



# 인공지능(AI)기반 비즈니스 프로세스 자동화 혁신 전략

---

- 한화그룹 혁신 사례를 중심으로 -

2018. 11. 08

한화시스템/ICT부문

# Agenda

## I. 시장 Trend 및 동향

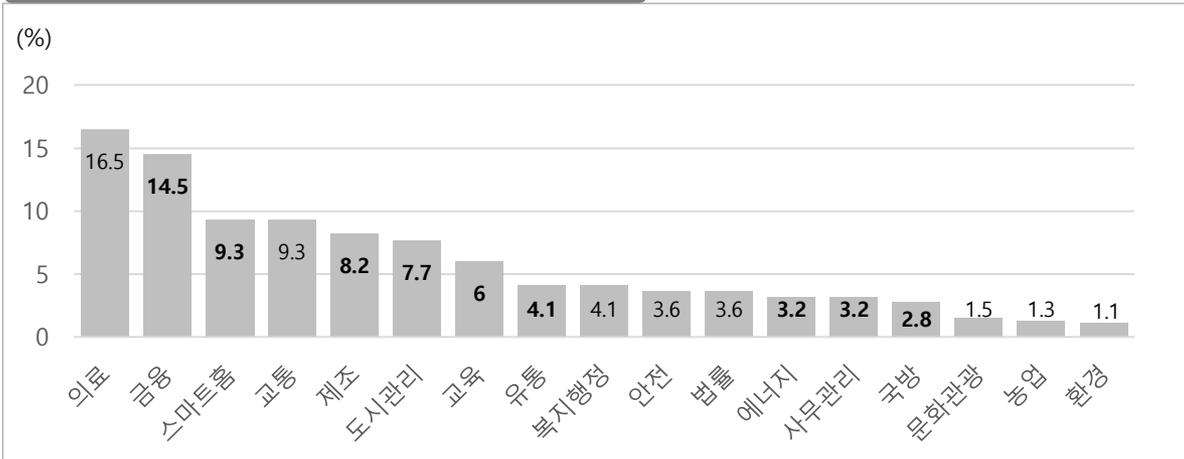
- AI 도입 현황
- AI 솔루션
- AI Driven Development
- AI 도입 Framework
- 선도사 동향

## II. 구축사례

## III. Lessons Learned

의료, 금융, 제조 중심으로 AI에 대한 투자가 확산되고 있으며, 산업별 AI수요 확대 및 관심도를 고려하여 차별화된 서비스 제공 등의 전략적인 시장 접근이 필요함

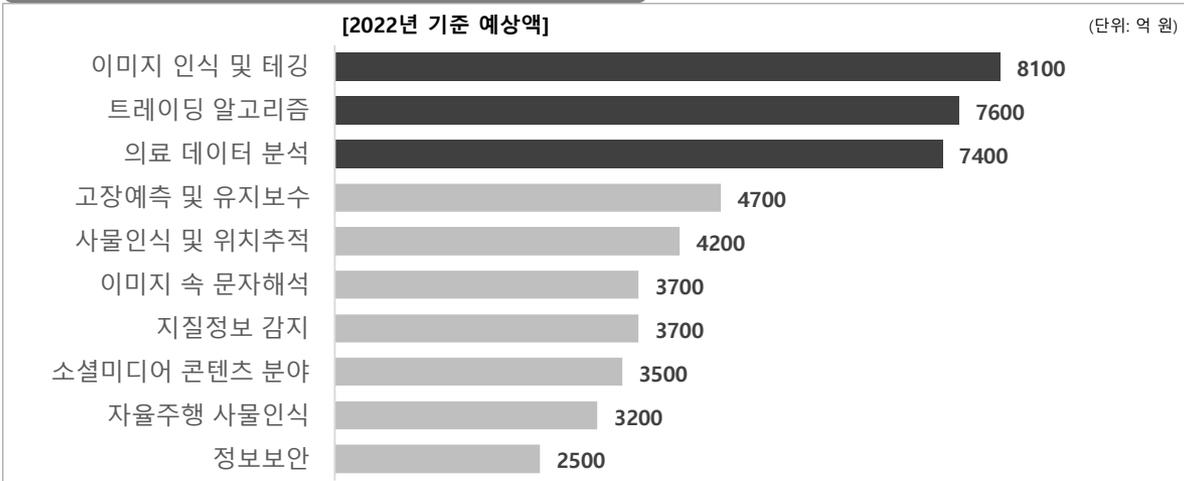
### 국내 산업별 도입현황



### 의료, 금융, 제조 중심 수요 확대

- 의료 (맞춤형 건강관리 서비스),  
금융 (고객맞춤 서비스, 금융시장 분석),  
스마트홈 (스마트 허브, 스마트 방법),  
교통 (자율주행, 교통혼잡 예측·제어),  
제조 (실시간 공정 제어 관리, 생산로봇)  
순으로 AI 도입 활발

### 기술 분야별 국내 시장 규모



### 인공지능 기술 연구 및 투자 확산

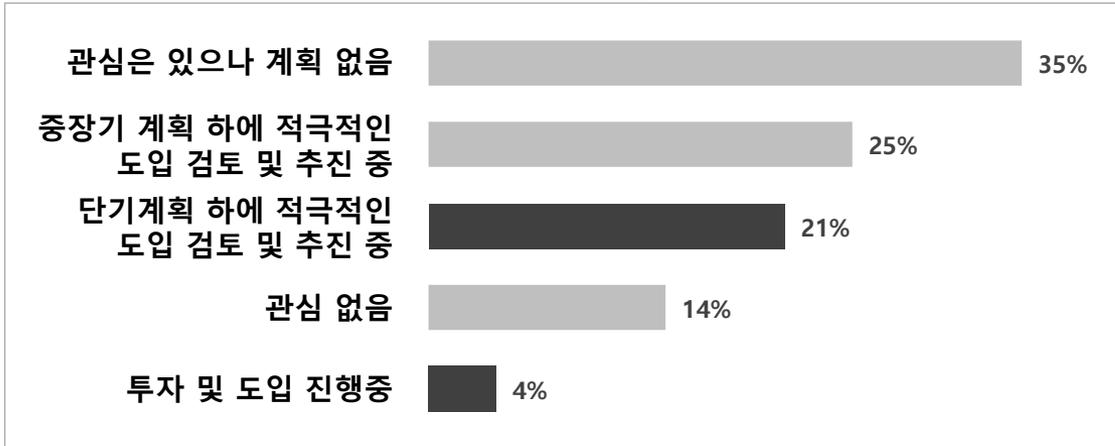
- 이미지 인식 및 태깅 분야는 약 8천 100억, 딥러닝 부문인 트레이딩 알고리즘은 약 7,600억에 달하는 등 시장 규모가 크게 증가할 전망

→ 딥러닝을 중심으로 인공지능 기술에 대한 투자 및 연구가 확산에 따른 대응 준비 필요

'18년 가트너 CIO Survey 결과, AI 기술 도입은 아직 초기단계이며, 업무 효율성 증가/사용자 경험개선/원가 절감 등의 이유로 도입을 추진함

