

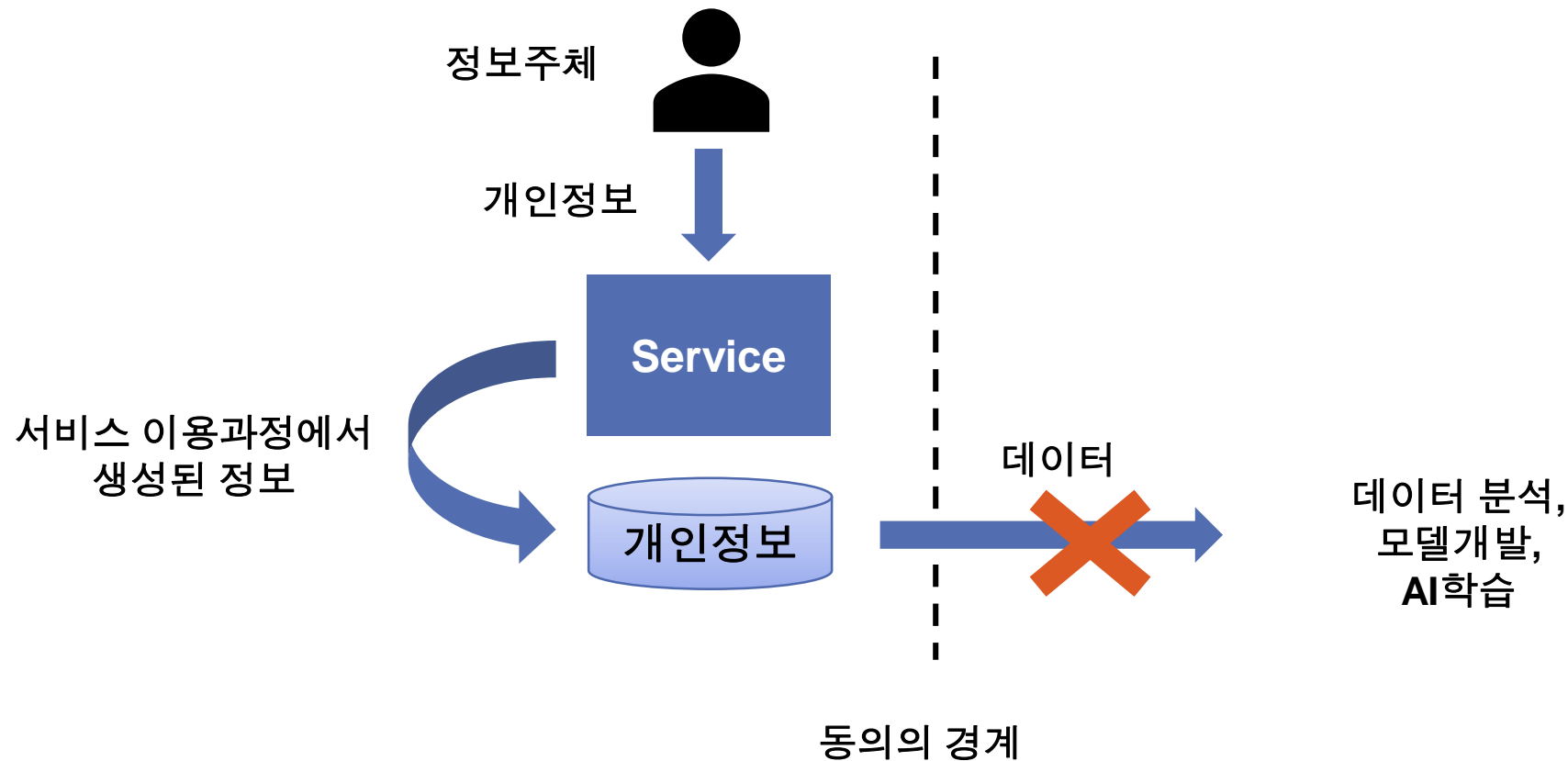
# 개인정보보호

# 내용

- ▶ 개인정보보호 개요
- ▶ 가명처리, 익명처리
- ▶ Privacy preserving machine learning

# 개인정보법

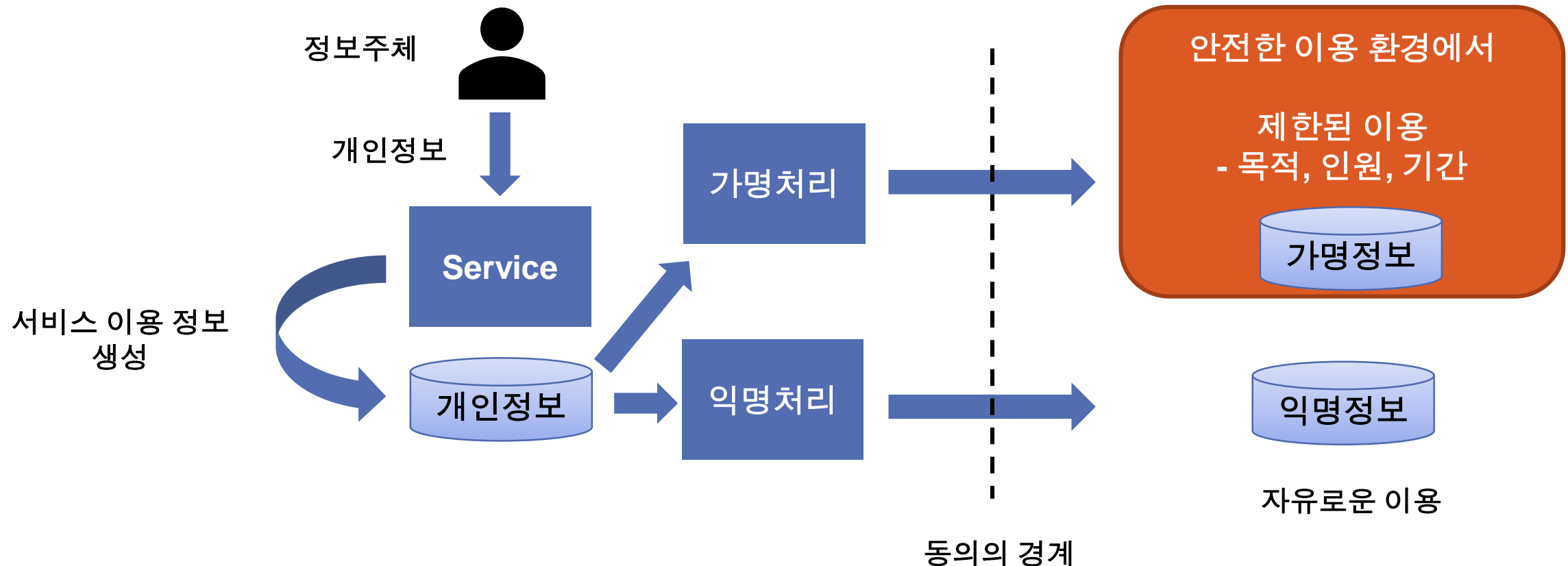
- ▶ 개인 정보는 정보주체의 동의없이 사용될 수 없음
  - 많은 데이터가 동의를 받지 못한 상태



# 데이터 3법

## ▶ 가명처리나 익명처리가 되면 동의없이 사용될 수 있음

- 익명정보 : 개인정보가 아니므로 제한없이 사용 가능
- 가명정보 : 통계 분석, 연구 등 목적, 여전히 개인정보이므로 안전한 환경에서 사용



# 익명정보, 가명정보

구분	개념	활용가능 범위
개인정보	특정 개인에 관한 정보, 개인을 알아볼 수 있게 하는 정보	정보주체로부터 사전에 구체적인 동의를 받은 범위 등의 내에서 활용 가능
가명정보	추가정보의 사용 없이는 특정 개인을 알아볼 수 없게 조치한 정보	다음 목적으로는 동의 없이 활용 가능 ① 통계작성(상업적 목적 포함) ② 연구(산업적 연구 포함) ③ 공익적 기록보존 목적 등
익명정보	더 이상 개인을 알아볼 수 없게 조치한 정보	개인정보가 아니기 때문에 제한 없이 자유롭게 활용

