및 제공





데이터 유형	목표	단위	등급 1	등급 2
이미지	분류	이미지	\$35	\$25
	경계 상자	경계 상자	\$63	\$49
	세분화	세그먼트	\$870	\$850
	회전 상자	경계 상자	\$86	\$60
	다각형/폴리라인	다각형/폴리라인	\$257	\$180
동영상	분류	5초 동영상	\$86	\$60
	객체 감지	경계 상자	\$86	\$60
	객체 추적	30초 동영상의 객체	\$686	\$480
	이벤트	30초 동영상의 이벤트	\$214	\$150
	텍스트	50단어	\$129	\$90
텍스트	감정	50단어	\$200	\$140
	항목 추출	항목	\$86	\$60



X 3600

구글 클라우드의 데이터 라벨링 서비스 기준, 약 200만 원 의 가치를 가진 데이터셋 제작 후 제공



## Reaction & Plan

AHAM 이용수 대표님:

인공지능 기반 스타트업의 데이터셋 관련 페인 포인트와 니즈를 해결할 수 있는 서비스.

추후 정식 서비스 출시 시 사용 의향이 있음.

: 실질적인 고객의 니즈와 시장성을 검증하기 위해 추상적인 시장 보다

과제를 제시해 주신 AHAM 사를 첫 번째 고객으로 설정하여 문제를 해결, 긍정적인 피드백





# Team



**박재현 (Entrepreneurs)**

- 동국대 컴퓨터공학과 4학년
- 사업계획서 및 오픈소스 분석 담당
- 창업관련 정부사업 지원 경험 다수



**채현욱 (Marketing)**

- 동국대 컴퓨터공학과 4학년
- 회계 서류 담당



**김경민 (Developer)**

- 동국대 컴퓨터공학과 4학년
- 인공지능 기술 개발 담당
- 특허법 관련 지식 보유 (변리사 준비)



**김인제 (Developer)**

- 동국대 컴퓨터공학과 4학년
- 고요한택시, 웨코 창업 경험 보유
- 시드투자 (1억 이상 규모) 유치 경험



# Q & A

# Thank you

본 프로젝트는 (주) 에이엠 의 지원을 받아 진행되었습니다.