### AI 도입 계획

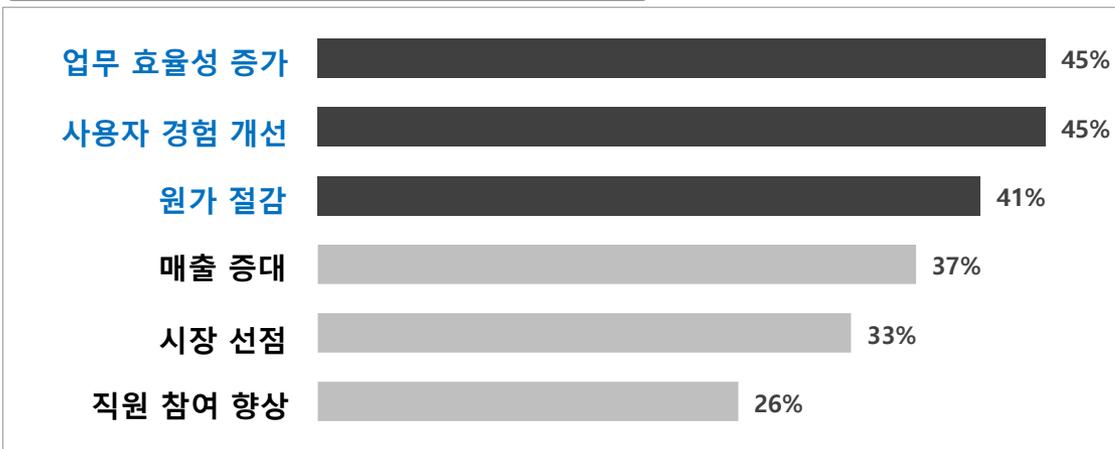
총 답변자 : 3,138 명



### AI 도입을 위한 검토 및 추진 단계

- AI에 대한 투자 및 도입을 진행 중이라고 답한 응답자는 4%이며, 21%는 도입을 추진 중이라 답함

### AI 도입 이유

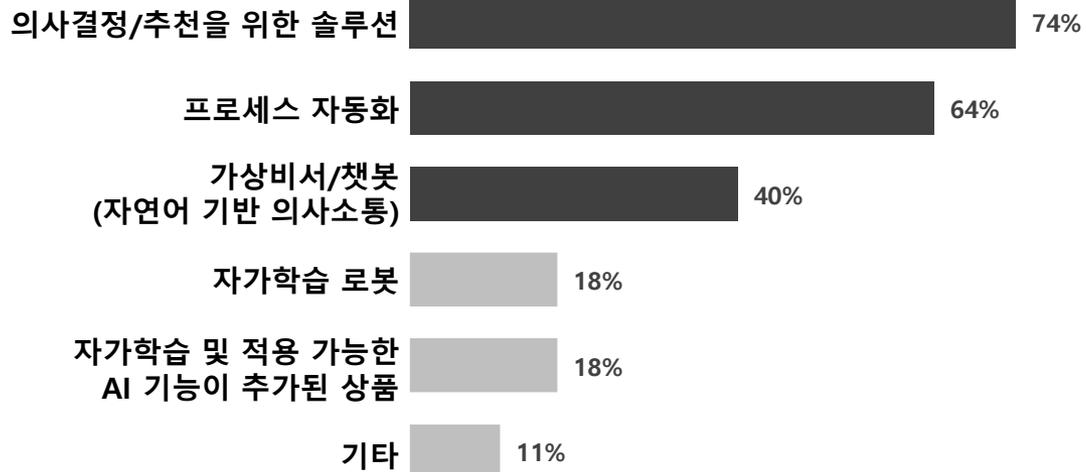


### 내부적 목적 달성

- 시장판도 변화나 경쟁사 대비 우위 확보 보다는 내부적인 업무 효율성 증가, 사용자 경험 개선, 자동화를 통한 원가 절감 목적이 큼

### AI를 이미 도입했거나 도입 예정인 회사의 AI 유형은 의사결정/프로세스 자동화/가상비서 등의 영역임

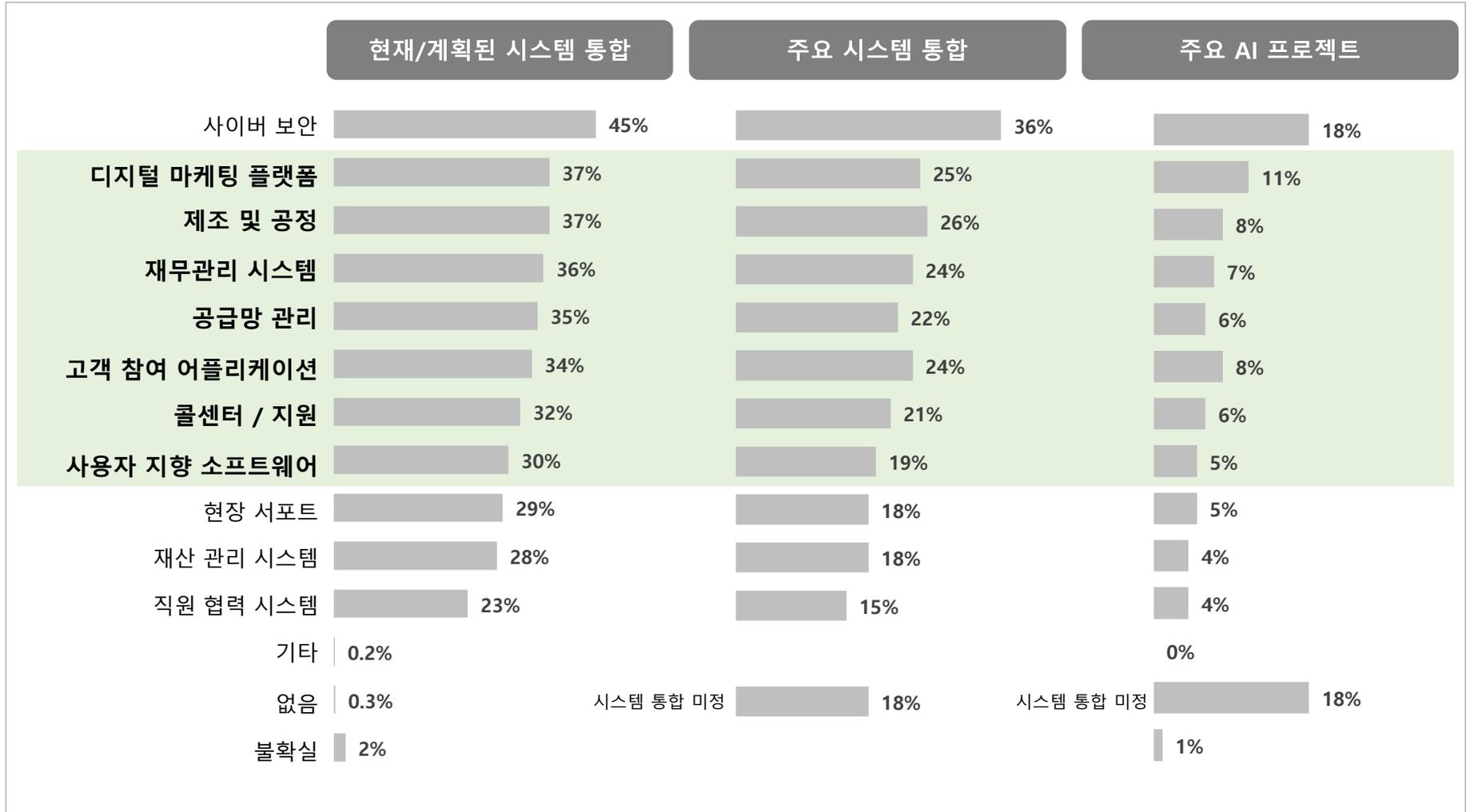
既 도입 / 도입 예정 AI 유형



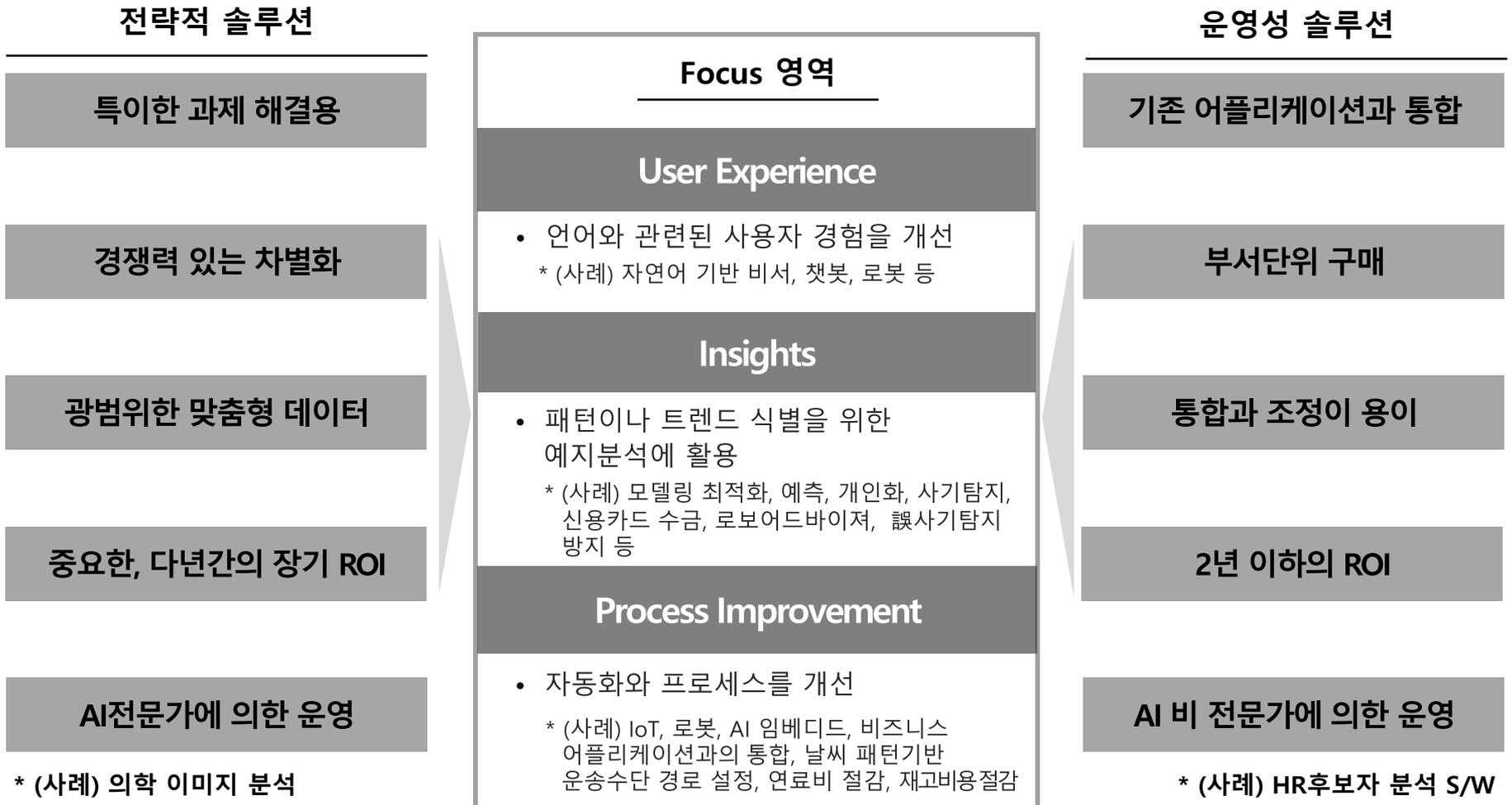
적극적인 기술 적용 추진 예정

- 의사결정/추천을 위한 솔루션, 프로세스 자동화, 가상 비서/챗봇 순으로 既도입 또는 도입 예정으로 나타남

디지털 마케팅 플랫폼, 제조/운영(생산), 재무관리, SCM, 고객 관련 어플리케이션, 콜센터 서비스/지원 어플리케이션은 AI와 통합되는 주요 업무 영역임