# 익명정보, 가명정보

< (예시) 원본 정보 >

성명	전화번호	성별	생년월일	보험가입건수
신사임당	010-1234-5678	여	1974.10.1.	3
권율	02-2345-6789	남	1990.3.26.	2
유관순	010-3456-4321	여	1969.5.28.	1
이수시	010-4567-9876	남	1993.11.3	2

< (예시) 가명처리된 정보 >

ID	성명	전화번호	성별	출생년도	보험 가입건수
9A00F1155584BA5DDFFC486DDD 7940431737C612651267FBD4716 FE93C46F6BA	신사임당	010-1234-5678	여	1974	3
C2E6376B9035D7067C8B68F25FA 34592F210D72E59B8E3F018C941 B391AB1D99	권율	02-2345-6789	남	1990	2
DACE2CCC9F459387EAE890D853 4955003F78B2B474C997CF2D990 573D4C3344F	유관순	010-3456-4321	여	1969	1
27B339D75FF1DCED2C29A866BA 5D61555D4C2E2C708F121AFABF3 4E5777AE498	이순신	010-4567-9876	남	1993	2

< (예시) 익명처리된 정보 >

성명	전화번호	성별	나이	보험 가입 건수	
권율	02-2345-6789	D	20대	2	동질집합 (k=3)
이순신	010-4567-9876	D	20대	2	
류성룡	010-7890-1234	D	20대	2	
안중근	010-6789-0123	D	30대	3	동질집합 (k=3)
이황	010-8901-2345	D	30대	5	
이아	010-9012-3456	D	30대	2	

# 속성분류

속성	가명·익명처리 대상 정보
식별자	성명, 상세주소, 전화번호, 생체인식정보, 전자우편주소, 사회관계망 서비스 주소, 주민등록번호, 여권번호, 운전면허번호, 외국인등록번호, 「정보통신망법」 제23조의3에 따른 본인확인기관이 특정 개인을 고유하게 식별할 수 있도록 부여한 정보, 특정 개인을 고유하게 식별하거나 동일한 신용정보주체를 구분하기 위하여 부여된 정보, 국내거소신고번호, 계좌번호, 신용카드번호, 건강보험증번호, 기기식별자, 자동차번호 등
개인식별가능 정보	성별, 나이, 주소, 우편번호, 직업(직업명 혹은 직업코드), 사건발생일자 (사망, 승인, 수술, 퇴원, 방문 등), 위치(우편번호, 건물명, 지역 등), 인종, 출생국, 모국어, 가시적 소수인종집단 지위(visible minority status), 결혼 여부, 학력, 범죄경력, 종교, 의료 진단명, 보험 가입정보 (보험 종류, 가입건수, 가입채널, 가입일, 보장금액 등), 신용대출 정보 (대출건수, 계약일, 대출액, 상환액, 연체율 등), 납입보험료, 추정소득, 추정주택가격, 보유차량 정보, 핵심고객 여부, 내부 신용등급, CB신용 점수 등

# 가명처리

추가정보(예 : 가명정보와 기존 식별자를 연결하는 매핑테이블 등)를 사용하지 아니하고는 특정 개인인 신용정보주체를 알아볼 수 없도록 개인신용정보를 처리하는 것을 말하는데, 그 처리 결과가 ① 어떤 신용정보주체와 다른 신용정보주체가 구별되는 경우 ② 하나의 정보집합물에서나 서로 다른 둘 이상의 정보집합물 간에 어떤 신용정보주체에 관한 둘 이상의 정보가 연계되거나 연동되는 경우 ③ 위와 유사한 경우로서 대통령령으로 정한 경우의 어느 하나에 해당하는 경우로서 법령에 따라 그 추가정보를 분리하는 등 특정 개인인 신용정보주체를 알아볼 수 없도록 개인신용정보를 처리한 경우를 포함한다(동법 제2호제15호).

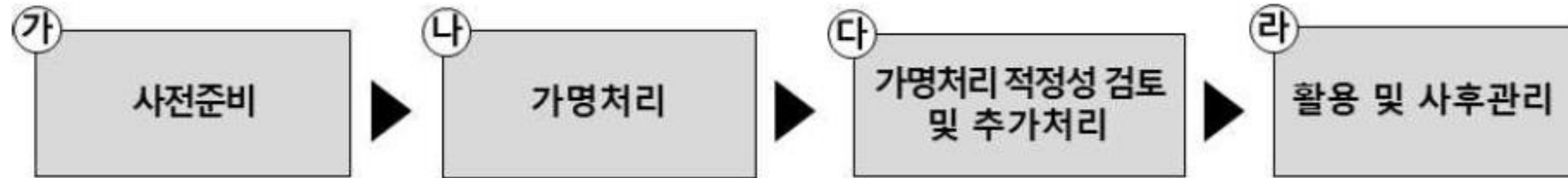


# 가명 정보

가명정보는 가명처리한 개인신용정보를 말하며, 어떤 신용정보주체와 다른 신용정보주체가 ‘구별’ 되더라도 특정 신용정보주체를 식별할 수 없는 경우에는 가명정보로 볼 수 있다. 또한, 가명정보는 하나의 정보집합물 내에서 또는 서로 다른 둘 이상의 정보집합물 간에 어떤 신용정보주체에 관한 둘 이상의 정보가 연계되거나 연동되는 경우를 포함하며, 별도의 매핑테이블 등 추가정보가 존재하는 경우에도 가명처리한 정보는 가명정보로 볼 수 있다(동법 제2조제15호).

- ※ 구별 : 성질이나 종류가 차이가 나는 것을 의미하는 것으로 특정 속성이 다른 속성과 구분되는 것을 의미함
- ※ 식별 : 분별하여 알아보는 것을 의미하는 것으로 속성을 통해 개인을 알아볼 수 있는 것을 의미함

# 가명처리 절차



## 가. 사전 준비

- 목적 구체화 : 통계, 연구, 기록 보존
- 처리대상 추출 : 처리 대상 테이블 추출

### 👉 필드 최소화

- 처리 및 활용 체계 구축 : 접근 통제 등 관리 방안 수립

### 👉 가명처리에서는 관리 방안이 핵심

- 가명정보 처리자와 원본정보 처리자의 분리
- 시스템 분리 : 가명정보를 원본정보와 대조해 볼 수 없게

# 나. 가명처리

## ▶ 위험도 측정

고려사항	세부 내용
가명처리 목적	- 통계작성(상업적 목적 포함), 연구(산업적 연구 포함), 공익적 기록보존 목적 여부 등 범위내에서 세부적인 목적
가명정보 활용 주체	- 내부 활용/내부 결합/외부 제공/외부 결합/외부 공개 여부 등
가명처리·이용 환경	- 처리 환경 및 이용(분석) 환경의 내부통제 수준, 재식별 의도 또는 능력 등 ※ 'II. 3. 다. 가명정보의 재식별 위험도 측정시 고려사항' 참고
가명처리 대상 데이터의 특징 분석	- 가명처리 대상 데이터의 특성 분석 - 데이터 속성(컬럼)을 식별자, 개인식별가능정보 등으로 분류 ※ 식별자, 개인식별가능정보 예시는 'I. 2. 나. 속성' 참고

## ▶ 수준 결정

- 각 조치 별 수준

## ▶ 가명처리

- 식별자 처리 (삭제 또는 대체) : 랜덤, 해시, 암호화
- 재식별 위험도가 높은 개인식별가능 정보에 대한 가명처리 : 일반화, 범주화, 상하단 코딩, 레코드 삭제 등

