

# Machine learning 을 활용한 Adaptive learning platform 구축

2018. 09. 07

김민우 대표

Kidaptive Asia



Edu-tech의 특징

&

Why Kidaptive?



# Edu-tech 프로젝트의 특징

Education + Technology

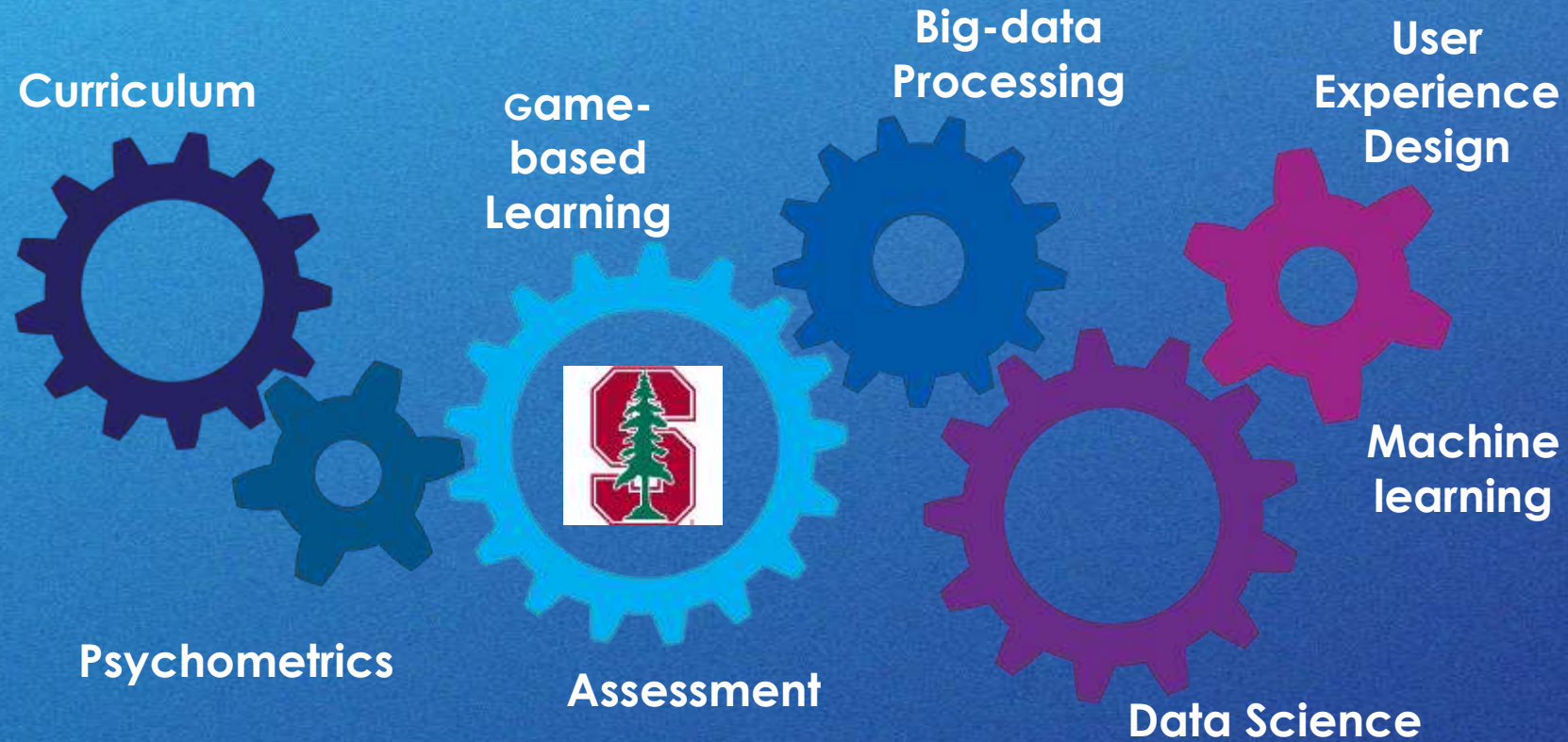
Contents  
Curriculum  
Meta data  
Assessment  
Psychometrics

Big data  
Cloud computing  
Machine learning  
Deep learning  
SDK

양 쪽을 모두 깊이 있게  
이해하고 있는  
강력한 팀이 필요함



# Kidaptive's talents





# 교육 빅데이터 분석을 위한 전문가 집단



요시나 베르하겐  
박사  
심리모델링 전문가



딜런 아레나  
박사  
맞춤형교육 전문가



데이빗 핫필드  
박사  
교육평가 전문가



웹 필립스  
박사  
심리언어 및 통계 전문가



우마 비쉬  
박사  
빅데이터분석 전문가



지안  
박사  
데이터 모델링 전문가



# 현직 Stanford University 교수로 구성된 자문단



댄 슈왈츠 교수

교육대학원 원장

[자문위원장]



에드 헤어텔 교수

교육대학 명예교수



엠마 러드윅 교수

정신분석학

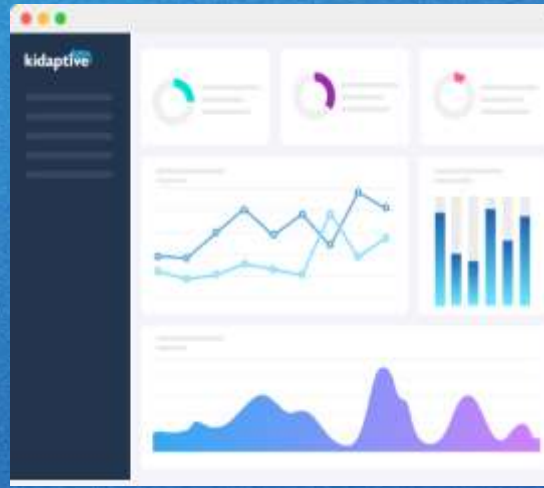


# Kidaptive ALP



## Real time Adaptivity

Bayesian-IRT 이론을 기반으로 개별 학습자에게 가장 필요한 다음 문제를 실시간으로 제공합니다.



## Actionable Insights

교사, 학습자, 관리자 등 학습의 주요 Stake holder들에게 학습자의 진도, 성취, 성향에 대한 분석과 추천을 포함한 다양한 Insights를 제공합니다.



## Personalized Learning

개별 학습자에게 커리큘럼 전반에 걸쳐 문항 뿐 아니라, Reading contents, 학습 영상 등 최적의 과정을 맞춤형으로 추천합니다.



formation | 8





# Machine Intelligence LANDSCAPE

## CORE TECHNOLOGIES

### ARTIFICIAL INTELLIGENCE

IBM WATSON MetaMind  
Numenta ai-one  
Cycorp Research nano  
Reactor SCALED INTELLIGENCE

### DEEP LEARNING

vicarious  
facebook  
Google  
SKYMIND  
Baidu  
ersatz  
SignalSense

### MACHINE LEARNING

rapidminer context  
Oxdata H2O DATA  
LiftIgniter  
Alteryx  
GraphLab  
Alteryx  
Sense  
Alteryx

### NLP PLATFORMS

cortical.io  
idibon  
LUMINOSO  
wit.ai  
Makubli

### PREDICTIVE APIS

AlchemyAPI  
Google  
bigm  
Indico  
ALGORITHMIA  
Expect  
PredictionIO  
Labs

### IMAGE RECOGNITION

clarifai MADBITS  
DNNresearch  
VISENZE  
lookflow

### SPEECH RECOGNITION

GRIDSPACE  
popUP archive  
NUANCE

## RETHINKING ENTERPRISE

### SALES

Preact  
RelateIQ  
CLARABIDGE  
infer  
AVISO  
NGDATA  
FRAMED  
ATTENTIVY  
causata

### SECURITY / AUTHENTICATION

CROSSMATCH  
EYEVEERY  
CYLANCE  
c-njur  
BITSIGHT  
bionym

### FRAUD DETECTION

sift science  
ThreatMetrix  
Brighterion  
SOCURE  
feedzai  
VERAFIN

### HR / RECRUITING

TalentBin  
predikt  
gild  
entelo  
Connectifier  
hiQ

### MARKETING

brightfunnel  
CommandIQ  
RADIUS  
Telkport  
bloomreach  
AIRPR  
people pattern  
freemove

### PERSONAL ASSISTANT

Siri  
Cortana  
tempo  
KASISTO  
VIV  
Google now  
cleversense  
Robinlabs  
fuse machines  
CLARA LABS

### INTELLIGENCE TOOLS

ADATAD  
Quid  
FirstRain  
Palantir  
Digital Reasoning

## RETHINKING INDUSTRIES

### ADTECH

METAMARKETS  
dstillery  
rocketfuel  
YieldMo  
ADBRAIN

### AGRICULTURE

BLUE RIVER  
ceresimaging  
THE CLIMATE CORPORATION  
TerraAvion  
KOKO  
tule

### EDUCATION

Declar  
KNEWTON  
Coursera  
kidaptive

### FINANCE

Bloomberg  
alphasense  
Dataminr  
Kenshco  
minedabrook  
BINATIX

### LEGAL

Lex Machina  
COUNSELYTICS  
JUDICATA  
DiligenceEngine  
brightleaf  
RAVEL  
Brevia