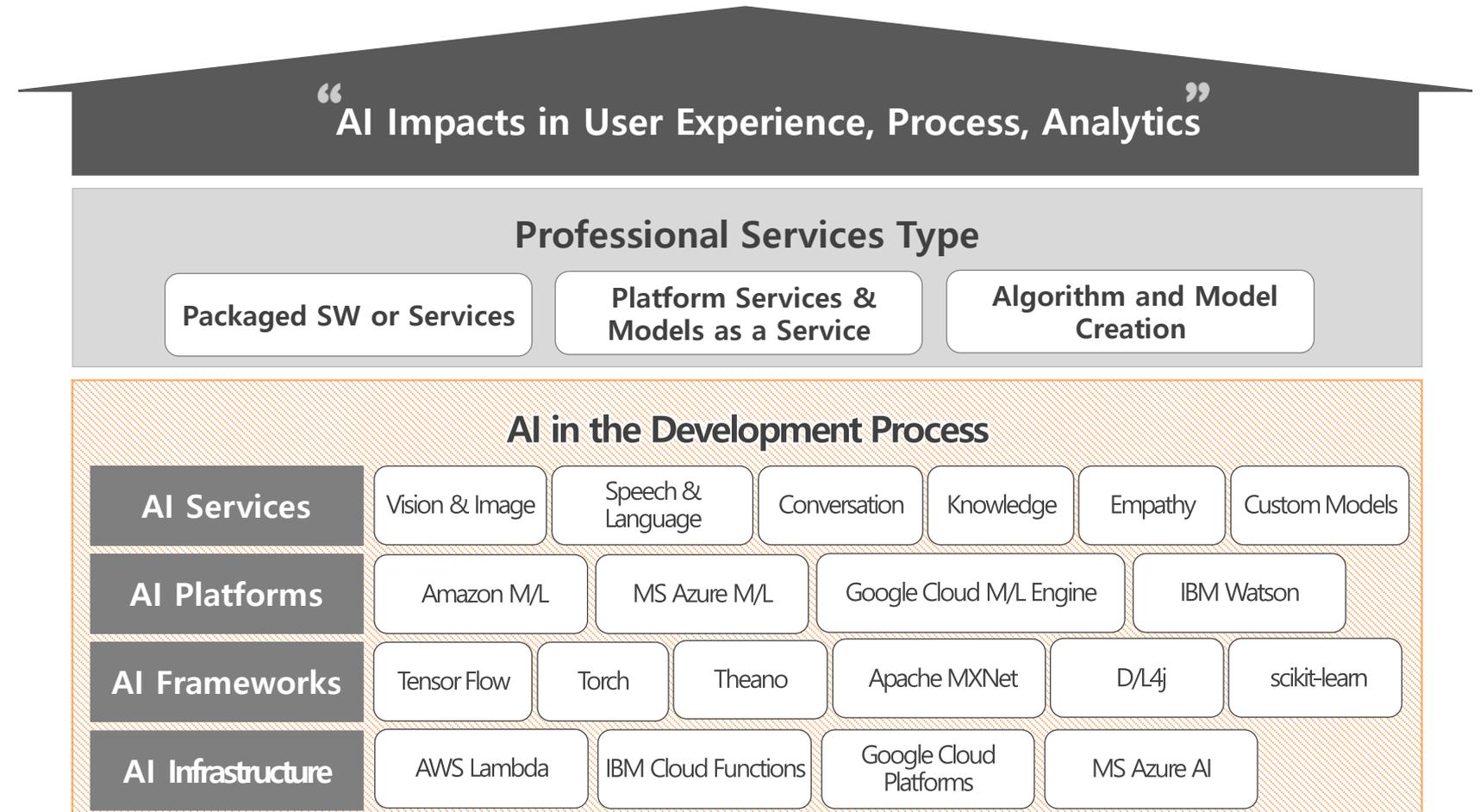


AI 솔루션은 전략적/운영성 솔루션으로 구분되며 User Experience, Insights, Process Improvement에 초점을 맞추고 있음



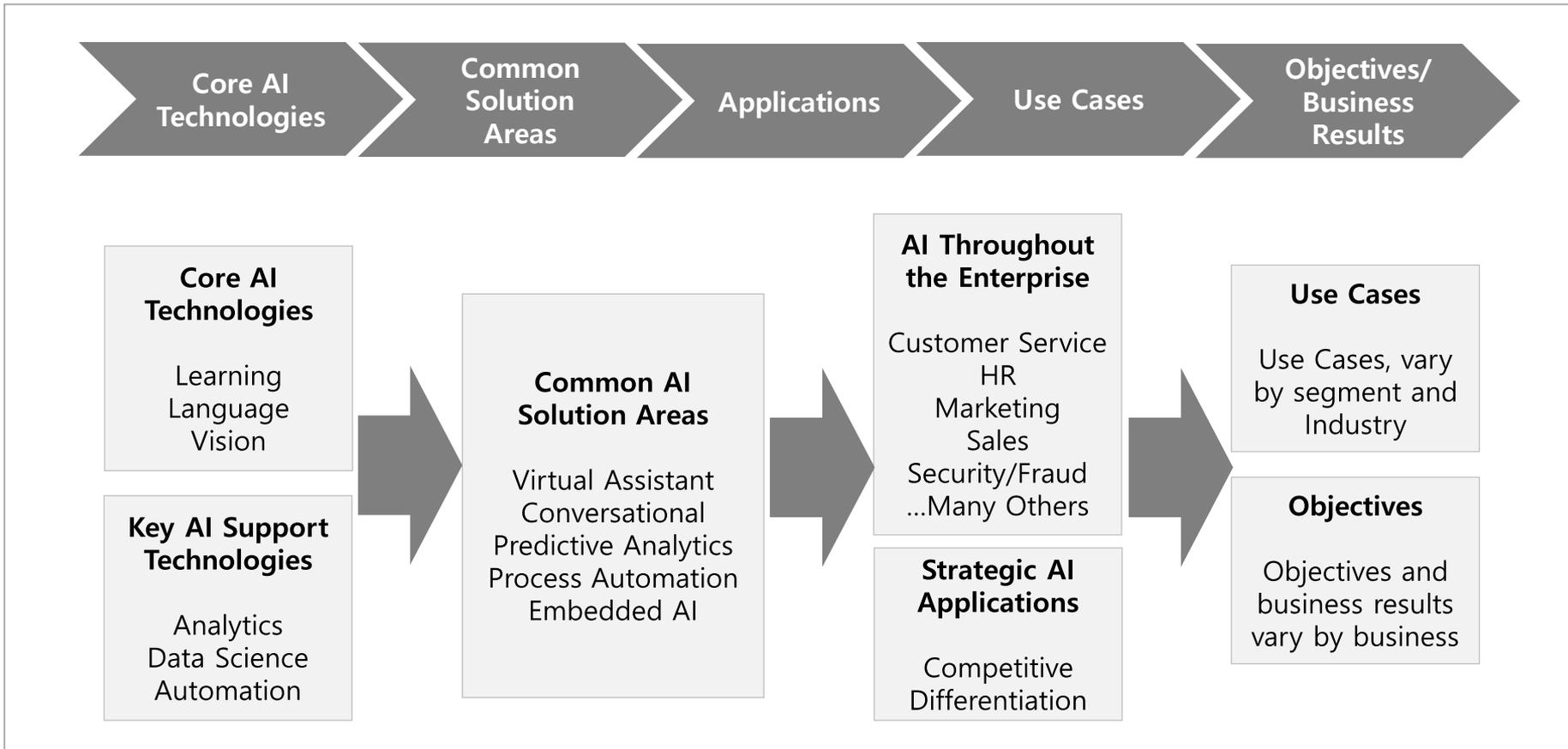
'22년까지 적어도 40%이상의 새로운 application 개발 프로젝트에 AI 개발이 함께 포함될 것으로 예상되며, AI를 활용한 개발 효과 향상에 대한 고민이 필요함