## 다. 적정성 검토 및 추가 처리

- 개인 식별 가능성 검토
- 추가 가명 처리


## 라. 활용 및 사후 관리

- 파기 등 관리

# 가명처리

## ▶ a.k.a De-identification

- 재식별될 수 있음
- 개인정보이므로 안전한 환경에서 사용되어야 함



		Non-Sensitive ( <b>Quasi ID</b> )			Sensitive
	Name	Age	Sex	Zip Code	Disease
1	Alice	28	F	13053	Heart Disease
2	Emily	29	F	13068	Viral Infection
3	Ken	21	M	13068	Cancer
4	David	23	M	13053	Viral Infection
5	Bob	31	M	13053	Heart Disease
6	Steve	37	M	13053	Viral Infection
7	Jane	38	F	13068	Cancer
8	Peter	35	M	13067	Cancer

Medical data

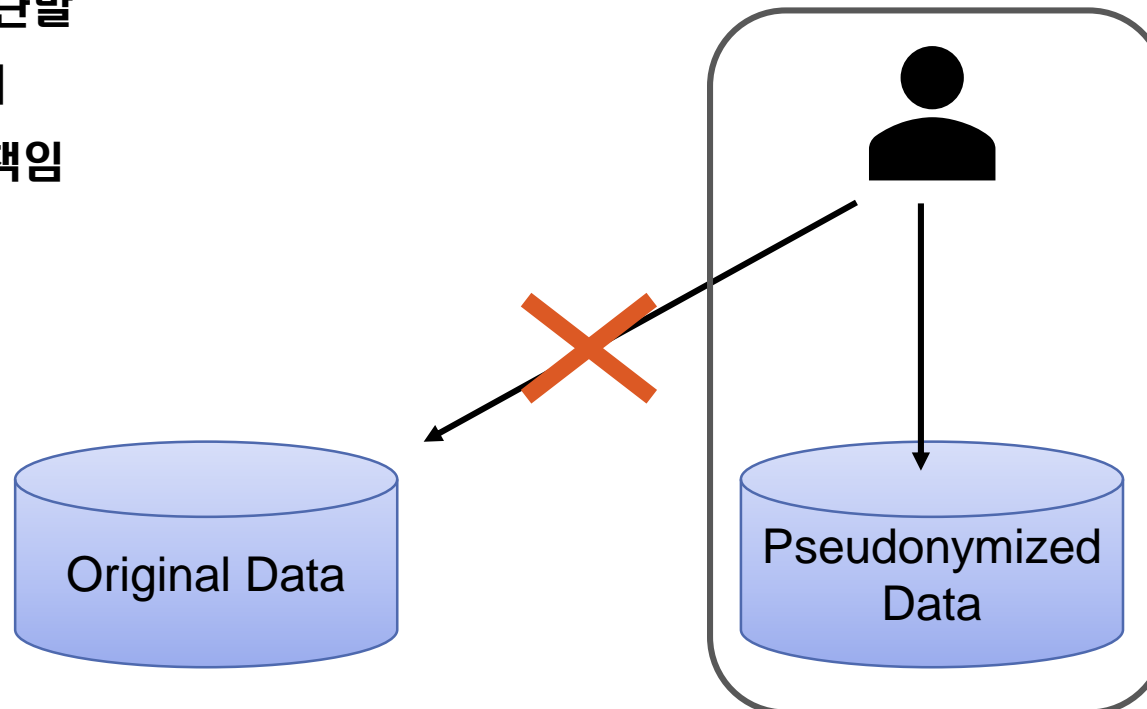
# 안전한 환경

## ▶ 개인정보 보호 조치, 환경

- 엄격한 접근제어, 로그
- 보안조치 : 망분리, 백신
- 보안절차 : 보안서약서, 보안 교육

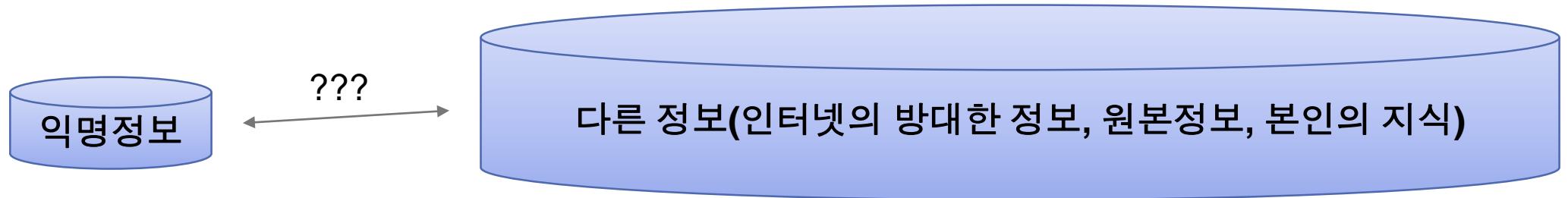
## ▶ 가명정보 처리자는 원본정보에 접근할 수 없어야

- 분리된 저장공간, 단말
- 처리자의 인적 분리
- 재식별 금지 법적 책임



# 익명정보

- ▶ 익명정보 - 시간 · 비용 · 기술 등을 합리적으로 고려할 때 **어떠한 다른 정보를** 사용하여도 더 이상 개인을 알아볼 수 없는 정보
- ▶ 재식별되면 안됨



	QI			SA
	Age	Sex	Zip Code	Disease
1	28	F	13053	Heart Disease
2	29	F	13068	Viral Infection
3	21	M	13068	Cancer
4	23	M	13053	Viral Infection
5	31	M	13053	Heart Disease
6	37	M	13053	Viral Infection
7	38	F	13068	Cancer
8	35	M	13067	Cancer

Name	Age	Sex	Zip Code
Bob	31	M	13053
Steve	37	M	13053
Jane	38	F	13068

**External Data -  
Voter list for Cambridge, Massachusetts**

# 재식별 방지를 위해

## ▶ 연결공격을 어렵게

- 연결되는 레코드가 K개 이상되게 : k-anonymity (기준)

범주화, 레코드 삭제 등의 수단을 통해 기준 달성

	Non-Sensitive			Sensitive
	Age	Sex	Zip Code	Disease
1	28	F	13053	Heart Disease
2	29	F	13068	Viral Infection
3	21	M	13068	Cancer
4	23	M	13053	Viral Infection
5	31	M	13053	Heart Disease
6	37	M	13053	Viral Infection
7	38	F	13068	Cancer
8	35	M	13067	Cancer

Original data



	Non-Sensitive			Sensitive
	Age	Sex	Zip Code	Disease
1	28-29	F	130**	Heart Disease
2	28-29	F	130**	Viral Infection
3	21-23	M	130**	Cancer
4	21-23	M	130**	Viral Infection
5	31-37	M	13053	Heart Disease
6	31-37	M	13053	Viral Infection
7	35-38	*	1306*	Cancer
8	35-38	*	1306*	Cancer

2-anonymous Medical Data

?

Name	Age	Sex	Zip Code
Bob	31	M	13053
Steve	37	M	13053
Jane	38	F	13068



# 동질성 공격

## ▶ Sensitive information exposed

👉 **L-diversity** : L different values of sensitive information in an equivalent class

	Non-Sensitive			Sensitive
	Age	Sex	Zip Code	Disease
1	28-29	F	130**	Heart Disease
2	28-29	F	130**	Viral Infection
3	21-23	M	130**	Cancer
4	21-23	M	130**	Viral Infection
5	31-37	M	13053	Heart Disease
6	31-37	M	13053	Heart Disease
7	35-38	*	1306*	Cancer
8	35-38	*	1306*	Cancer

} equivalent class

2-anonymous Medical Data

# 현황

- ▶ 가명처리, 익명정보 -> 금융권 등에서 활발히 적용
- ▶ 타 분야도 도입 시작

# 속성 분류 및 익명처리 적용기준

1) 신용정보회사등은 익명처리의 대상이 되는 정보를 식별자, 개인식별 가능정보로 분류한 후 적절한 익명처리 기법을 적용하여야 한다.