### MANUFACTURING

SIGHT MACHINE  
MICROSCAN  
IVISYS  
BOLD REPAIR

### MEDICAL

Parzival  
Genescent  
grand round table  
transcriptic  
ZEPHYR  
bina  
TUTE

### OIL AND GAS

kaggle  
TACHYUS  
Futura  
AYASDI  
biota

### MEDIA / CONTENT

Outbrain  
SAILTHRU  
NarrativeScience  
Prismatic  
newsle  
wavii  
YPOJO  
ai

### CONSUMER FINANCE

Affirm  
inVenture  
BILLS GUARD  
LendUp  
LendingClub  
Kabbage

### PHILANTHROPIES

DataKind  
thorn  
DATA GUILD

### AUTOMOTIVE

Google  
T  
Continental  
Cruise

### DIAGNOSTICS

enlitic  
lumiat  
3SCAN  
ENTO

### RETAIL

BAY SENSORS  
PRISM SKYLARS  
celect  
euclid

## RETHINKING HUMANS / HCI

### AUGMENTED REALITY

Augmented Intelligence  
APX  
METRA  
blippar  
layar

### GESTURAL COMPUTING

THALMICLABS  
omek  
LEAP  
eyeSight  
GestureLink  
3Gear  
nod

### ROBOTICS

intel  
iRobot  
anvi  
LIVING ROBOTICS  
SoftBank  
Dynamics  
robotics

### EMOTIONAL RECOGNITION

affectiva  
BEYOND-VERBAL  
EMOTION  
cogito

## SUPPORTING TECHNOLOGIES

### HARDWARE

NVIDIA  
QUALCOMM  
rigetti  
XILINX  
NERVANA  
Arctia Learning

### DATA PREP

TRIFACTA  
tamr  
Paxata  
Alation

### DATA COLLECTION

diffbot  
kimono  
CrowdFlower  
WorkFusion  
Cinnote  
import



# 변화의 시대

*Featured by 4차 산업혁명*



Cloud

Alpha-go

Machine learning

Big data

AI

IoT

Deep learning

4차 산업혁명



"지금 직업의 47%는  
20년 내에 사라질 것"

영국 옥스포드 대학교, 칼 오스본 교수





뉴스

증권

부동산

정책·금융

기업

오피니언

뉴스 >

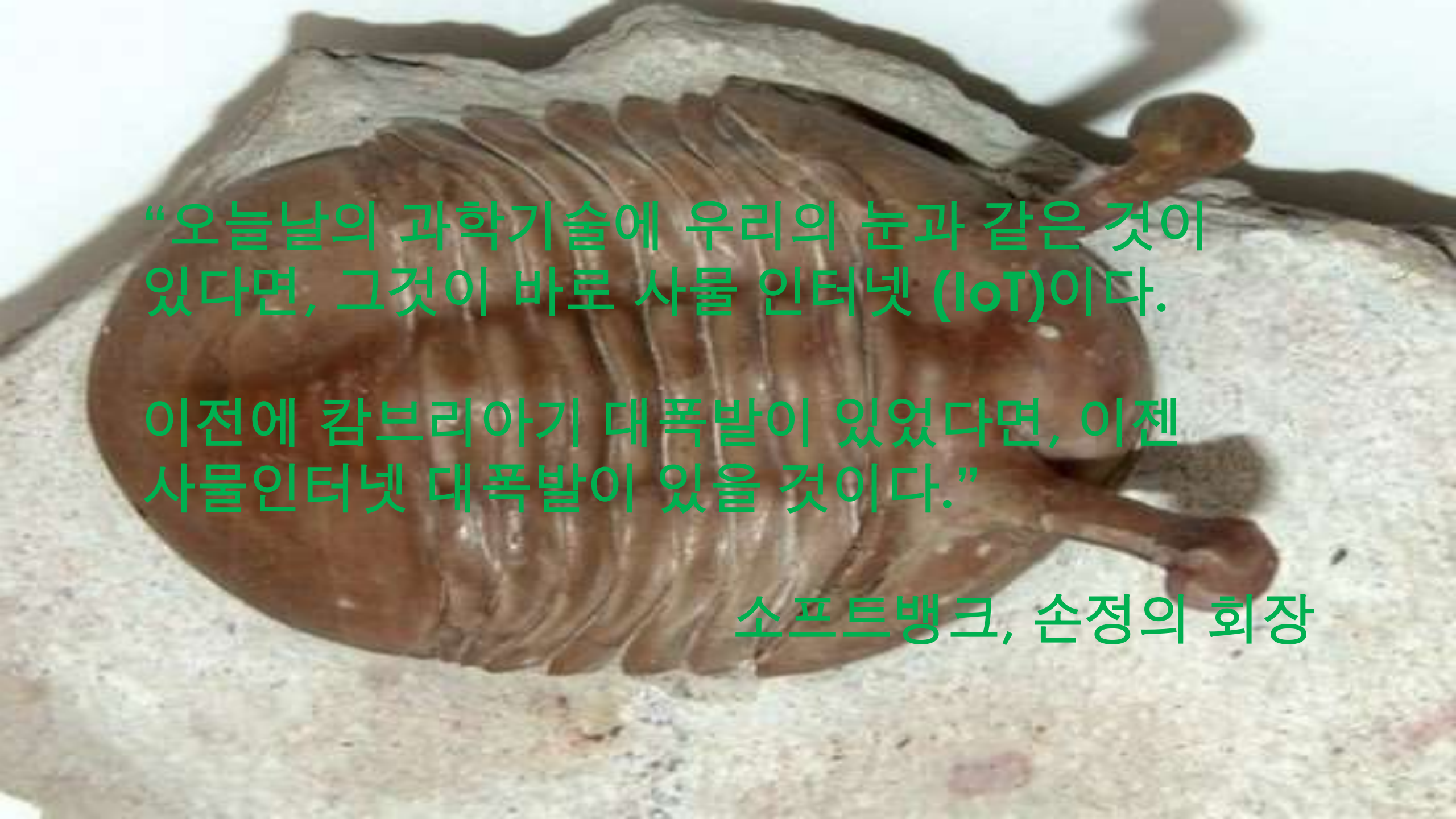
[이코노미조선] '100엔 초밥' 만들어 낸 '초밥  
로봇'...일본 회전초밥 체인점 '구라스시'

아나운서 72%

요리사 96%







“오늘날의 과학기술에 우리의 눈과 같은 것이 있다면, 그것이 바로 사물 인터넷 (IoT)이다.

이전에 캄브리아기 대폭발이 있었다면, 이젠 사물인터넷 대폭발이 있을 것이다.”

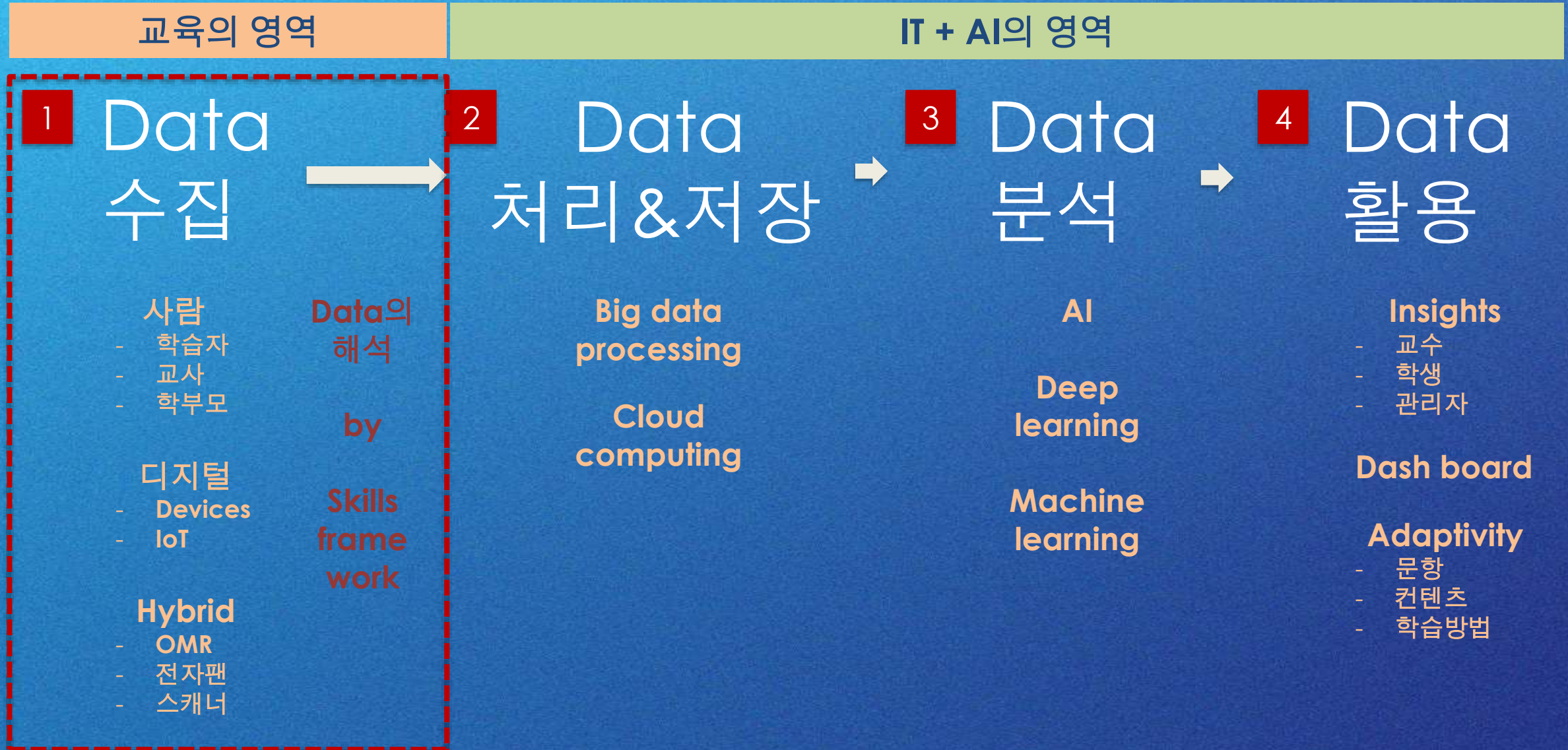
소프트뱅크, 손정의 회장



# Edu-AI service process



# AI Ed-tech project process (Ed + Tech)





# 1 Data 수집

## Key questions

[정의] 어떤 Data를 모을 것인가?