### 개념도



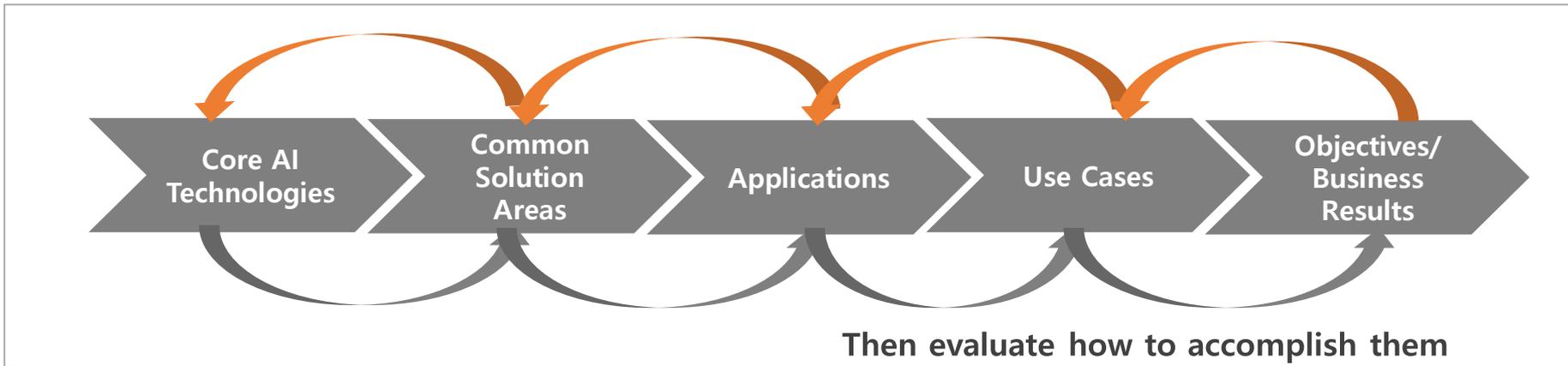
AI 도입 Framework은 기업에서 AI를 도입할 때 활용하는 방법으로, Core 기술부터 Common Solution, Application, Use Case, Objective로 구성됨

### AI 도입 Framework

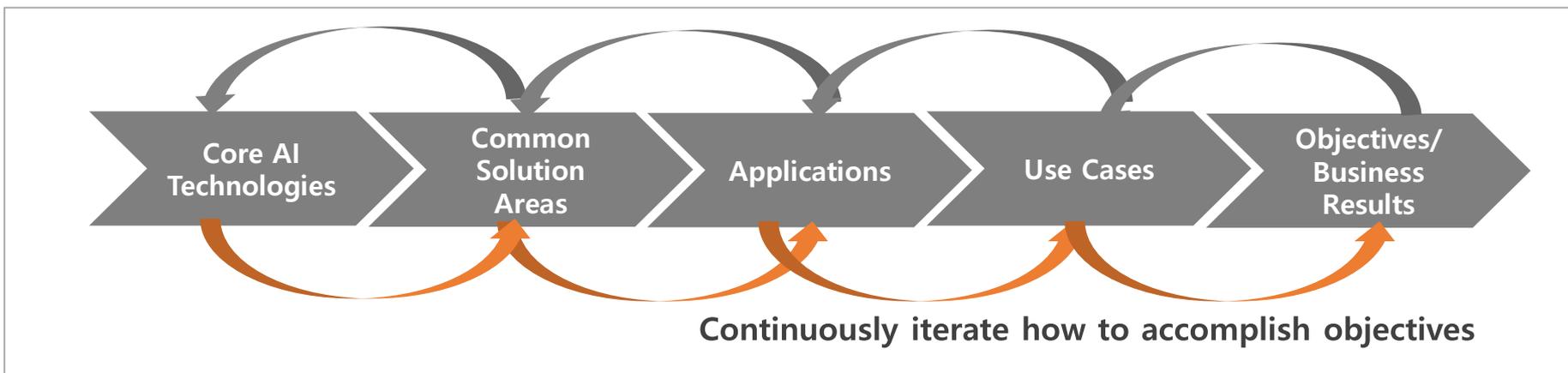


AI 도입 목적을 먼저 정의한 후, 그 목적에 필요한 AI 코어 기술, 솔루션, 어플리케이션을 검토하는 방법과, 구축 경험을 바탕으로 목적을 달성하는 방안을 지속적으로 반복해야 함

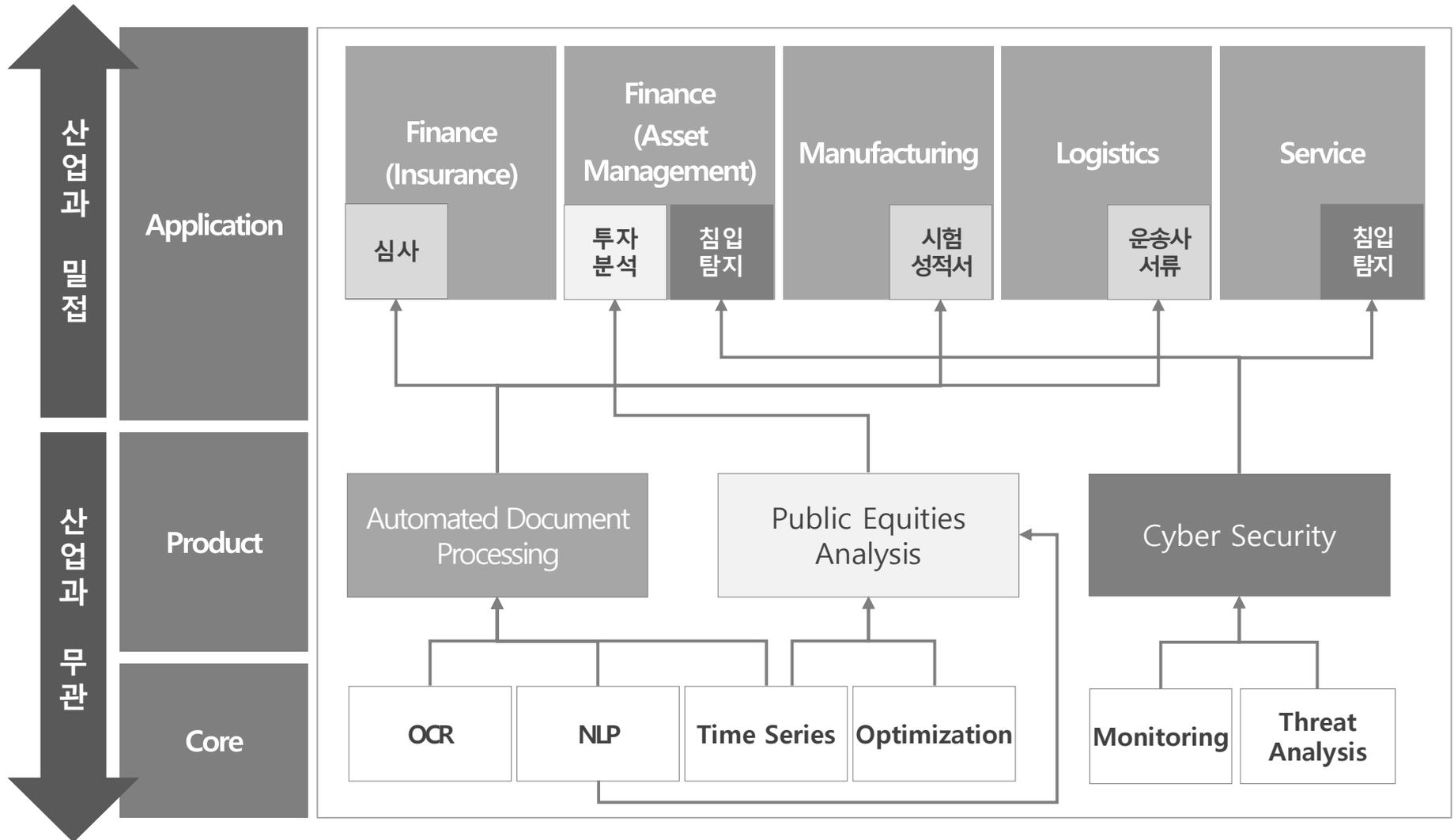
### 1 Define objectives first



### 2 Gain Experience, and Review Objectives



AI 기술 구조를 Core, Product, Application으로 구분함. OCR, NLP 등의 Core를 개발한 뒤 Core를 조합하여 Product를 구성하고, 각 산업에 적용하기 위한 Application 개발에 활용함



# Agenda

## I. 시장 Trend 및 동향

## II. 구축사례

- 프로젝트 개요
- 프로젝트 수행과정
- 전략적 파트너

## III. Lessons Learned

AI 도입을 통해 심사 프로젝트를 효율화하여 프로세스 소요 시간을 단축하고, 정확도를 향상시키며, 심사 비용을 절감하고자 함

### 내용

#### 정의

- 본 과제는 보험 가입자가 병원에서 진료를 받은 후 의료비를 청구하는 Claim 업무 프로세스에 관한 내용임
- Claim 업무란 보험 가입자가 청구서, 진료비 영수증, 진단서 등의 필요 서류를 갖춰서 보험금을 청구하는 절차를 말함
- 통원, 입원과 그 이외의 담보가 있지만, 프로젝트 범위는 통원으로 한정함

#### 이슈/ 요구사항

- Claim이 증가하고 있으나 인적 기반 처리로 인해서 프로세스 처리 시간이 증가함
- 심사자 개인 차이로 인해 심사 정확도의 상향 평준화가 필요함
- Claim 처리 외주 비용이 증가함