※ 익명처리 대상 정보의 분류는 익명처리 목적 및 이용·제공 환경 등에 따라 달라질 수 있음

2) 익명처리시 식별자는 삭제하여야 하며, 부득이하게 정보 이용 목적상 필요한 경우에는 적절하게 익명처리를 한 후 이용하여야 한다.

3) 개인식별가능정보 중 개인 식별 가능성이 높은 속성은 그 정도에 맞추어 익명처리 수준을 높이는 등의 조치를 취하여야 한다.

※ k-익명성 모델 등의 기법도 적용 가능(붙임 1의 '9. 프라이버시 보호 모델' 참조)

4) 개인식별가능정보 중 개인 식별 가능성이 낮은 속성은 활용 목적, 해당 정보의 특성, 다른 정보와의 결합 등을 고려하여 필요시 동질성 공격, 배경지식 공격 등의 다양한 위협을 제거하기 위하여 추가로 익명처리 기법을 적용하여야 한다.

# 속성 구분

구분	속성	위험	익명처리 기술
ID	식별자	신용카드사에서 개인을 식별하기 위한 ID이므로 개인이 특정될 가능성이 있음	삭제
성명		식별자로 개인을 특정할 수 있음	삭제
카드번호		식별자이며 다른 사업자도 보유하고 있을 수 있으므로 개인이 식별될 위험이 존재	삭제
전화번호		식별자이며 다른 사업자도 보유하고 있을 수 있으므로 개인이 식별될 위험이 존재	삭제
성별	개인식별 가능정보* (k-익명성 적용)	생년월일, 주소 등의 정보와 조합하여 개인이 식별 될 수 있음	생년월일, 주소 등의 정보를 익명처리하여 위험 해소
생년월일		주소, 성별 등의 정보와 조합하여 개인이 식별 될 수 있음	생일을 삭제하고, 나이를 연령대로 범주화(20대/30대 등)
주소		성별, 생년월일 등의 정보와 조합하여 개인이 식별 될 수 있음	동 이하 주소는 삭제, k-익명성 수준을 만족하지 못할 경우 구 이하 주소 삭제 등 프라이버시 보호 모델 수준을 만족하도록 조치

# 속성 구분

직업	개인식별 가능정보 (k-익명성 미적용)	다른 사업자도 보유하고 있을 수 있으며 다른 정보와 조합하여 개인이 식별될 위험이 존재	자영업, 공무원, 회사원, 기타로 일반화
연봉		수입이 너무 많거나 적은 경우 다른 정보와 조합하여 개인이 식별될 위험이 존재	연봉의 분포를 고려하여 3단계로 범주화
내부신용등급		이미 등급화 된 정보이며 개인에 민감한 정보	내부신용등급을 5단계로 범주화
연체잔고		개인의 소득을 추정할 수 있는 민감한 정보	연체잔고의 분포를 고려하여 10단계로 범주화
결제기관		익명정보를 제공받을 사업자가 필요한 정보가 아님	삭제

# 가명 익명처리 실습

▶ 14.ipynb

# PRIVACY PRESERVING MACHINE LEARNING

## ▶ Crypto based

- Homomorphic encryption
- SGX
- Garbled circuit
- Secret sharing

## ▶ Perturbation based

- Differential privacy
- Dimension reduction

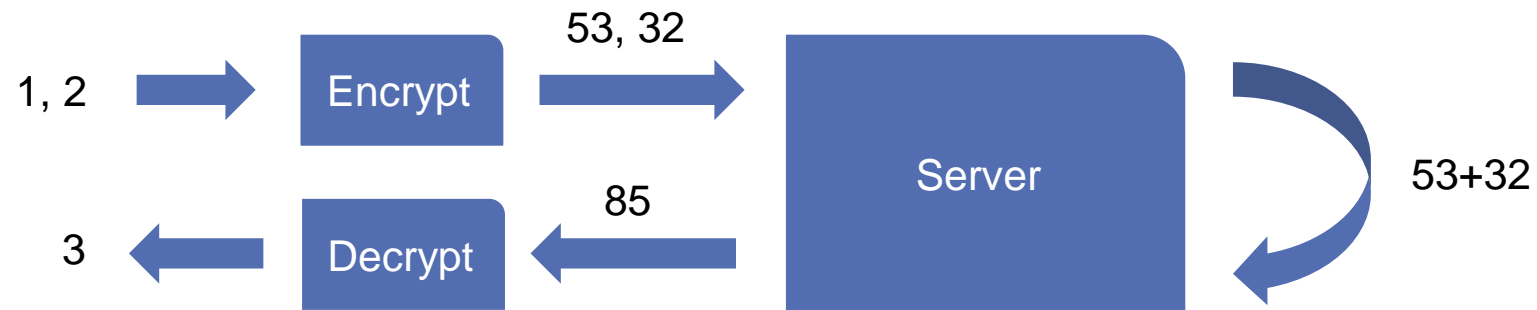
## ▶ Etc

- Synthetic data
- Federated learning

# HOMOMORPHIC ENCRYPTION

## ▶ 암호화된 상태에서 연산

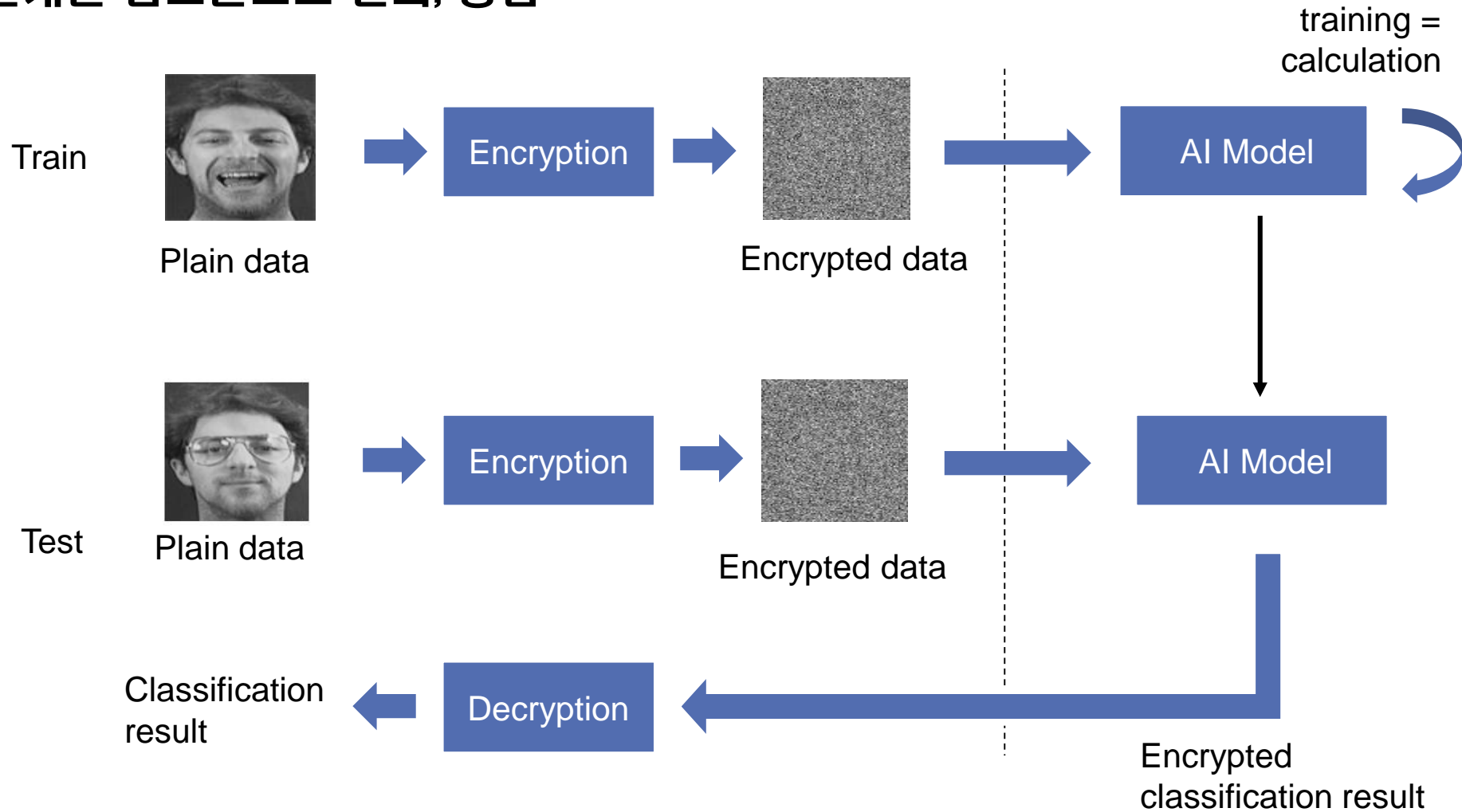
- 일반적 연산 : +, \*
- 특수연산 : 검색, 정렬, ..