Behavior, Academic

[생성] 어떻게 Data를 모을 것인가?

채널, 방법, 주기

[구분] 어떻게 Data를 분류할 것인가?

Meta data tagging  
방법론 및 기준



# 1 Data 수집

## [생성] 생성 방법

### Full digital 방식

Digital device 활용 (Tablet, Smart phone, PC)

직접 Internet 방식, SDK 활용방식

문항당 소요 시간, 집중시간 등

세밀한 정보 파악 가능

실시간 Feed back 및 맞춤형 구현

인프라에 상당한 투자가 필요함

### Hybrid 방식

OMR 카드 → 중앙 Server

Smart OMR (답안입력 only)

좌표 인식 Paper

이미지 Scanner

빠른 시행 가능

기존 수업 방식에 최소한의 변화

세밀한 학습자 정보 및 실시간 Feed back 등 구현에 제약



# 1 Data 수집

## [정의] 데이터의 종류

### Academic data (학습데이터)

교육적 속성을 포함한 데이터  
과목에 따라 속성이 달라짐  
예) 문항 A: 교양수학-연산-단답형-  
1학년1학기-미분

세밀한 학습적 분석이 가능  
학습자 약점 분석 및 맞춤형 진단에 용이  
상당한 수준의 초기 투자 필요

### Behavior data (행동데이터)

학습자의 행동 습성을 포함한 데이터  
과목과 상관없이 수집 가능  
예) 문항 A: 풀이횟수-정오여부-풀이시간

학습자의 습관을 기반으로 분석  
초기 실행 용이 하고 전공 및 학년 등  
커리큘럼에 따른 제약 없음  
특정 교과목 강.약점 분석에 제약 존재



# 1 Data 수집

## [구분] Meta data의 Tagging

### Manual tagging

과정별 Skills frame work  
Contents에 초기 Meta data 부여  
서비스를 운영하면서 고도화

과목별 별도 작업  
초기에 많은 노력이 필요함  
세밀한 분석과 측정이 가능

### Auto tagging

초기 High level Meta data 혹은  
none meta data  
서비스를 운영하면서 학습자의  
반응을 귀납적으로 분석하여 진행

빠른 시행 가능  
AI 기술의 발전에 따라 고도화 가능  
특정 기준에 부합하는 성취도 개선에  
제약



# Case 1

## [Academic data 중심]



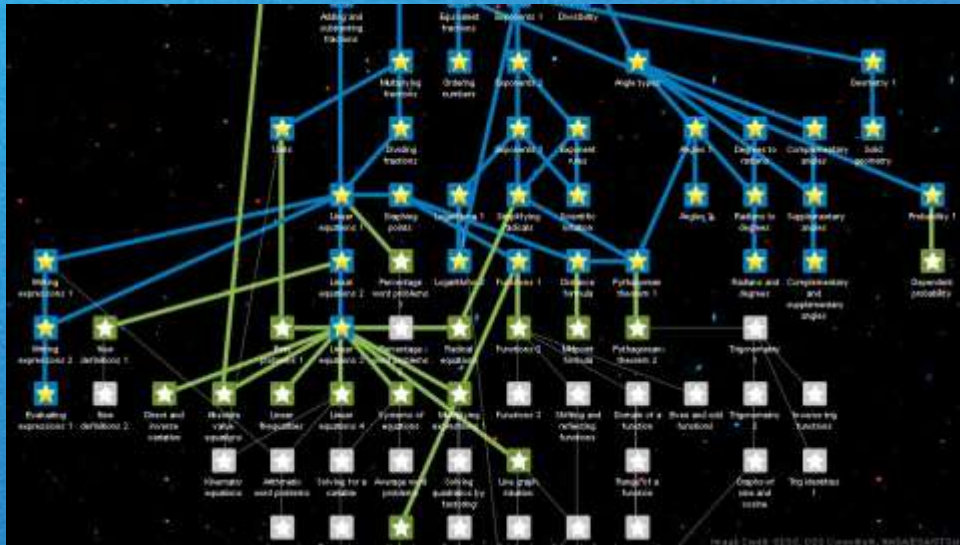
# Project 진행 방식





# Skills frame work

## 개념 중심 Frame work



이론적 기준 map을 활용  
개념간의 상관 관계를 통해, 맞춤형  
교육을 실시하면서 문제 지점을  
발견

## 역량 중심 Frame work

수로 쓰시오.

(1) 팔백

(2) 칠백오십

위어 세는 규칙을 찾아 빈칸에 알맞은 수를 써넣으시오.

(1) 180 → 280 →  →  → 580

(2)  → 760 → 790 →  → 810

□ 안에 알맞은 수를 써넣으시오.

(1) 60보다 40 큰 수는 입니다.

(2) 80보다  큰 수는 100입니다.

지수가 십구십에서 100이하의 정수를 지려고 합니다. 100이하의 정수 중에, 10이하의 정수의 2배도 될 수 있는 정수는 모두 몇 개입니까?  
(1) 지수가 지려고 하는 정수는 모두 몇 개입니까?  개

(2) 지수가 지려고 하는 정수는 모두 몇 개입니까?  개

Mathematics

Number Recognition & Place Value

Number Sequence, Counting Sets

Addition & Subtraction

Problem Solving

문제를 풀기 위해 해당되는 역량을  
중심으로 high level mapping  
서비스를 제공하면서 상관관계를  
고도화



# Skills frame work (Early learning)



## LANGUAGE & LITERACY

### Language Development

- Receptive Language
- Expressive Language

### Alphabet Knowledge

- Letter Recognition & Alphabet Sequence
- Letter Sound Production

### Phonological Awareness

- Phoneme Recognition
- Blending & Segmentation

### Phonics & Word Recognition

- Decoding
- Spelling
- Word Recognition

### Vocabulary

- Word Acquisition
- Word Relationships

### Reading

- Reading Aloud
- Reading Comprehension

### Writing & Print Conventions

- Book & Print Familiarity
- Narrative & Expository Writing
- Grammar
- Punctuation & Capitalization



## MATHEMATICS

### Numbers & Counting

- Number Recognition & Place Value
- Number Sequence
- Counting Sets
- Comparing Quantities
- Addition & Subtraction
- Representing Operations

### Measurement & Data

- Making Measurements
- Telling Time
- Estimating
- Data Collection & Representation
- Sorting & Classifying

### Geometry & Spatial Sense

- Shape Recognition & Attributes
- Shape Manipulation
- Composition & Decomposition
- Position, Location, & Direction



## SCIENCE

### Health

- Health & Hygiene Practices
- Health & Wellness Knowledge

### Natural Science

- Human Body
- Plants & Animals
- Weather & Seasons

### Physical Science

- Forces & Motion
- Properties of Materials & Objects

### Environmental Science

- Natural Resources
- Physical Environments

### Scientific Skills & Method

- Scientific Method
- Engineering & Technology



## PHYSICAL DEVELOPMENT

### Sensory Skills

- Auditory Skills
- Visual Skills
- Tactile Skills

### Motor Skills

- Fine Motor Skills
- Gross Motor Skills



## THINKING SKILLS

### Reasoning & Problem Solving

- Patterns
- Deductive Reasoning
- Analogical Reasoning
- Perspective Taking

### Executive Function Skills

- Working Memory
- Persistence
- Flexible Thinking
- Planning



## SOCIAL-EMOTIONAL SKILLS

### Interacting with Others

- Friendship & Play Skills
- Understanding Others' Emotions

### Understanding the Self

- Expressing & Managing Emotions
- Confidence & Initiative
- Regulating Behaviors