### 추진 목적

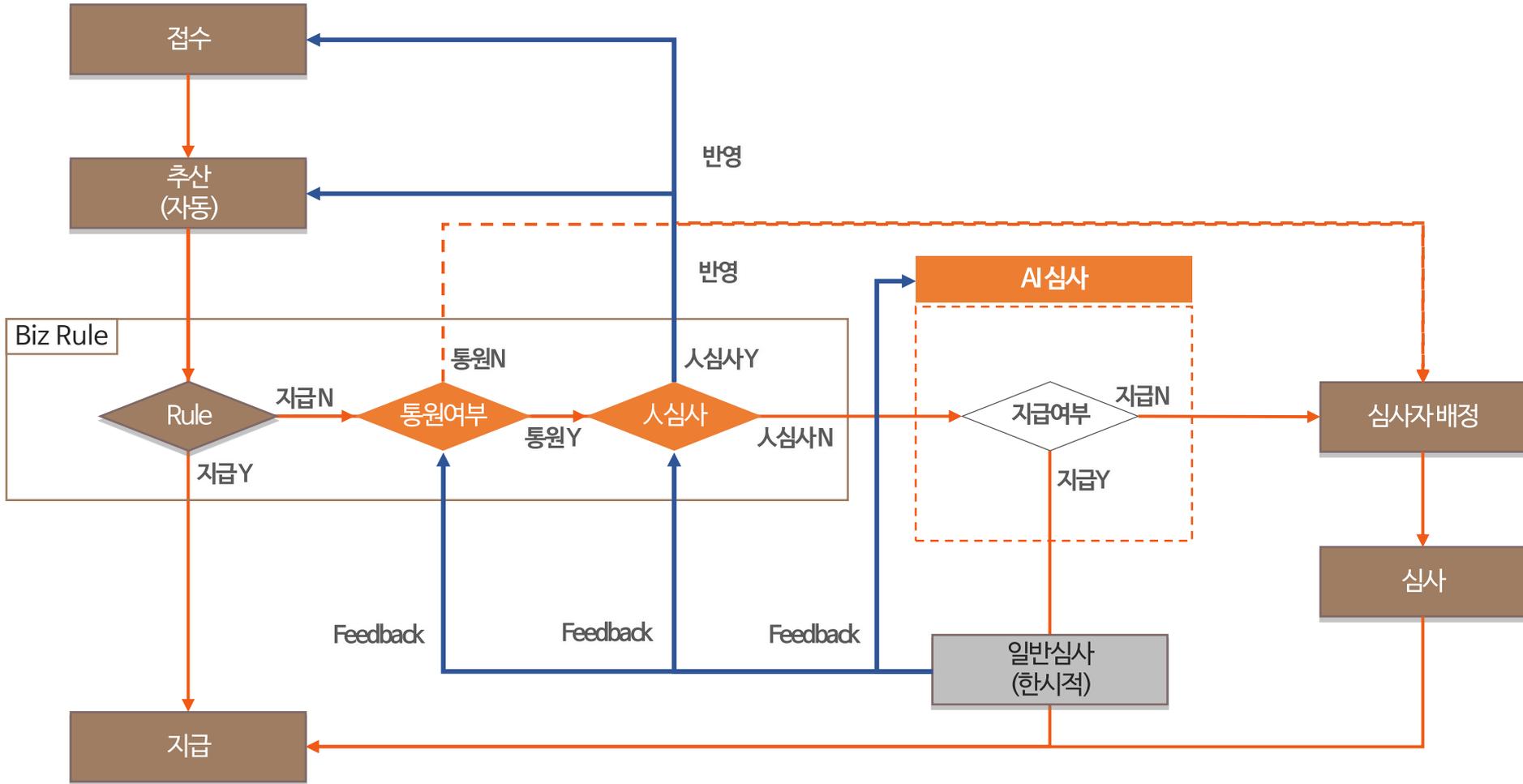
AI 기반의 심사 프로세스  
선진화를 통한 Biz 경쟁력 강화

업무표준화/자동화를 통한  
심사 프로세스 단축

시스템 기반의  
심사 정확도 향상

외부 심사 축소를 통한  
심사 비용 절감

AI 심사 프로세스 도입을 통해 자동으로 지급 여부를 판단하며, AI심사 이전에 통원여부와  
 인심사 여부를 걸러내는 단계를 추가함



## AI 프로젝트는 크게 데이터 준비, 모델링, 평가의 과정으로 구분됨

### 데이터 준비 학습Data구성 및 최적화

### Modeling Data 검증 및 모델 개발

### 평가 Matrix 및 심사원 평가

#### 현업인터뷰

- 보상, 심사, IT장기, 현장담당등 인터뷰 진행

#### 자료 분석

- POC결과, 손사 ERD모델링 등 Biz 분석

정형 데이터

바로 활용 가능한 데이터

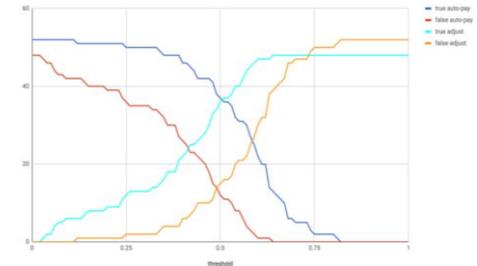
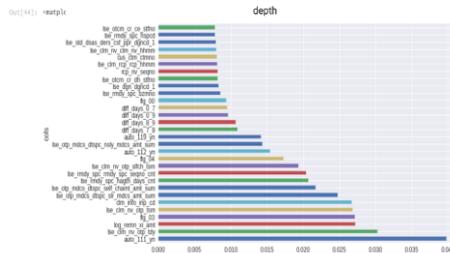
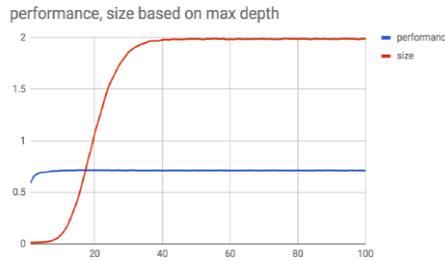
- 클레임번호
- 접수일시
- 발병일시
- .
- .
- .

비정형 데이터

(반정형) 전처리 필요 데이터

- 사고내용
- 담보상세
- 보장내용

- Generalized linear models
- Logistic regression
- Elastic Net regression
- Bayesian regression
- Random forests
- Support vector machines
- Gradient boosted machines
- Neural networks (1 hidden layer)
- Deep learning neural networks

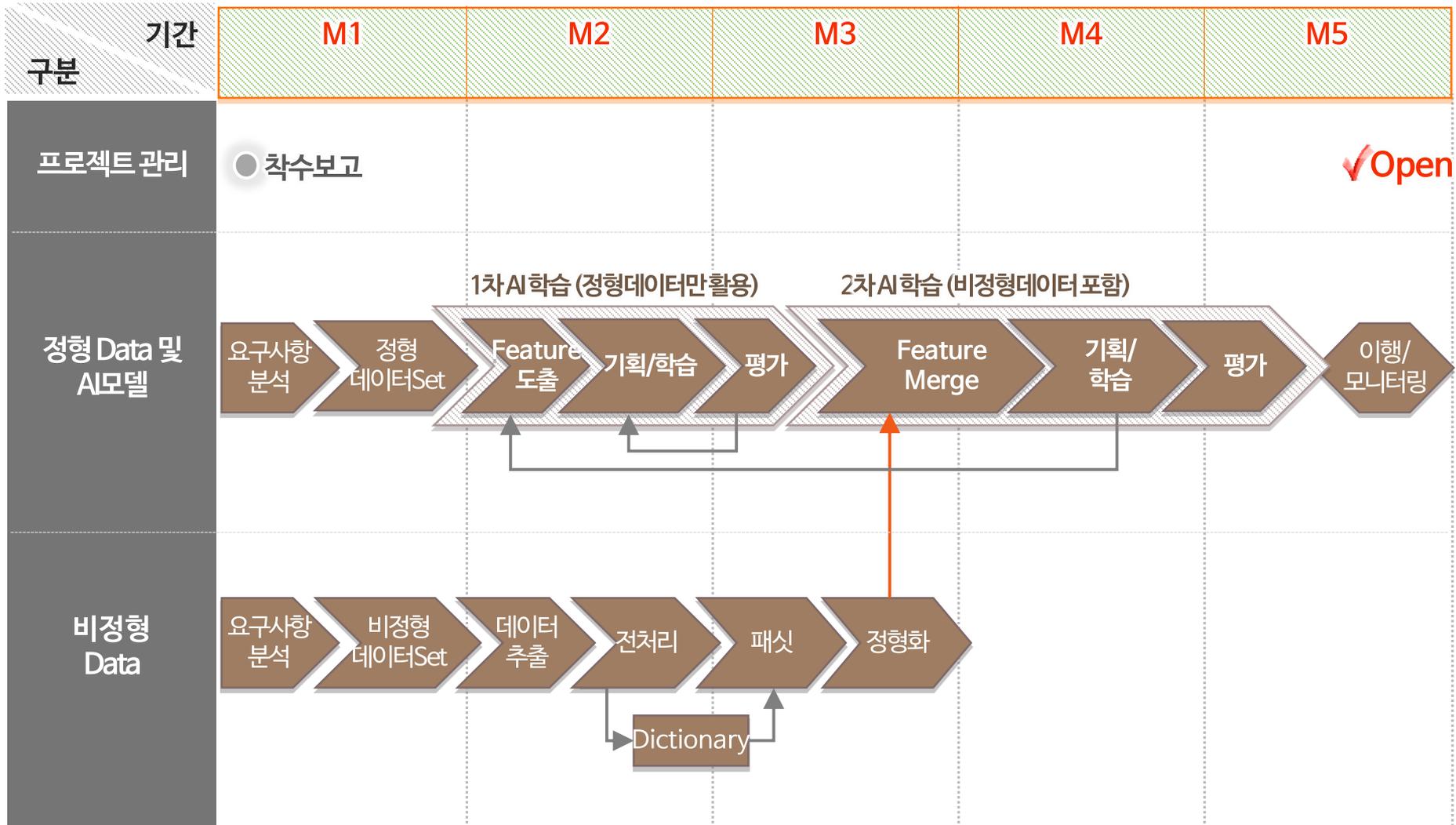


Matrix 분석		모델		Total
		Y	N	
데이터	Y	40	10	50
	N	10	40	50
Total		50	50	100



심사원 검토

전체 프로젝트는 5개월과 안정화 기간을 거쳤음. 정형 데이터와 비정형 데이터 Set을 도출하고, 반복적인 학습을 통해 모델의 정확도를 높여가는 과정임