# 동형암호활용 AI 학습

- ▶ 학습 단계는 암호문 또는 평문
- ▶ 활용 단계는 암호문으로 질의, 응답

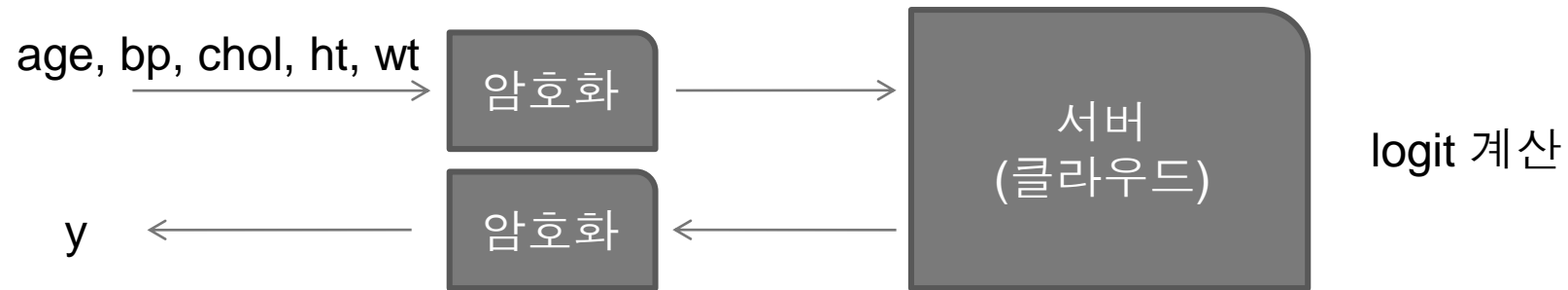


# HOMOMORPHIC ENCRYPTION

## ▶ 예제

- 심장병 발병 예측

$$y = \frac{e^x}{e^x + 1}, \quad x = 0.072 \cdot age + 0.013 \cdot bp + 0.008 \cdot chol - 0.053 \cdot ht + 0.021 \cdot wt$$



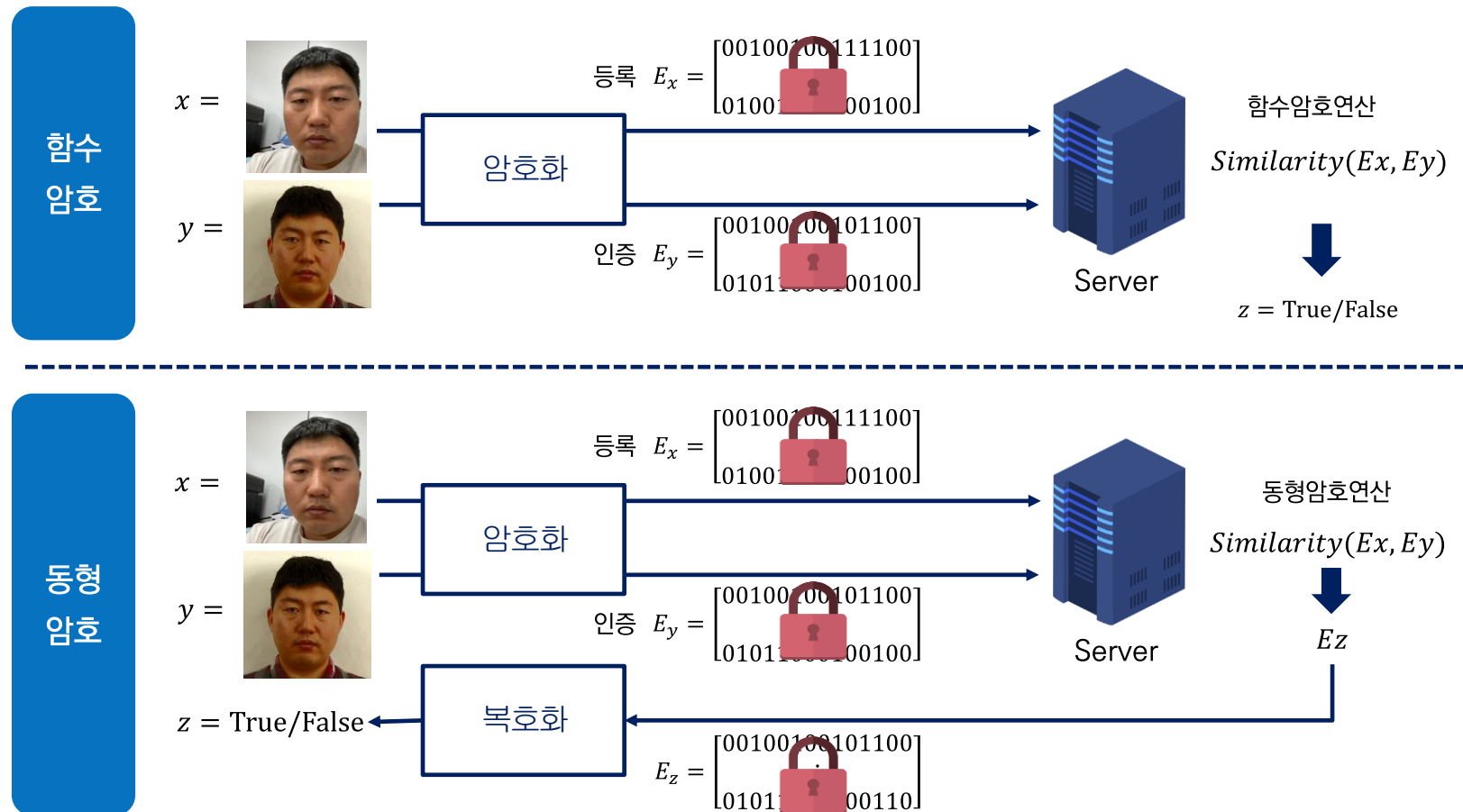
- 서버는 x, y 다 모름

## ▶ Test time, not training time

Model	Runtime (sec)		$\lambda$	Accuracy (%)	Dataset
	Total	Amortized time			
CryptoNets [14]	570	$69.580 \times 10^{-3}$	80	99.00	MNIST
E2DM [22]	28.590	$450.0 \times 10^{-3}$	80	98.01	MNIST
FCryptoNets [13]	39.100	39.100	128	98.71	MNIST
A*FV	5.160	$0.630 \times 10^{-3}$	82	99.00	MNIST
A*FV	5.710	$0.340 \times 10^{-3}$	175	99.00	MNIST
CryptoDL [20]	11,686	11,686	80	91.50	CIFAR-10
FCryptoNets [13]	22,372	22,372	128	75.99	CIFAR-10
A*FV	304.43	0.372	91	77.55	CIFAR-10