## CREATIVE ARTS

### Music & The Arts

- Music
- Drawing & Painting

### Imaginative Play & Storytelling

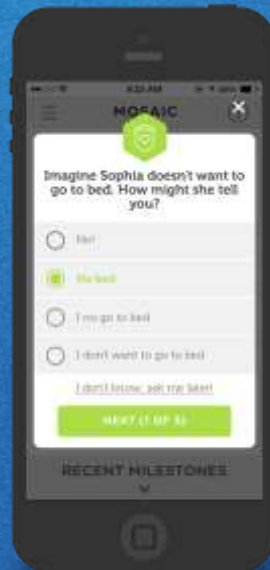
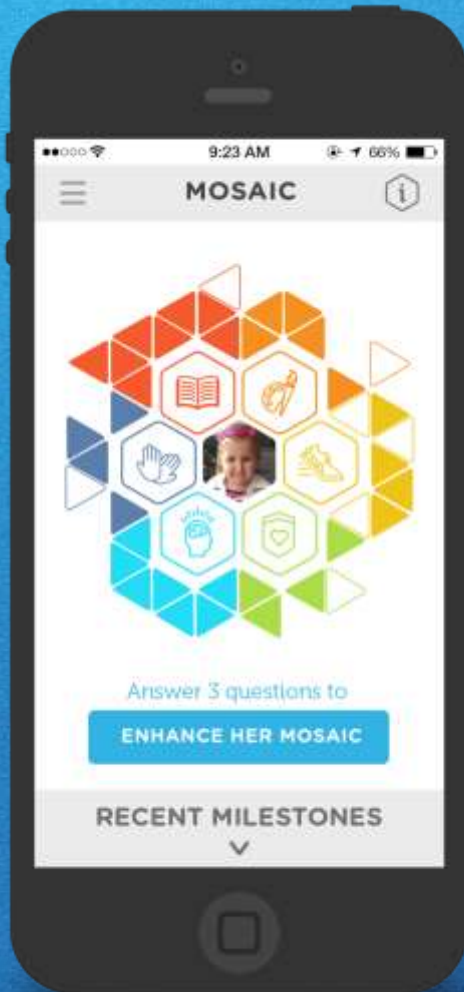
- Pretend Play
- Storytelling



# Frame work 활용 사례 (Learner's Mosaic)

시계열에 걸쳐, 동일 기준으로의 성취도를 지속적으로 측정 가능  
Data pipe lines를 확장하면 더 정교하고 종합적 측정 가능

Frame work의 변경 등 유연성이 부족함  
신규 콘텐츠 추가 시 추가 작업 지속 필요





# Case 2

## [Behavior data 중심]



# Project 진행 방식





# Event의 정의

수집, 활용하는 정보의 최소 단위

동일한 현상도 Event의 정교함에 따라 분석의 깊이가 매우 달라짐

동일한 교과서를 100쪽 중 80쪽까지 읽은 두 학생

## 학생 A

Page 당 소요 시간이 점점  
길어지다가 80 page에서 멈춤

책이 지루함 (맞지 않음)

## 학생 B

동일한 속도로 읽다가 78 page – 80  
page를 반복해서 계속 읽음

해당 Page에 관심이 있음 (혹은)  
모르는 내용이 있음



# Case in depth (W社)

[Behavior data 중심]



# Case [Summary]



# W社 개요

- 국내 1위의 Tablet based 교육 기업
- 50만 학습자 + 17만 Daily active user
- 6개월간 축적 Events 111억개
- 기존 콘텐츠의 활용을 위해, Behavior data 중심으로 분석 서비스 론칭 (2018.02)

과목 및 커리큘럼에 무관한 행동 데이터를 전과목에서 수집



# W社 Events

- 학습 시작/종료 : 무엇, 시작/종료 시점
- 문항 풀이: 문항 정보, 시도 횟수, 정오답 여부, 입력한 답안, 채점여부, 풀이소요시간
- 강의관련: 영상 정보, 학습 시간, 마지막 학습 위치 등
- 힌트관련: 관련 문항 정보, 힌트 종류, 학습 시간 등

과목 및 커리큘럼에 무관한 행동 데이터를 전과목에서 수집



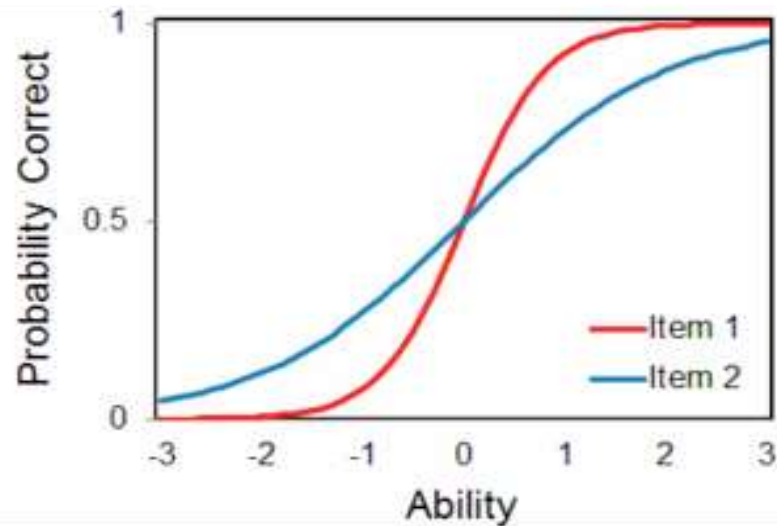
# W사 Event (분석 체계)





# 적용 이론

IRT(Item Response Theory)

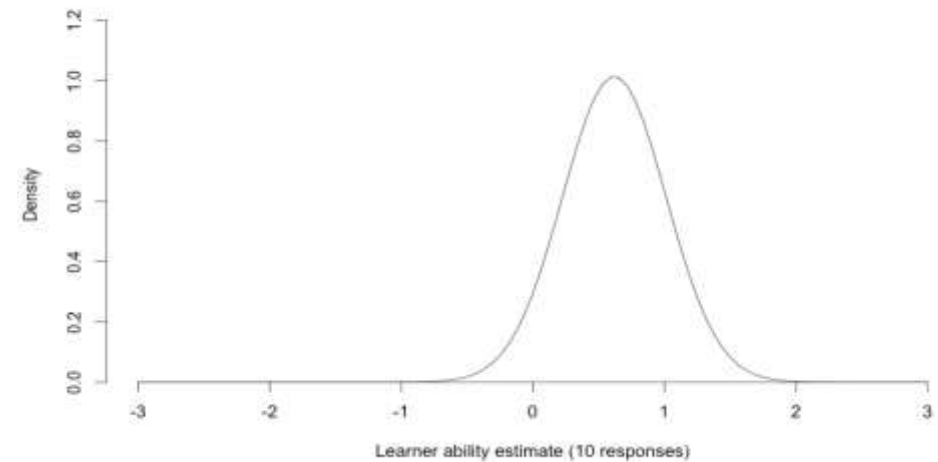


$$P(Y_{ik} = 1) = \Phi(-b_k + a_k \theta_i)$$

Annotations for the equation:

- $P(Y_{ik} = 1)$ : P(k문항에 맞게 답할 학습자 i의 확률 값)
- $b_k$ : k문항의 난이도
- $a_k$ : k문항의 변별력
- $\theta_i$ : i학습자의 능력치

Bayesian model



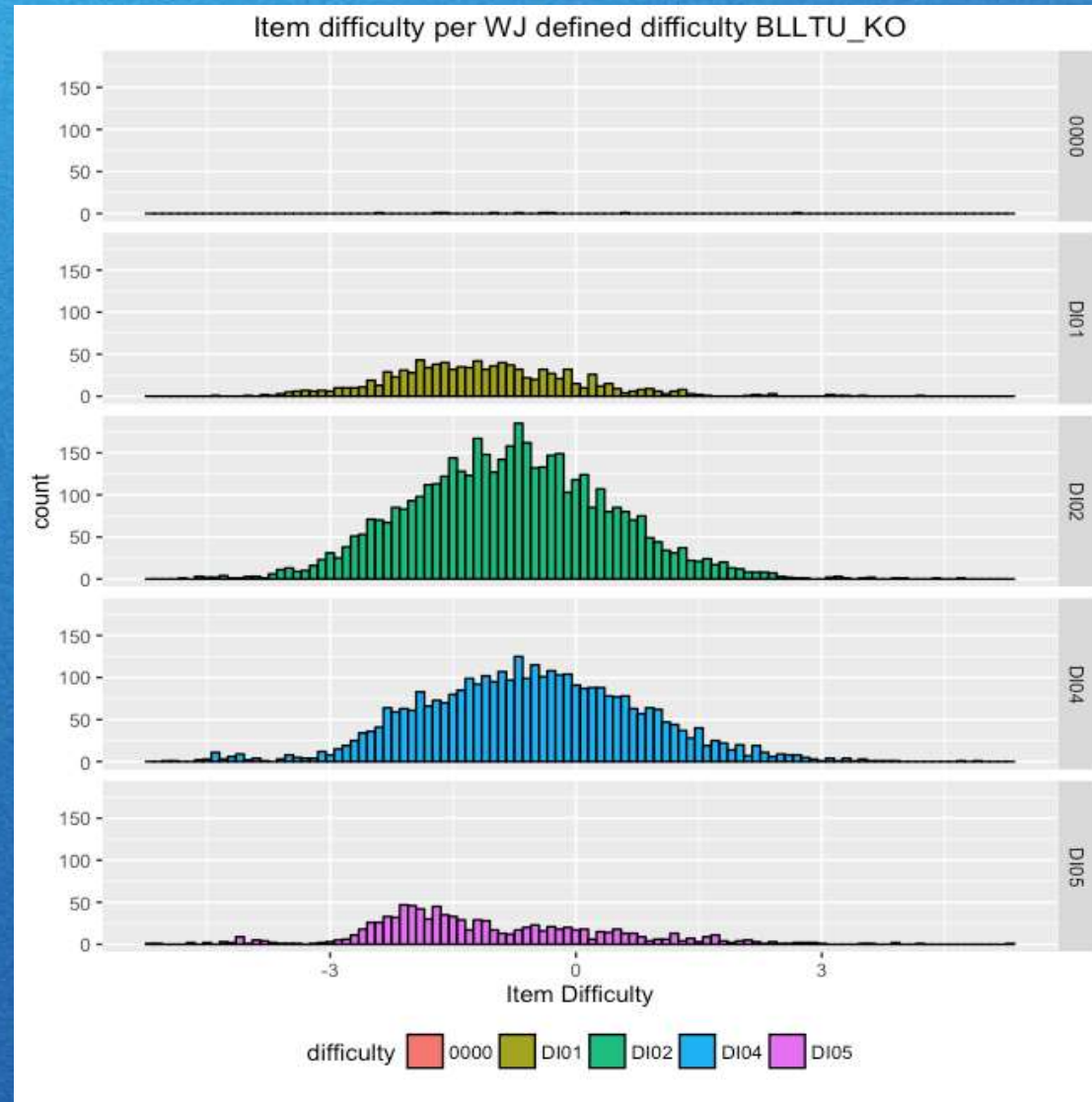
지속적인 측정과 Peer group을 대상으로 한 Machine learning 분석을 활용, 학습자의 능력치에 대한 정확도를 향상 시킴