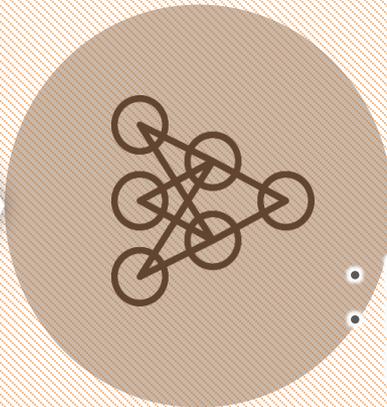
전체 테이블 중 인터뷰와 프로세스 분석을 통하여 주요 테이블을 추출 한 후, 데이터 분포도 분석과 항목별 중요도 분석을 통하여 모델구현을 위한 Feature를 도출함

## Business상 유의미한데이터



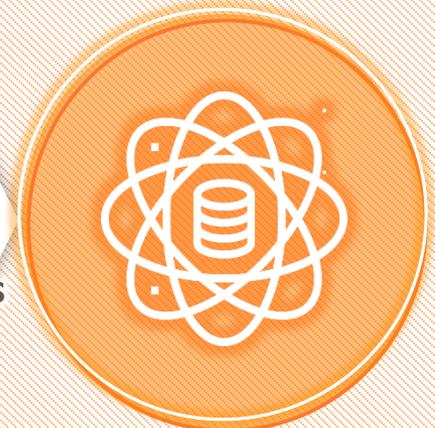
**전체 Data**  
(100개 테이블)

## 분석 및 학습관점의 유의미한 데이터



**Learning Data Set**  
(60개 테이블, 2,000컬럼)

- Pre-Analysis
- Feature Engineering



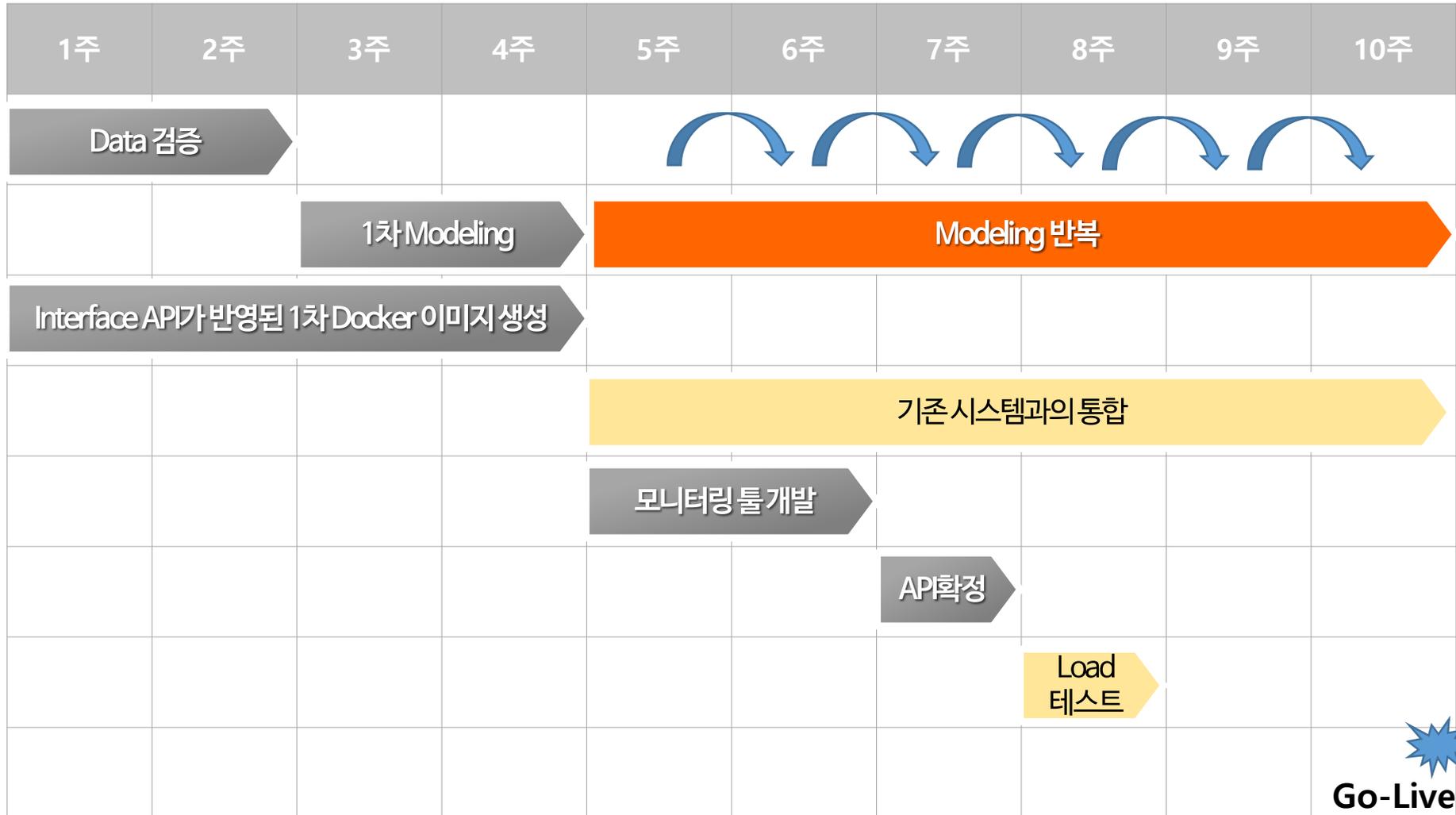
**모델용 Feature 도출**  
(정형 350 컬럼, 비정형 30컬럼)

모델용  
Feature 구성  
항목

- 고객
- 직업
- 보험계약
- 기타



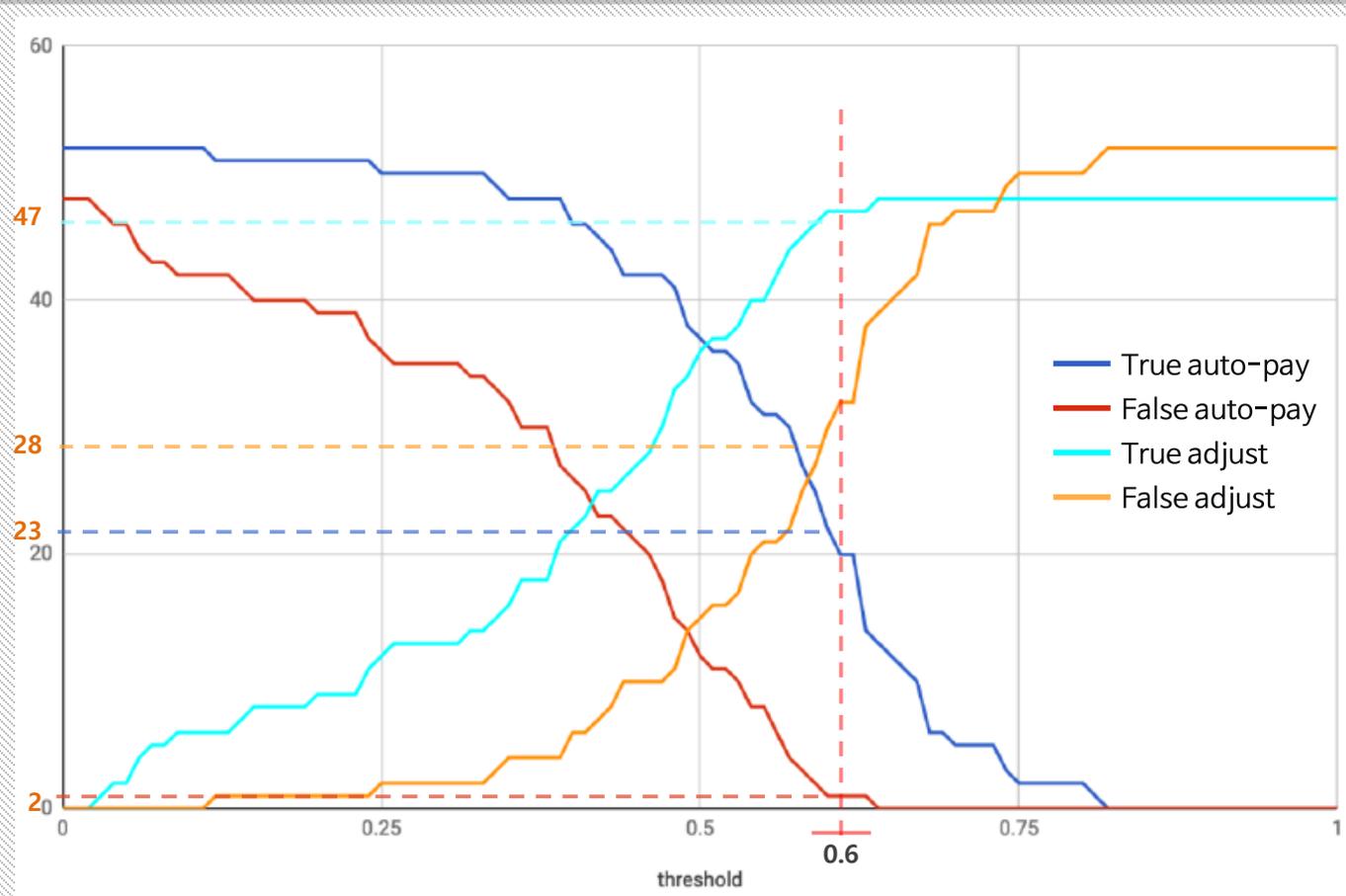
데이터Set에 맞는 알고리즘 후보군들을 선정하고, 가장 정확도가 높은 알고리즘을 활용하여 고도화 작업을 진행하는데 학습은 꼭 필요한 데이터만으로 진행해야 함