# 함수암호

## ▶ Functional encryption vs. homo...

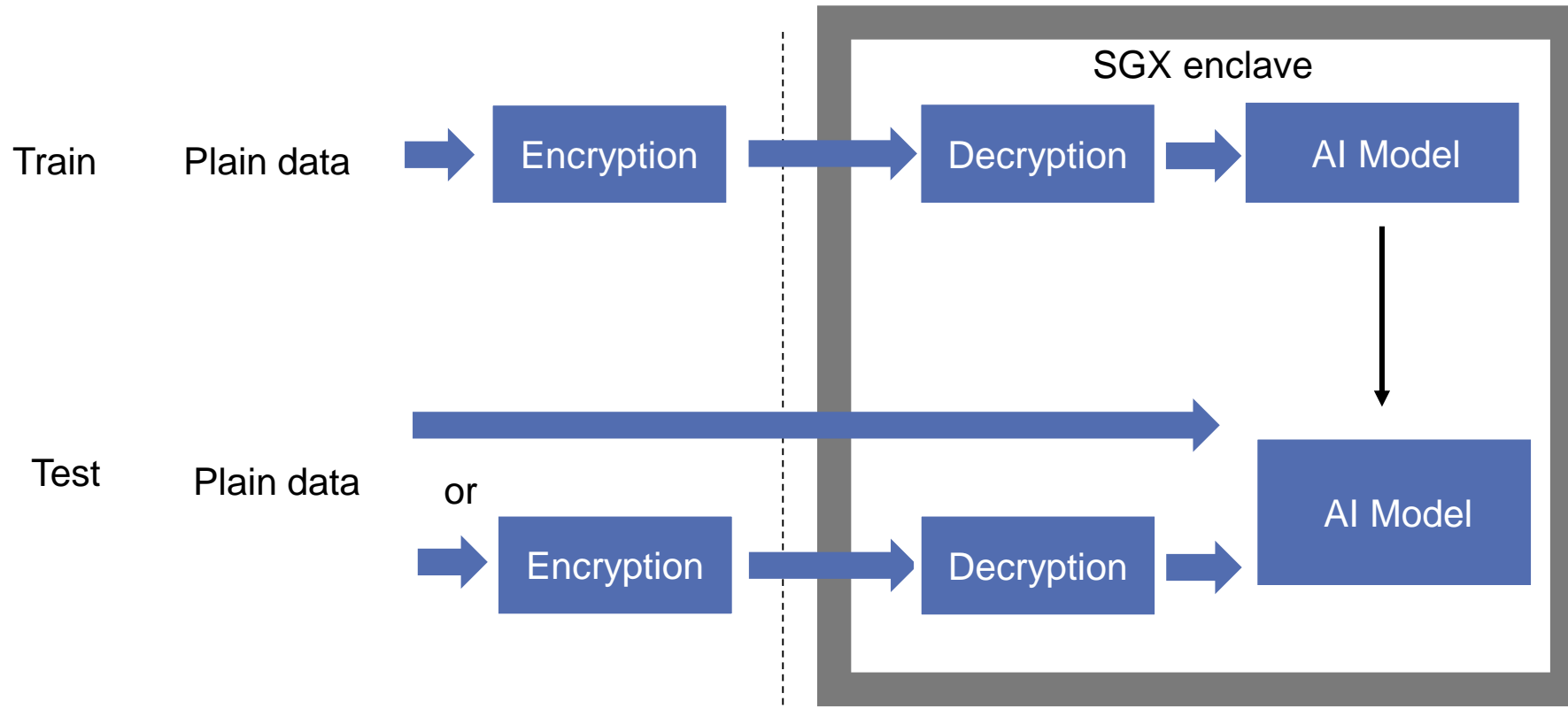


클라이언트에서 복호화된 연산결과를 얻기 때문에 인증에 부적합

# SGX

## ▶ 칩 내부에서 복호화 (칩 외부에서는 평문을 알 수 없음)

- 제한된 용량
- 별도의 코딩 ( 어려움 ) 필요
- 부채널 공격 등 존재



# DIFFERENTIAL PRIVACY

## ▶ 투표 전

- 찬성 : 10, 반대 : 5

## ▶ 당신이 투표 후,

- 찬성 : 10, 반대 : 6

## ▶ 당신의 표결 내용이 노출



# 차분 공격

## ▶ 2 DB 간의 특정인의 포함여부가 차이가 나는 것을 알 때

Q1 employee DB (anonymized)

	QI			SA
	Age	Sex	Zip Code	Disease
1	35-38	M	130**	Cancer
2	35-38	M	130**	Viral Infection
3	35-38	M	130**	Heart Disease
4	35-38	M	130**	Viral Infection
<b>5</b>	<b>35-38</b>	<b>M</b>	<b>130**</b>	<b>Viral Infection</b>
6	35-38	M	130**	Cancer

Q2 employee DB (anonymized)

	QI			SA
	Age	Sex	Zip Code	Disease
1	35-38	M	130**	Cancer
2	35-38	M	130**	Viral Infection
3	35-38	M	130**	Heart Disease
4	35-38	M	130**	Viral Infection
5	35-38	M	130**	Cancer

**exposed**

If you know a man who quit  
in Q2 and 36 years old

# 대응 방안

## ▶ 노이즈 추가

- E.g 가짜 레코드 추가 or oversampling

Q1 employee DB

	QI			SA
	Age	Sex	Zip Code	Disease
1	35-38	M	130**	Cancer
2	35-38	M	130**	Viral Infection
3	35-38	M	130**	Heart Disease
4	35-38	M	130**	Viral Infection
5	35-38	M	130**	Viral Infection
6	35-38	M	130**	Cancer
7	35-38	M	130**	Viral Infection