# Case [사전 작업]

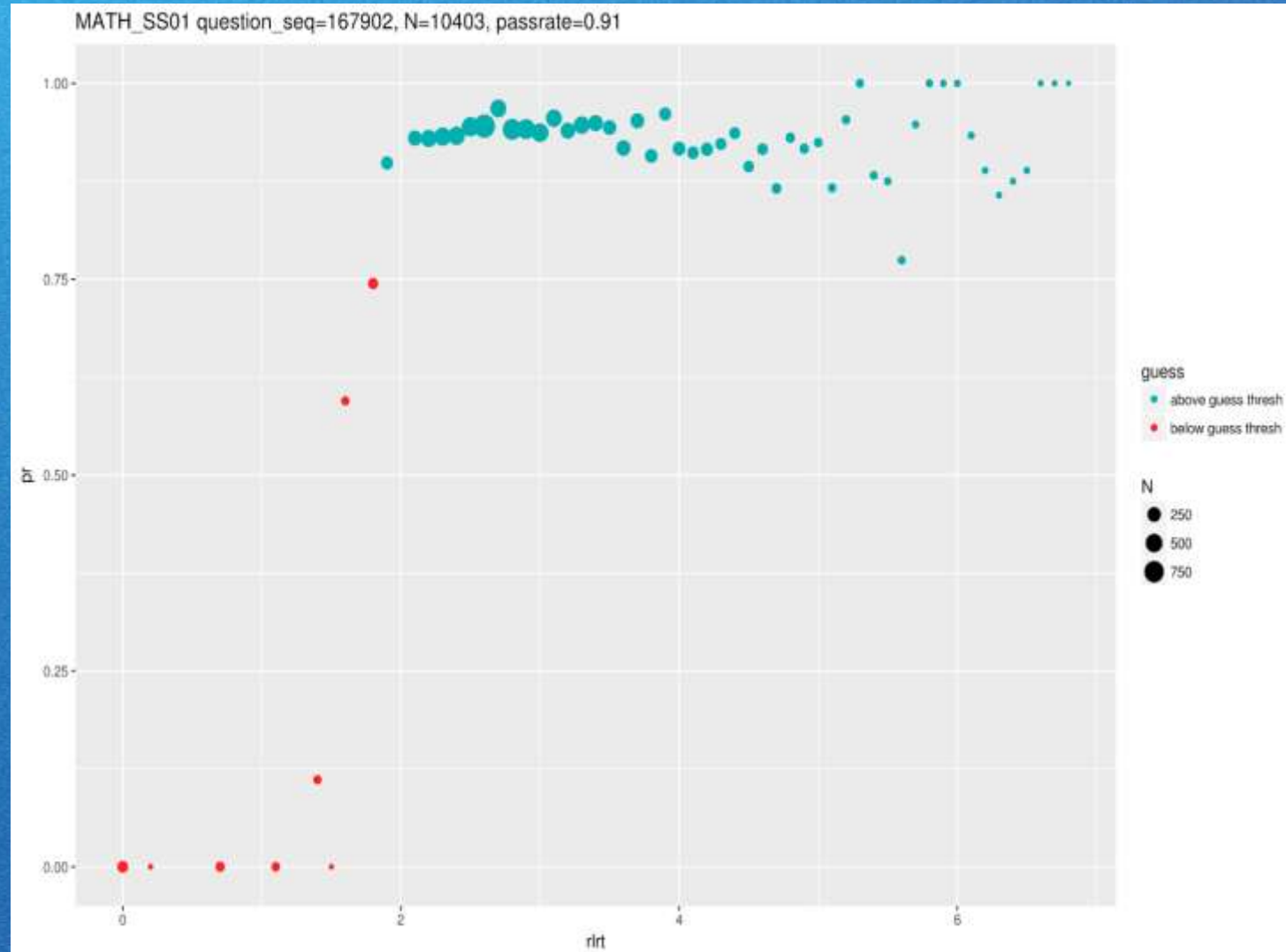


# 난이도 부여



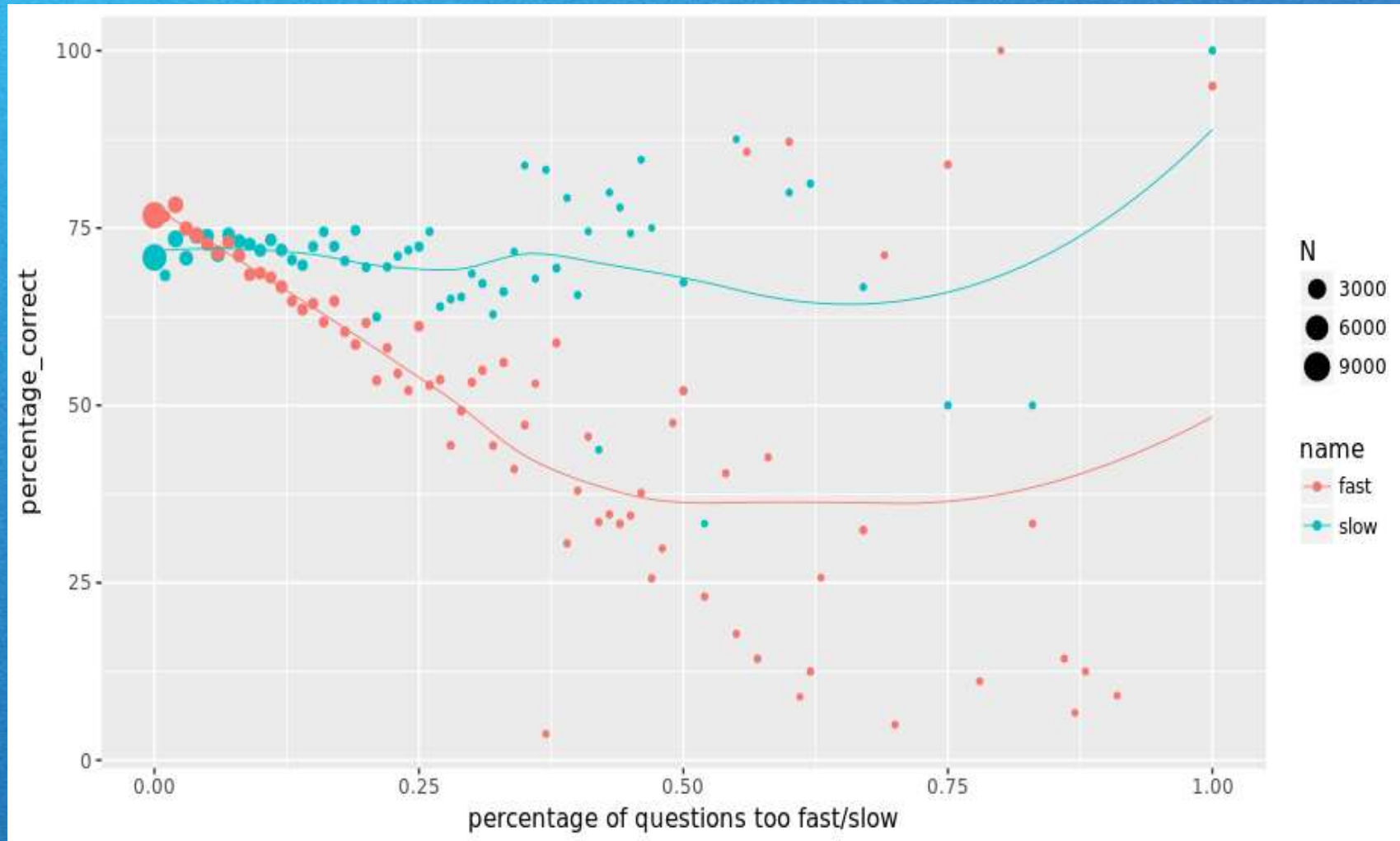


# 시간-정답률 상관관계 분석 1/1





# 시간 - 정답률 상관관계 2/2



X축: 응답시간 너무 빠름/너무 느림으로 나타난 문항의 비율  
Y축: 정답률

파란 선: 느림 / 붉은 선: 빠름



Case [서비스; USER]



## 틀린 문제 분석

다원 학생의 이번 테스트에서 틀린 13 문제에 대한 분석입니다.