Go-Live

모델의 평가 Matrix rate를 분석하여 Threshold를 선정하는데, 각 지표의 Trade off를 고려해야 함

평가 Matrix rate - Classification



평가 Matrix\* (Threshold 0.6)

		모델		Total
		Y	N	
실 데이터	Y	40	10	50
	N	10	40	50
Total		50	50	100

항목	계산식	결과값
탐지율 (Accuracy)	40+40	80%
	실제 Y, N 값을 맞춘 비율	
정확도 (Precision)	40/50	80%
	Y라고 예측한 것 중 실제 Y의 비율	
재현율 (Recall)	40/50	80%
	실제 Y 중 Y로 예측한 비율	
오검출율 (Fall-out)	10/50	20%
	실제 N을 Y로 예측한 비율	

\*Confusion matrix 성능 지표 산출식: <http://www.waytoliah.com/1222>

여러 버전의 학습 모델을 만들어 분석/검증하고 결과를 다시 모델에 반영/재학습하는 과정을 거쳐서 최적의 모델을 선정함

## ✓ 모델 마스터 관리

No	Date	Version	Remark
1			
2			
3			

- Model history 관리
  - 모델별 적용 알고리즘, Feature 관리
  - 결과값 분석

## ✓ 그룹별 특성치 통계 분석

		모델		Total
		Y	N	
실 데이터	Y			
	N			
Total				100



- IT Engineer 평가 - 그룹별 특성치 분석

## ✓ 결과 반영 및 재학습

```

In [71]: train_df_raw, test_df_raw = sklearn.model_selection.train_test_split(df, test_size=0.1,
                                     shuffle=True, random_state=42)

In [84]: X_train=train_df, train_y, nan = proc_df(train_df_raw, pred_name)

# Scize train_df, train_y, nan, mapper = proc_df(train_df_raw, pred_name, do_scale=True)
CPU times: user 4.47 s, sys: 1.48 s, total: 5.95 s
Wall time: 5.94 s

In [75]: Wcfe
w = RandomForestClassifier(n_estimators=40, min_samples_leaf=1, max_features=None, oob_score=True, n_jobs=-1)
m.fit(train_df, train_y)

# m = LogisticRegression(solver='lbfgs', C=1.0, solver='lbfgs', n_jobs=-1)
# m.fit(train_df, train_y)
CPU times: user 1h 18m 25s, sys: 28.5 s, total: 1h 18m 49s
Wall time: 26m 35s

In [76]: Wcfe.score(train_df, train_y), m.oob_score,
# Scize m.score(train_df, train_y)
CPU times: user 24.2 s, sys: 3.84 s, total: 27.3 s
Wall time: 10 s

Out[76]: (0.999674854749084, 0.7060283661792587)
    
```



- 평가 결과 반영하여 재학습
  - Feature 변경
  - 변수, 알고리즘 변경 적용

## ✓ 검증인력 투입/분석

1	검증인력	검증인력 번호	성별	공역이 변동 한 사유	확인 결과1	확인 결과2	세부적인 확인 방법	일
2	201609109656	1 A20117709324000	Y	고용사고버튼	분석완료	LTRE00	확인된 'OO' 부분에 고용사고라고 되어 있음	3
4	20160710168	2 A20118510421000	N	보상확수	보상확인	LTRE03	확인된 '능력기준' 부분이 '180'인데 실제 항목은 200이 됨	3
5	20161215812	1 A2011738174000	N	보상확수	보상확인	LTRE00	기존의 보상인력을 보고 총 총합을 통해 확인하여	3
6	201609109656	1 A2007104898000	Y	임금입력비율	보상확인	LTRE03	확인된 임금인력 부분의 임금인력의 경우, 임금이 월총보통	3
7	201412108687	6 0600405383	N	임금입력비율	미지	ABC0003	임대차의 경우 월수령 항목을 확인	20
8	20150417957	3 Q2005010089						
9	20150417957	4 Q2005010089						
10	201703104807	1 A20094121030000						
11	201709104164	1 A20151307085000						
12	201709103985	1 A20167020197000						
13	201709104505	1 A20167045405000						
14	201610103409	5 A20095875525000						
15	201302104220	11 A2005201087000						
16	20141110686	2 A20094072069000						
17	201503101188	5 A20071018420000						
18	20150417925	10 A200711401991000						
19	20150417942	2 A2012730120000						
20	201512109957	7 C1010000710						
21	201512109957	7 Q2000012795						