Q2 employee DB

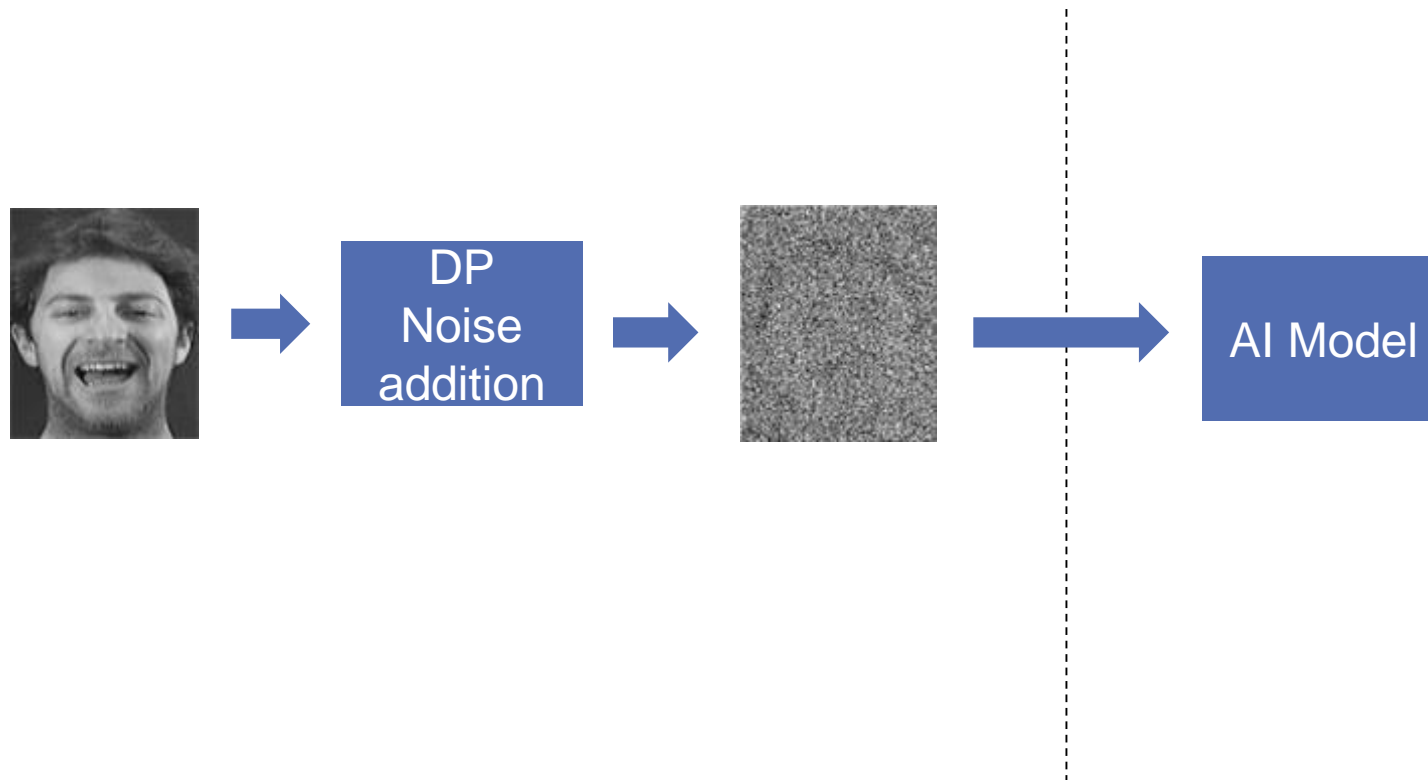
	QI			SA
	Age	Sex	Zip Code	Disease
1	35-38	M	130**	Cancer
2	35-38	M	130**	Viral Infection
3	35-38	M	130**	Heart Disease
4	35-38	M	130**	Viral Infection
5	35-38	M	130**	Cancer
6	35-38	M	130**	Viral Infection
7	35-38	M	130**	Viral Infection



# AI 학습에 차분 프라이버시 적용

## ▶ 학습데이터에 노이즈 추가

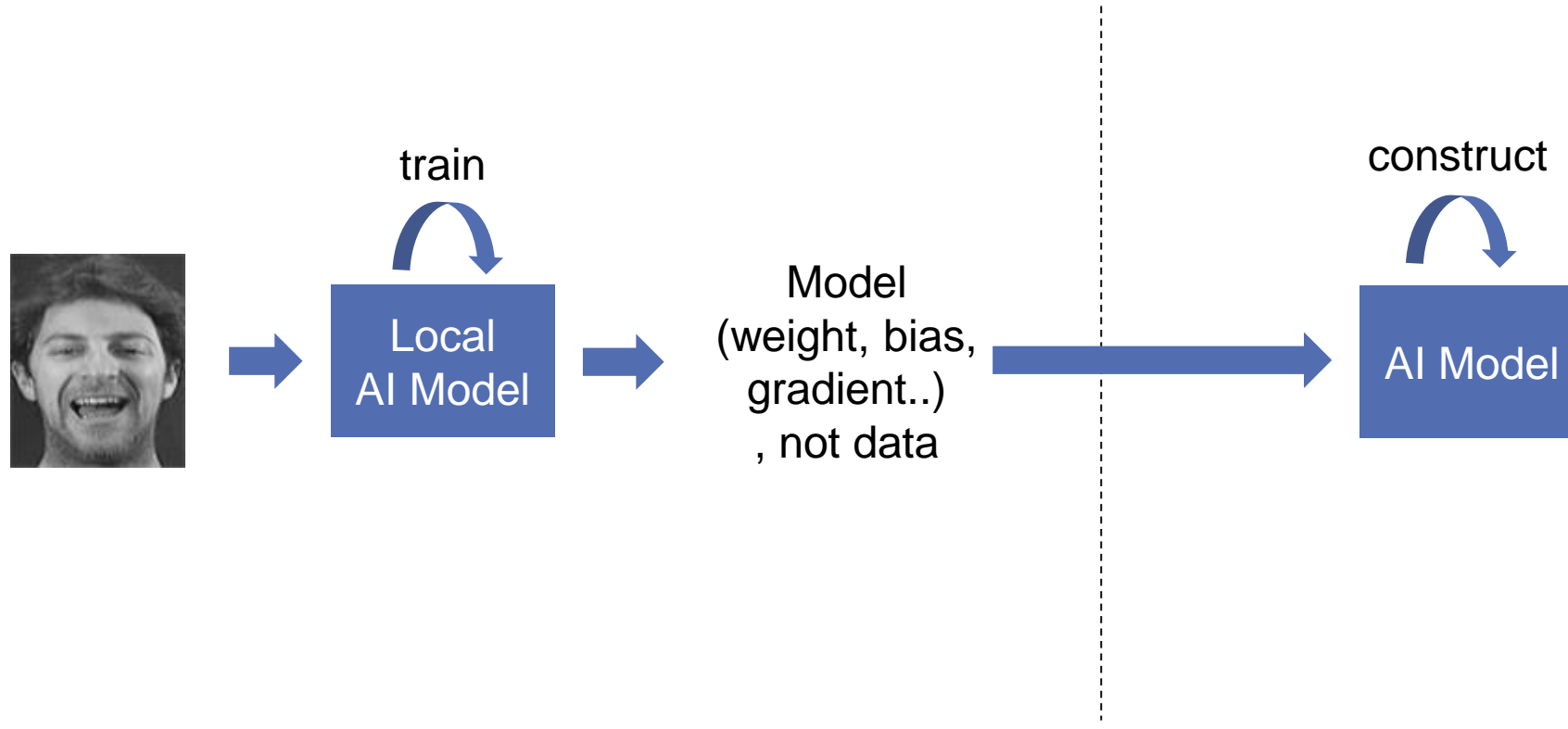
- 성능저하



# 연합학습

## ▶ 분산학습

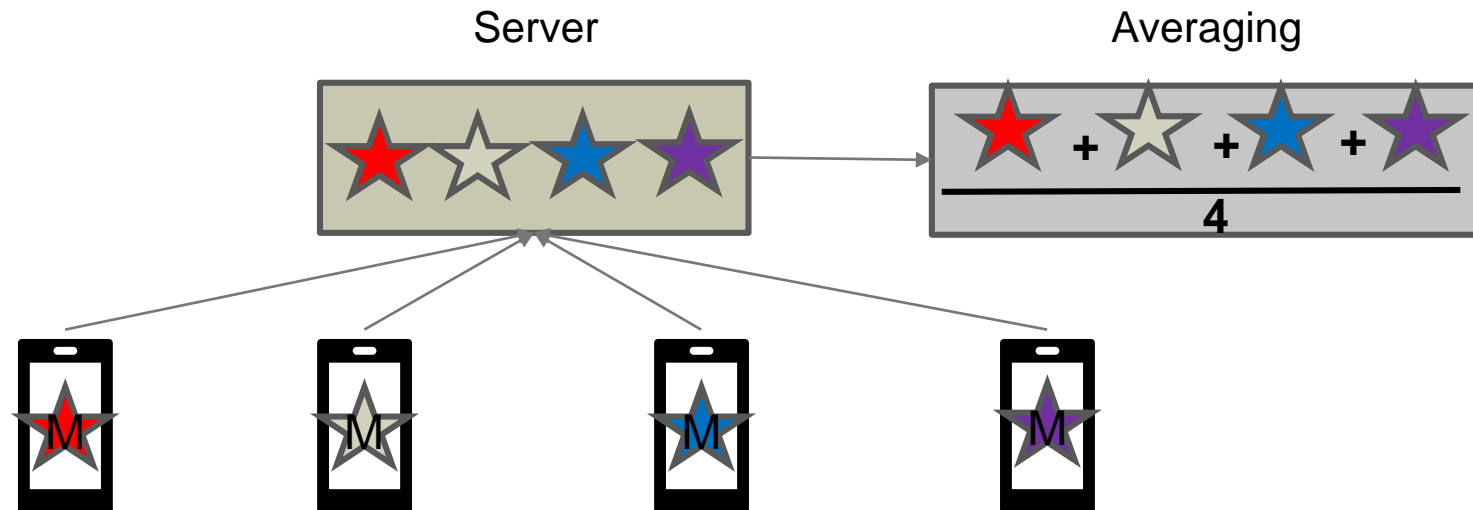
- local model 모델을 학습
- 모델 파라미터를 전송
  - Weight, bias, or gradient <= 개인정보가 아님
- 파라미터를 이용하여 모델을 구성