맞힐 수 있지만 틀린 문제



몰라서 틀린 문제



찍어서 맞은것으로 예상되는 문제



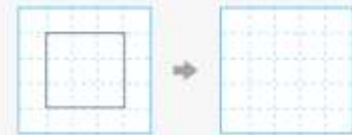
## 틀린 문제 분석 (상세)

← 맞힐 수 있지만 틀린 문제 (3)

Q1

이 문제는 난이도가 낮고 문제 풀이시간도 적정 풀이시간과 비슷했지만 결국 틀렸습니다. 혹시 문제 풀이 시 실수가 있었던 것은 아닌지 확인해 주세요.

다음 도형을 화살표 방향으로 밑었을때 생기는 모양을 그리세요.



다원학생의  
해답 난이도

쉬움

문제 난이도

중간

풀이시간

다원 학생

작성 풀이시간

3분 50초

4분



## 행동분석 1

다원 학생의 이번 테스트 중 **12개의 문항에서 고쳐야할 습관**을 발견했습니다.

특히, 문제 풀이에 2회 이상 틀리고도 해설 강의를 보지 않은 문제 8개였습니다.



## 행동분석 2

첫 2주 동안 다원 학생은 틀린 문제에 대해 힌트를 확인하지 않았습니다. 하지만 가장 최근 2개의 시험에서는 틀린 문제의 80%에 대해 힌트를 확인했습니다. 다원 학생의 개선된 학습 태도를 칭찬해 주세요!





## 개인별 난이도 분석

다원 학생이 다음 테스트의 15,19번 문제에서 어려움을 느낄 것으로 예측됩니다. 문제를 풀 때, 신중하게 풀 수 있도록 지도해 주세요.



## 시험대비 점수 예측

다원 학생의 최근에 본 3개의 시험결과를 분석하여 예측한 주니어 토플 점수는 845점입니다.

다원 학생은 Reading Comprehension 영역에서 쉬운 문제를 빨리 풀어 틀리는 실수를 하는 경향이 있으니 실제 시험에서는 이점에 주의하여 차분하게 시험에 임하도록 지도해주세요.