- 검증인력 건별 검증 수행
  - 자동심사여부
  - 지급액변경사유분석

# 파트너사

요수아 벤지오 교수가 공동 창업한 Element AI사와 협업을 통해 프로젝트를 수행함



**안레쿤**

뉴욕대 교수  
인공지능 연구소장  
Facebook AI

**제프리 힌튼**

토론토대 교수  
석학 연구원  
Google AI

**요수아 벤지오**

몬트리올대 교수  
공동창업자  
Element AI

**앤드루 응**

前스탠퍼드대 교수  
인공지능 연구소  
Landing AI

**ELEMENT**

16년 8월 설립 총 펀딩 금액 100M USD,  
17년 5월 기업가치 295M USD 평가



# Agenda

I. 시장 Trend 및 동향

II. 구축사례

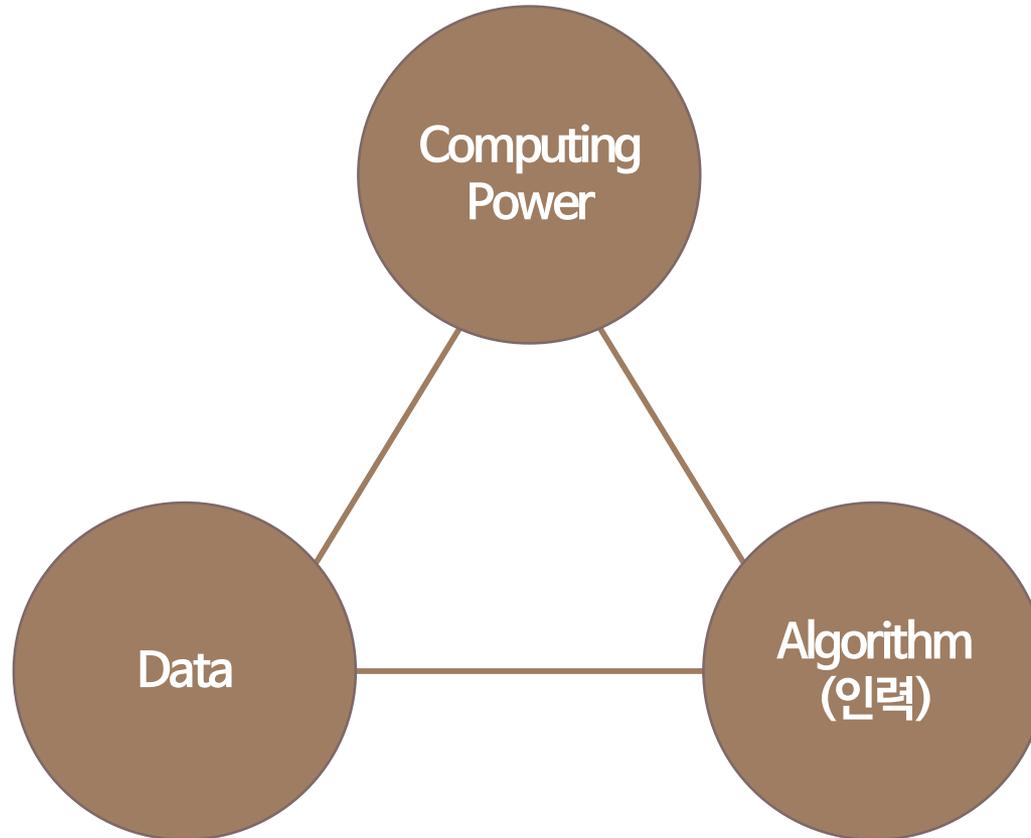
**III. Lessons Learned**

- Machine Learning의 3대 요소
- Role & Responsibility
- AI의 활용방안

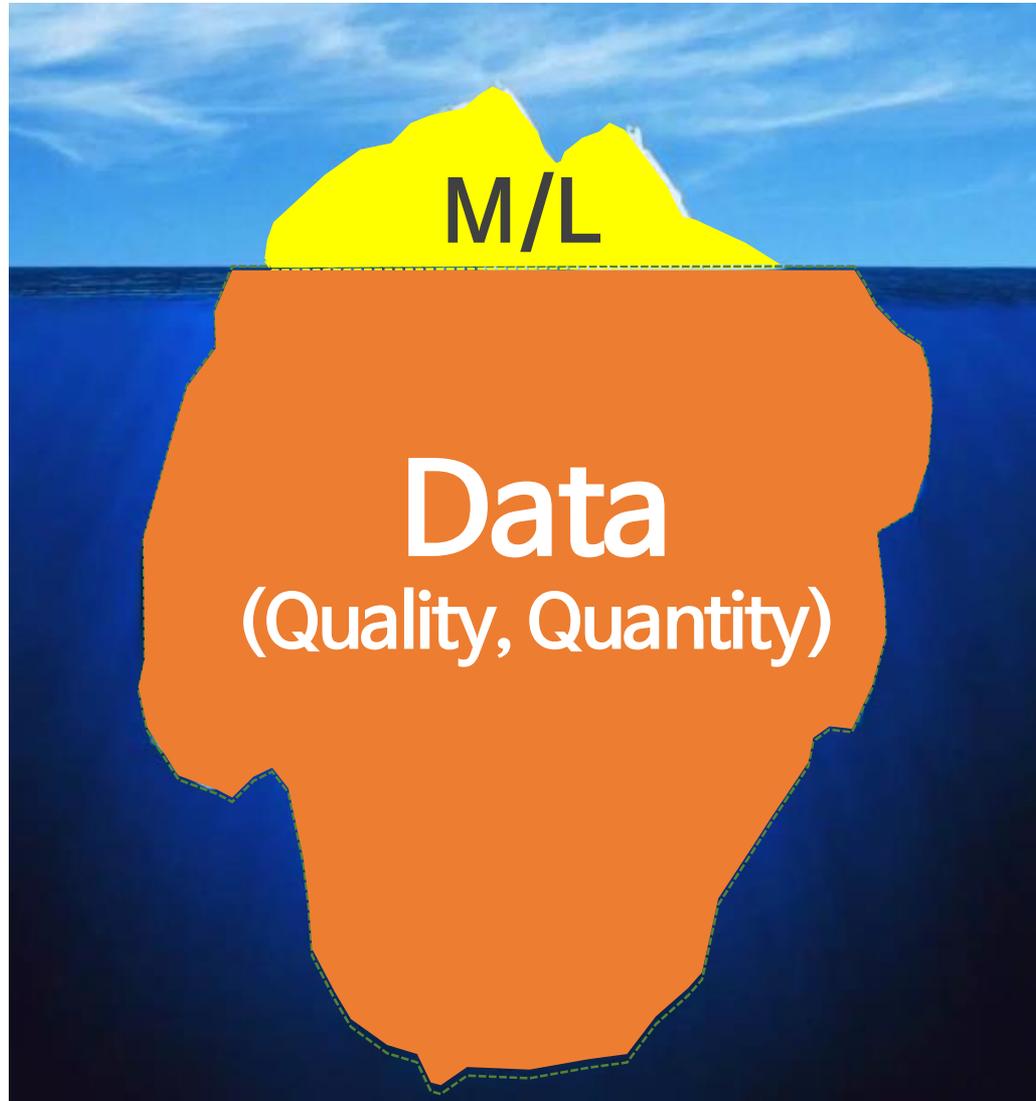
# Machine Learning의 3대 요소

---

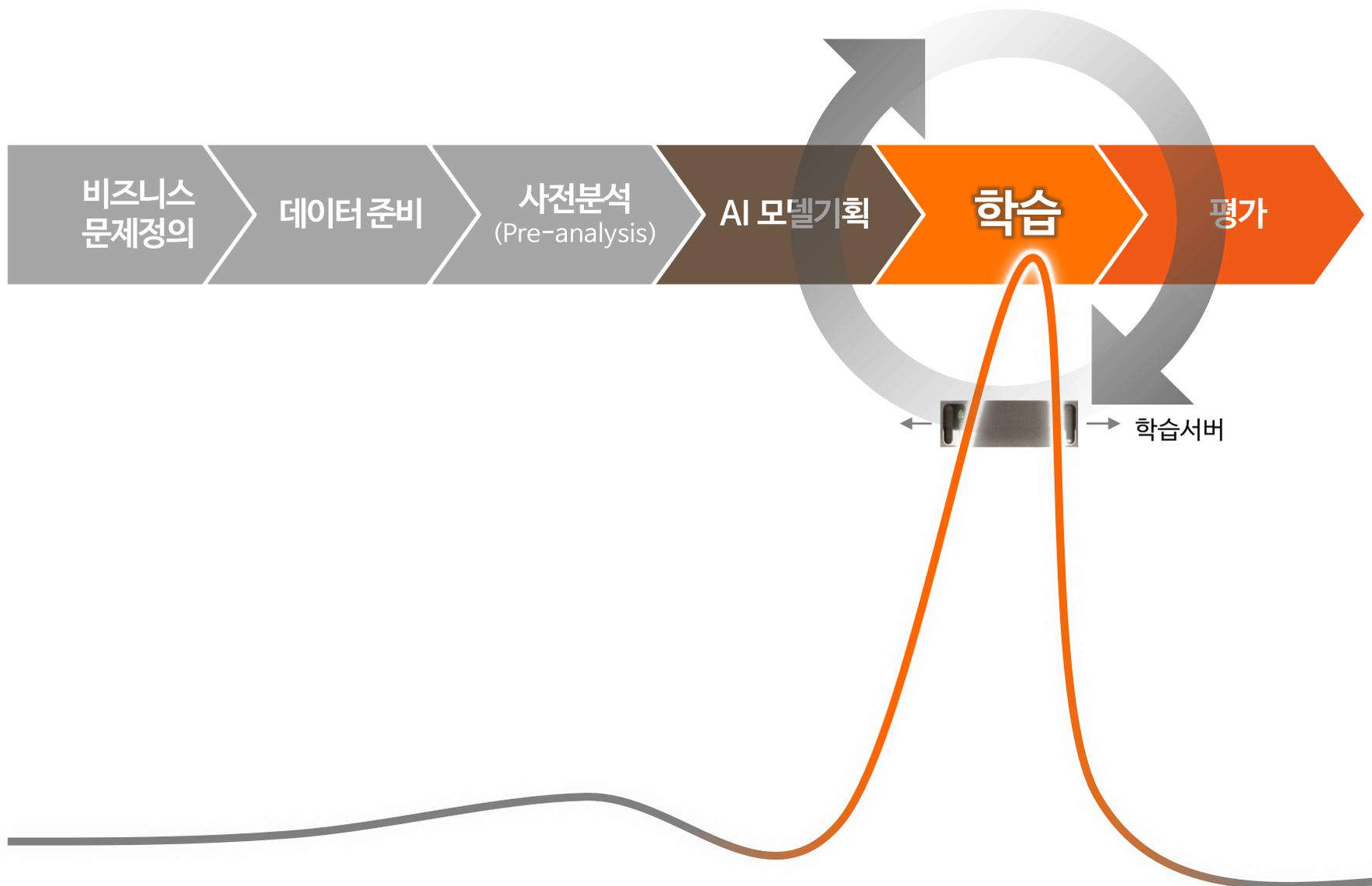
Machine Learning의 3대 요소에는 Computing Power, Data, Algorithm이 있으며, 3가지 요소가 조화를 이뤄야 효과적임



모델의 품질은 Data의 Quality와 Quantity에 의해서 좌우됨



GPU가 장착된 서버는 다양한 알고리즘과 모델의 학습을 매우 효과적으로 수행하게 함



# Role & Responsibility

알고리즘과 모델을 다룰 수 있는 인력뿐만 아니라, 산업과 업무 프로세스를 잘 이해하는 컨설턴시를 보유한 인력과 Data를 전처리하고 잘 다루는 인력도 필수적임

## Consulting

- 고객 Needs 분석
- 현업 Process 분석
- 시나리오 및 키워드 도출
- 학습데이터 셋 도출
- Label 선정 기준 수립
- KPI 수립  
(Performance 평가 기준)

- Business 이해
- 커뮤니케이션 능력
- AI 개요
- 수치 분석 및 이해
- 통계학

## Data Engineer

- 데이터 아키텍처 분석
- 학습 데이터 셋 추출
- 데이터 전처리

- SQL
- Python
- Pandas

## Data Analyst

- 데이터 분석
- 알고리즘 선택
- Feature Engineering
- 데이터 학습

- Python
- Pandas
- Sklearn 등 M/L 알고리즘

## 사람에게 어려운 것은 기계에게 쉽고 기계에게 어려운 것은 사람에게 쉽다?

### 사람에게 어려운 것

- 방대한 양의 데이터 처리
- 많은 양의 자료를 읽고 학습하기
- 쉬지 않고 일하기

### 기계에게 어려운 것

- 사진을 보고 개와 고양이를 판단 하는 것
- 아는 지, 모르는 지를 바로 아는 것
- 모르는 걸 물어보기

# AI 활용방안



# 감사합니다

한화시스템/ICT부문

디지털혁신팀 정해진 부장

[haejin.jeong@hanwha.com](mailto:haejin.jeong@hanwha.com)