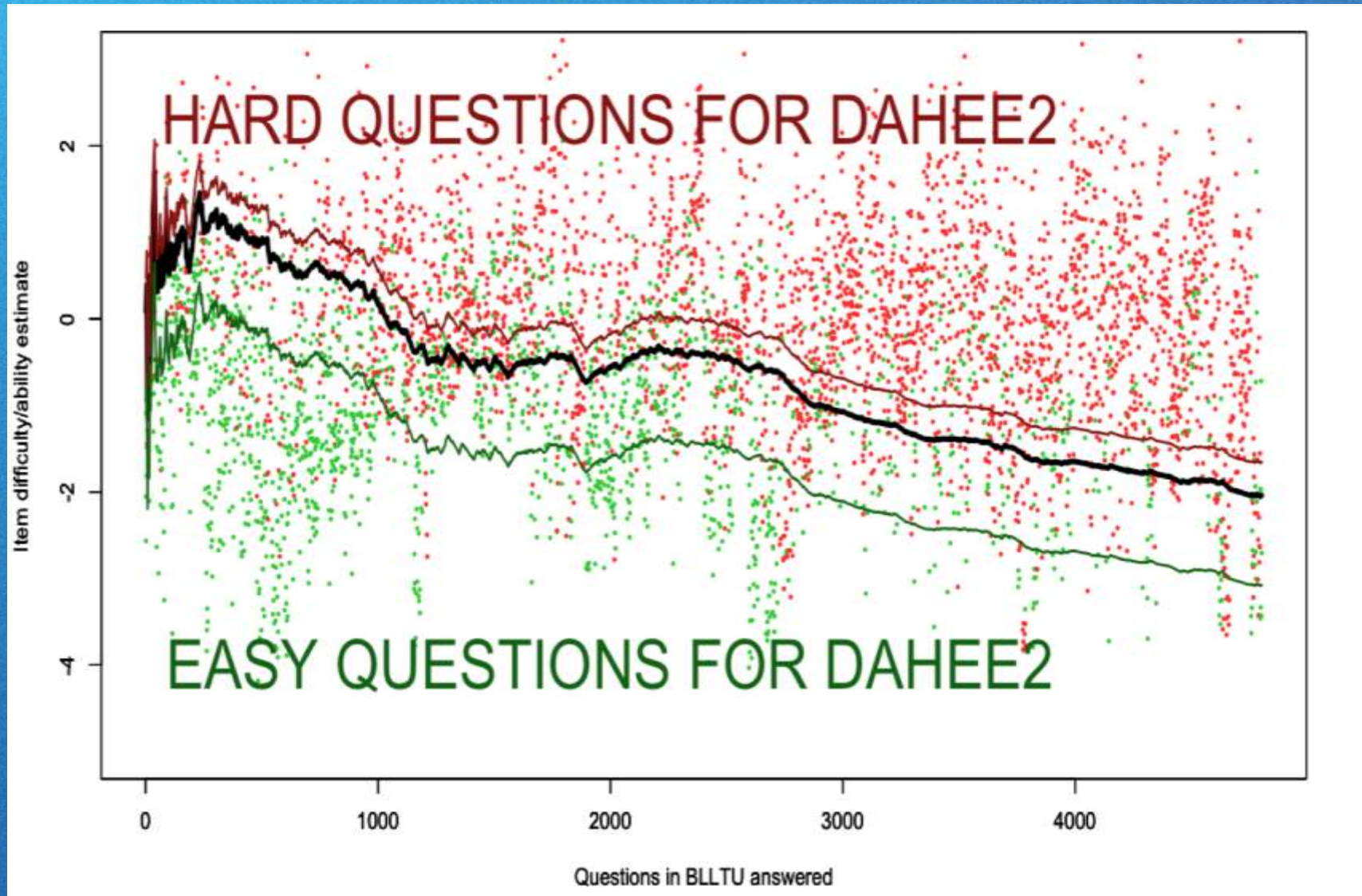




# Case [서비스; 관리자]

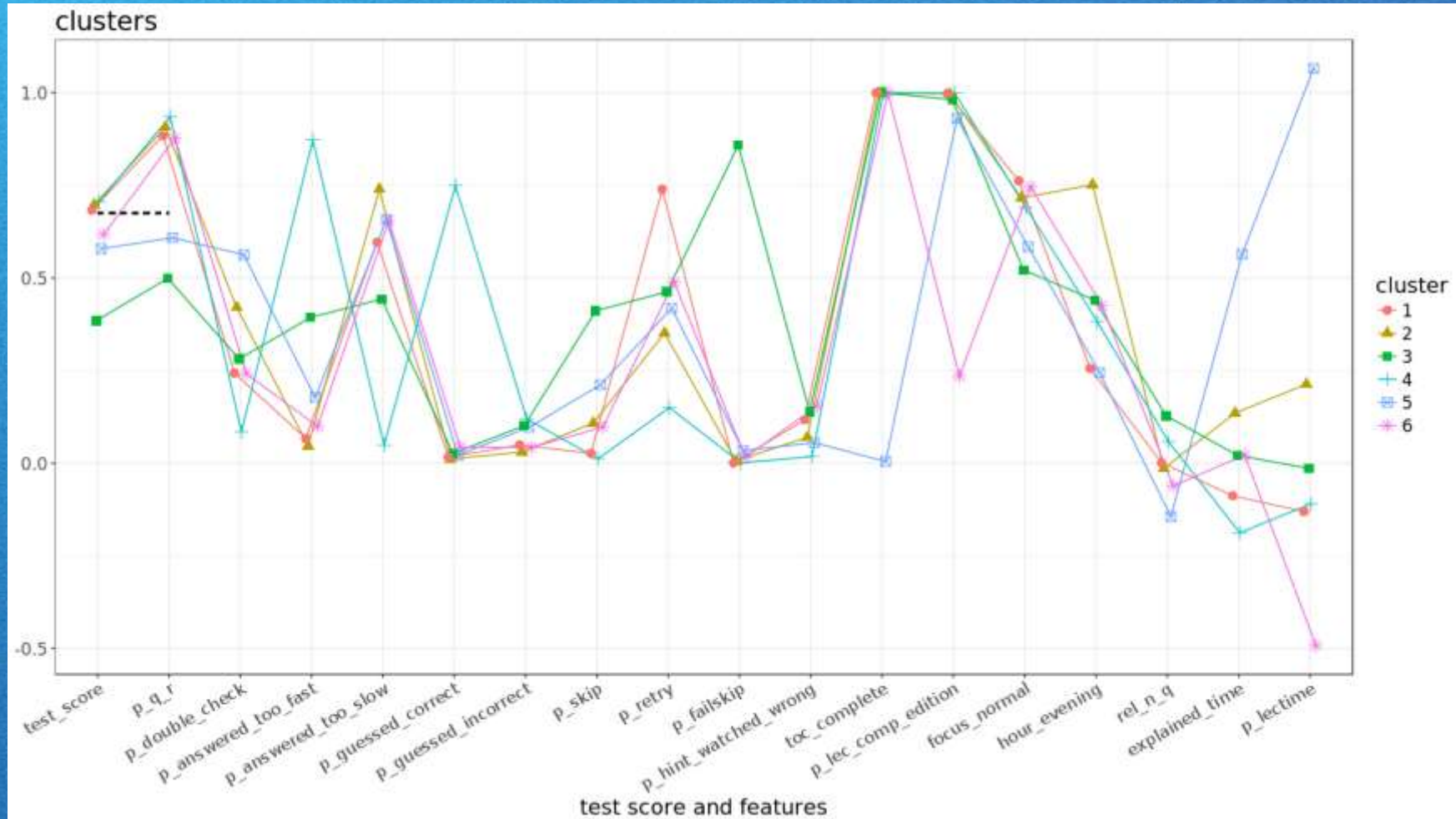


# 역량 Tracking





# Learner clustering



행동 특성에 따라 학습자를 8개 클러스터로 구분



# Leaner clustering

클러스터/ 퍼포먼스	정의 & 기준	각클러스터로 분류되는 %	상	중	하
CL1	<b>연습문항 점수가 좋지 않으나 열심히함</b> 강의(>=.7)와 연습 목차(>=.9)를 대부분 완료함, 적정시간에 응답함, 틀린 문항을 상대적 으로 많이 재시도함 (>=.25)	27.3%	31.3%	54.6%	14.1%
CL2	<b>연습문항 점수가 좋지 않으나, 대부분의 학습을 완료함, 하지만 재시도는 잘 안함</b> 연습문항 점수가 좋지 않음 (<.85) 강의(>=.7)와 연습 목차(>=.9)를 대부분 완료함,적정시간에 응답함, 틀린 문항을 상대적 으로 많이 재시도하지 않음	6.9%	31.1%	46.7%	22.2%
CL3	<b>강의 또는 연습문항 목차를 완료하지 않음</b> 연습문항 점수가 좋지 않음, 대부분의 강의 또는 연습문항 목차를 완료하지 않음, 적정시 간에 응답함	13.7%	27.2%	55.1%	17.7%
CL4	<b>대부분의 연습문항 목차를 완료하였고 좋은 연습문항 점수를 보임</b>	16.6%	61.0%	30.1%	8.9%
CL5	<b>호 내 강의들을 거의 완료하지 않음</b> 연습문항 점수가 좋지 않음, 강의 완료율이 .3을 넘지 않음, 너무 빠르게 풀지 않음	4.5%	38.9%	45.4%	15.6%
CL6	<b>느림</b> 연습문항 점수가 좋지 않음, 너무 빠르게 풀지 않음비교적 느리거나(>=.15) 또는 지나치 게 오래 집중함 (>=.25)	17.8%	33.8%	51.5%	14.7%
CL7	<b>지나치게 빨리 풀고, 한 가지의 고쳐야할 습관을 가짐</b> 연습문항 점수가 좋지 않음 (또는 좋으나 대부분의 연습문항 목차를 완료하지 않음) 적어도 40%의 문항들을 지나치게 빨리 답함	11.0%	11.4%	31.0%	57.6%
CL8	<b>지나치게 빨리 풀 &amp; 고쳐야할 습관 없음 (클러스터 7에 비해)</b> 연습문항 점수가 좋지 않음,적어도 40%의 문항들을 지나치게 빨리 답함. 지나치게 빨리 풀 문항 중, 25% 미만을 틀림, 재시도율 >= .25, 빠르게 건너뛴 비율 <= .3	2.3%	27.8%	50.4%	21.8%



# ALP Ecosystem



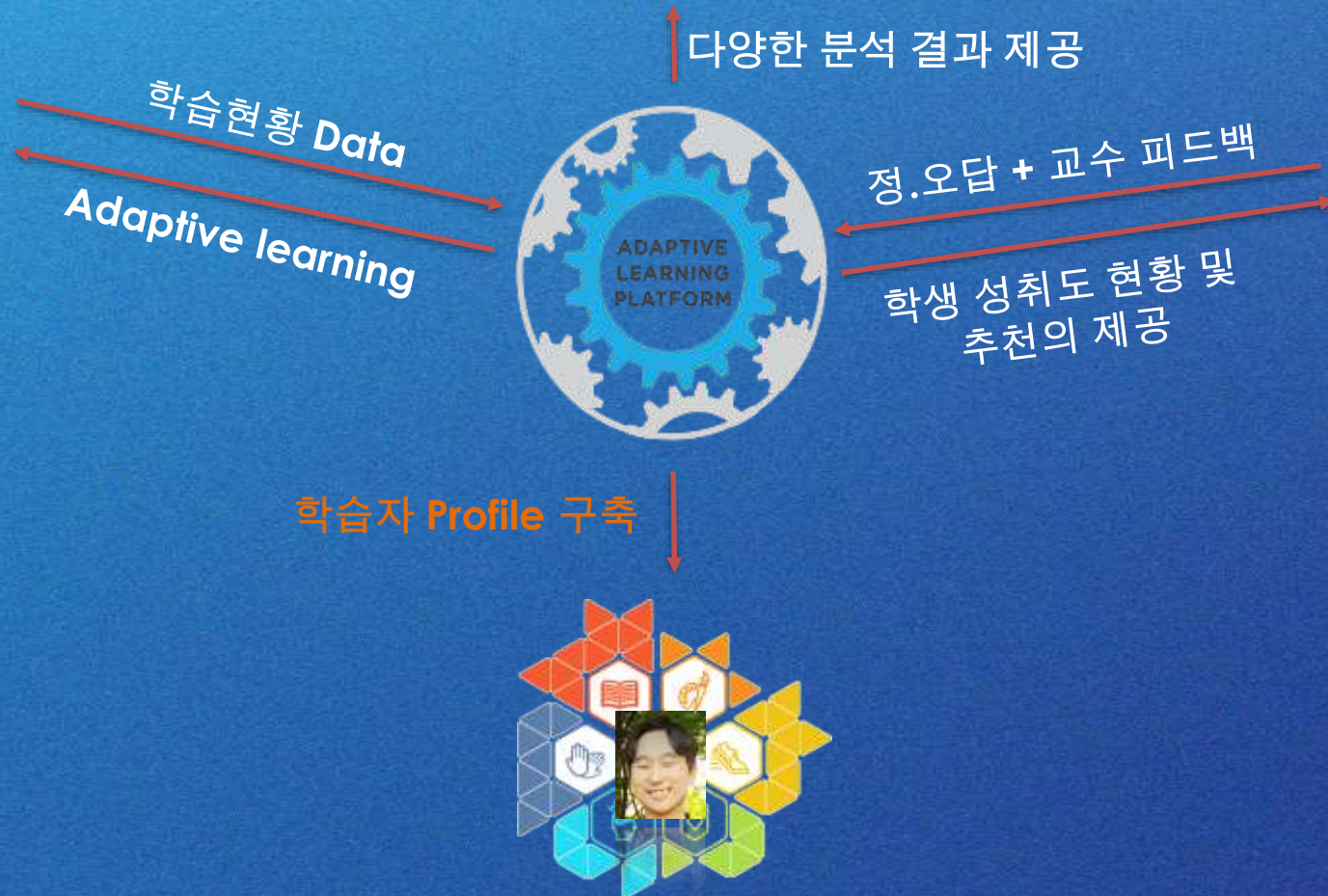
# ALP Ecosystem

## Self-learning



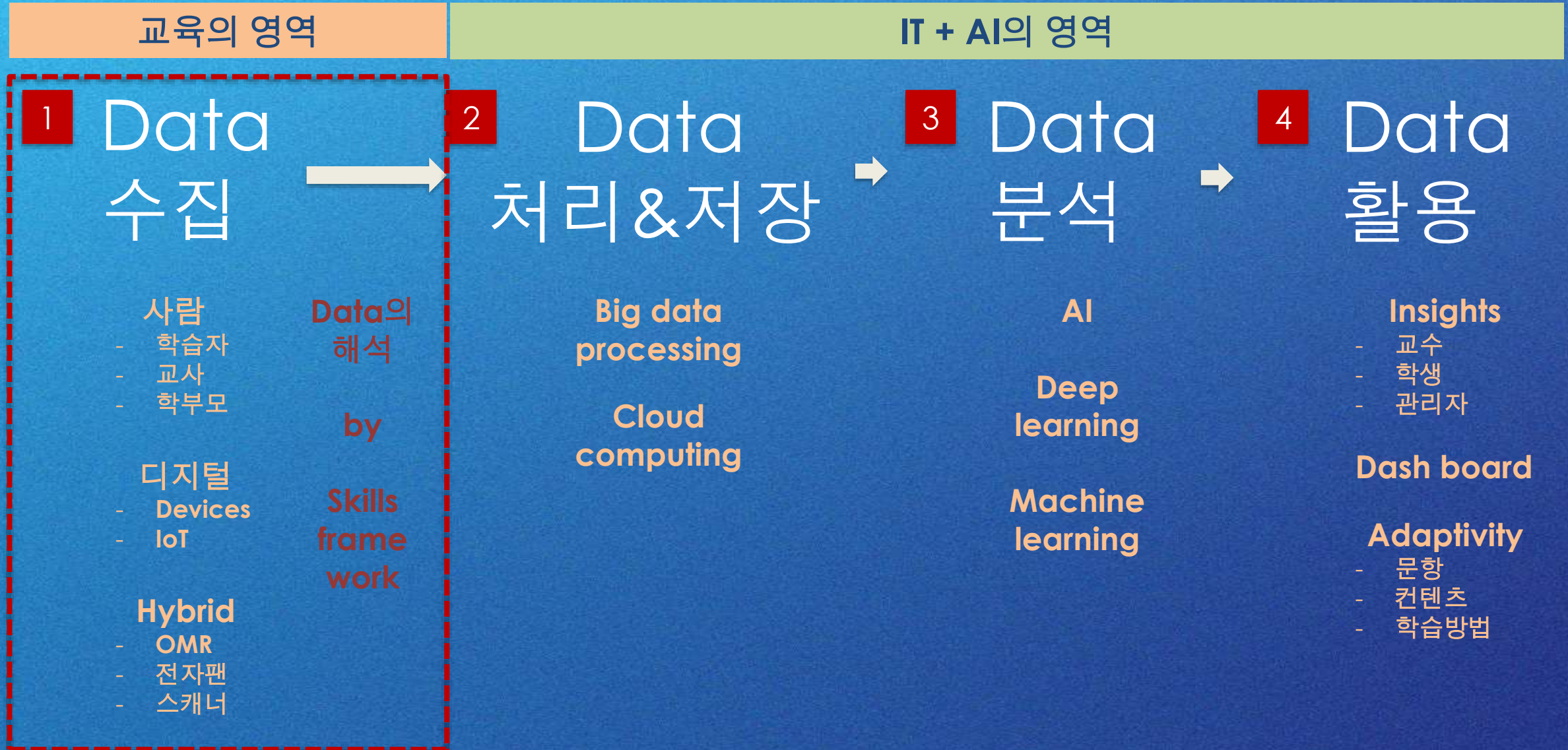
Dash board

## Classroom





# AI Ed-tech project process (Ed + Tech)





목표와 측정할 Data에 대한 정의  
의미와 가치가 있는 Data의 생성과 수집  
고도와 된 Tech의 활용

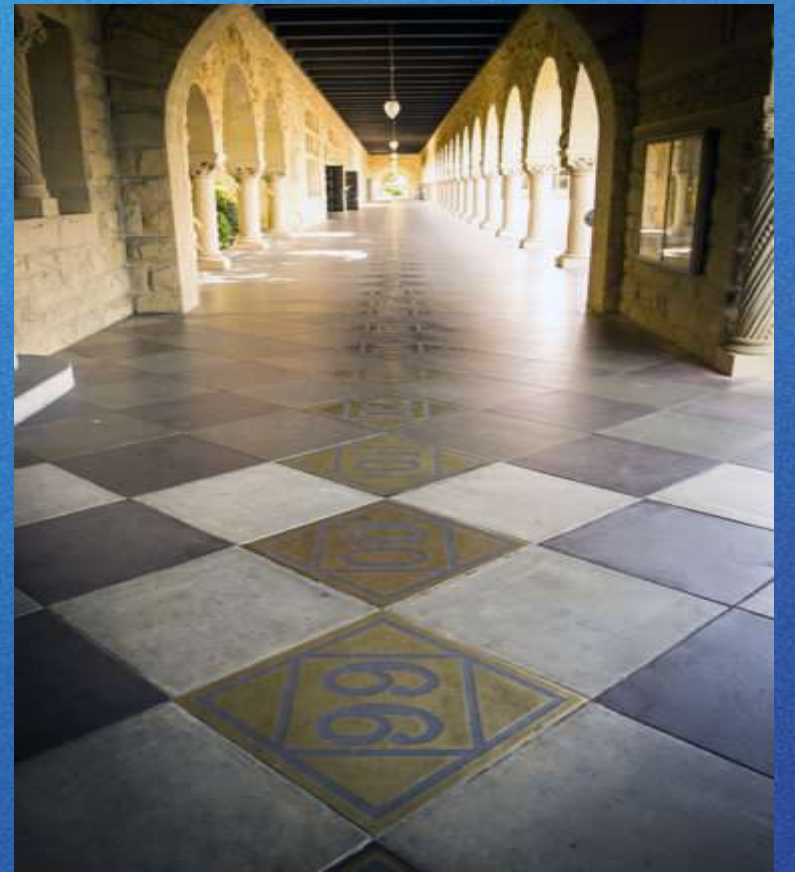
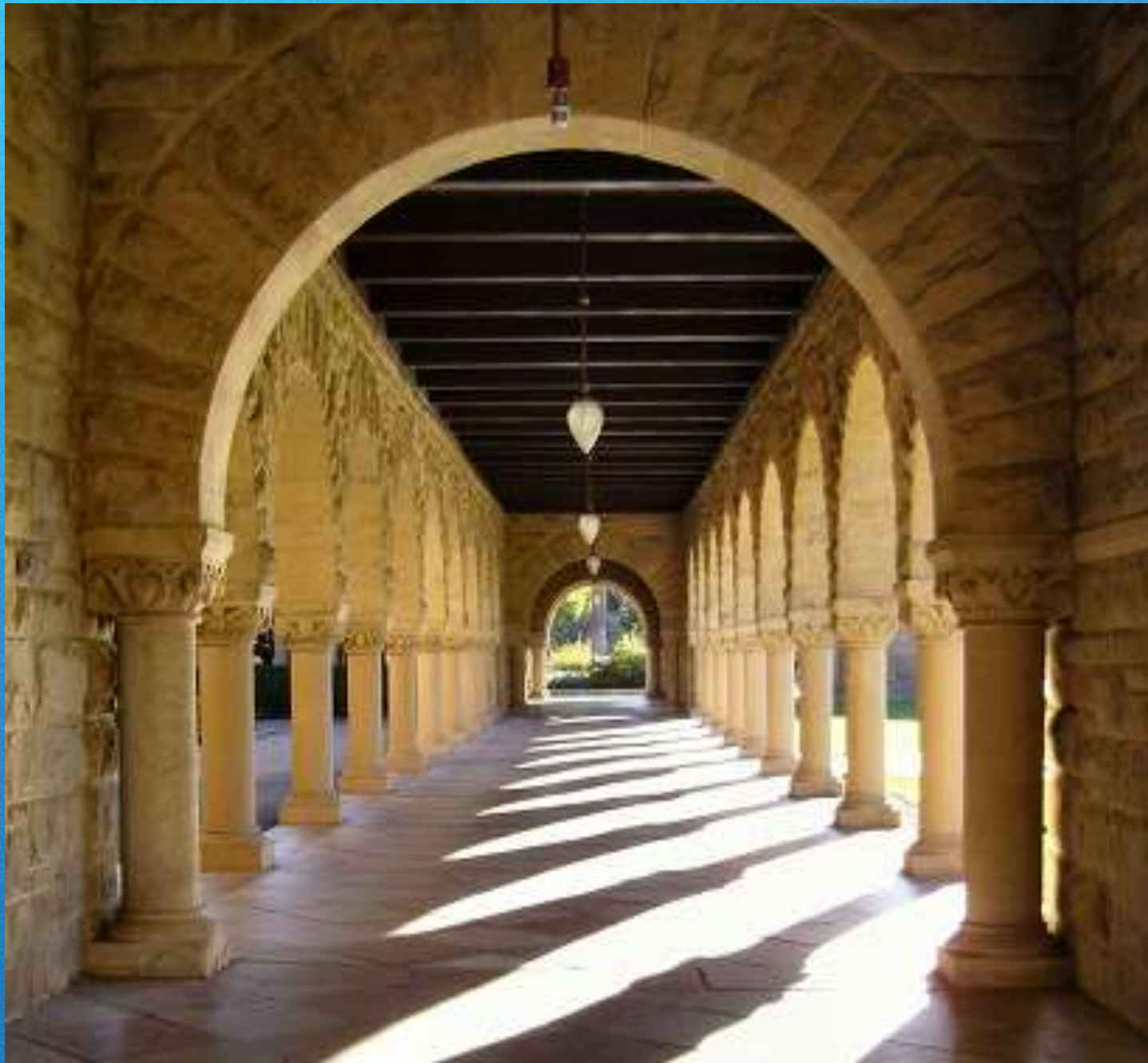
의지와 좋은 Data + AI = High quality service



# Clients









# Thank you

양지원 매니저 [jiwon.yang@kidaptive.com](mailto:jiwon.yang@kidaptive.com)  
정다희 팀장 [dahee.jung@kidaptive.com](mailto:dahee.jung@kidaptive.com)  
김민우 대표 [min.kim@kidaptive.com](mailto:min.kim@kidaptive.com)