# 연합학습

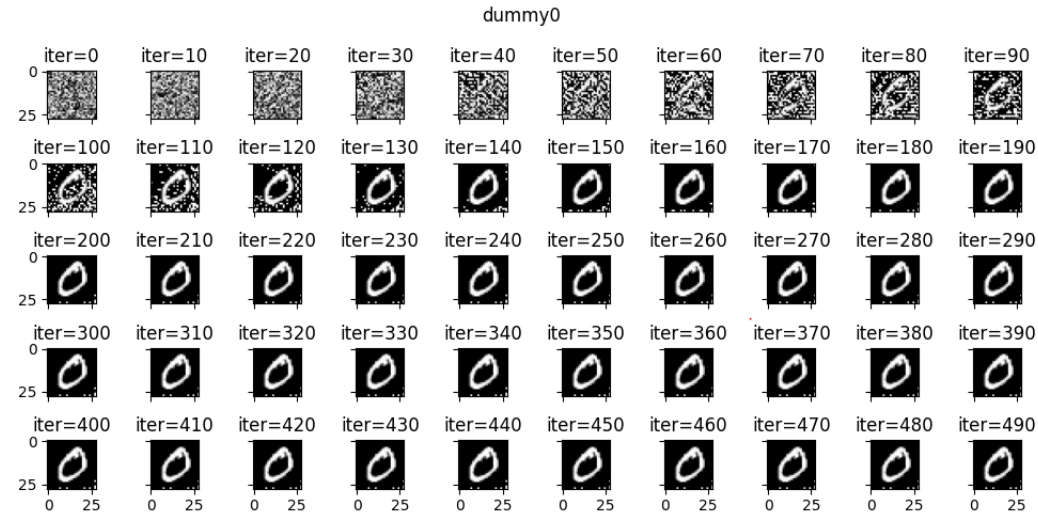
## ▶ 중앙집중형 분산학습

- 서버는 global model을 배포
- 클라이언트는 자신의 데이터로 수신한 global model 을 학습
- 서버는 클라이언트로 파라미터를 모아서 자신의 global model을 업데이트
- 위 과정을 반복



# FEDERATED LEARNING

## ▶ 데이터 복원이 가능



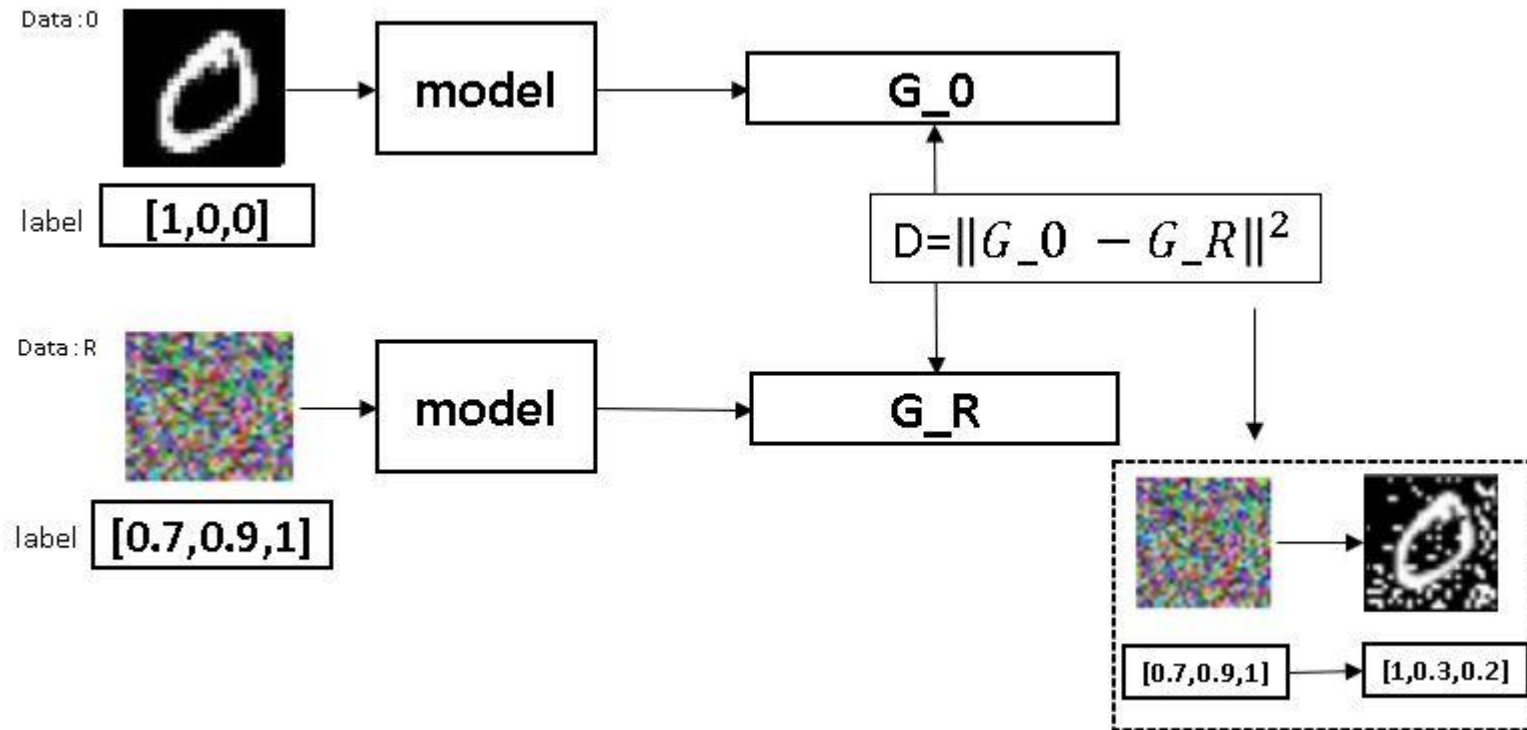
## ▶ 모델 오염 공격도 가능

- 데이터 오염 보다 쉬움
- 백도어를 만들거나 모델 성능 저하를 목표

# RECONSTRUCTION ATTACK

## ▶ 원리

- 랜덤데이터의 그래디언트를 원본의 그래디언트에 가깝게 데이터를 최적화

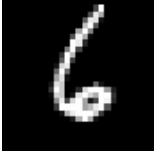


<Deep leakage from gradient, NeurIPS 2019>

# RECONSTRUCTION ATTACK


## ▶ example

Ground truth



Gradient = 10

Random data



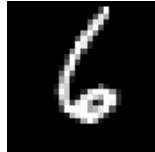
Gradient = 2

<Iteration = 0>

→  $D = \|10 - 2\|^2 \rightarrow 64$




Ground truth



Gradient = 10

Random data



Gradient = 6

<Iteration = 100>

→  $D = \|10 - 6\|^2 \rightarrow 16$

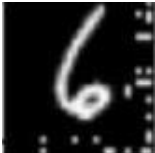


Ground truth

$D = \|10 - 2\|^2 \rightarrow 64$

Gradient = 10

Random data



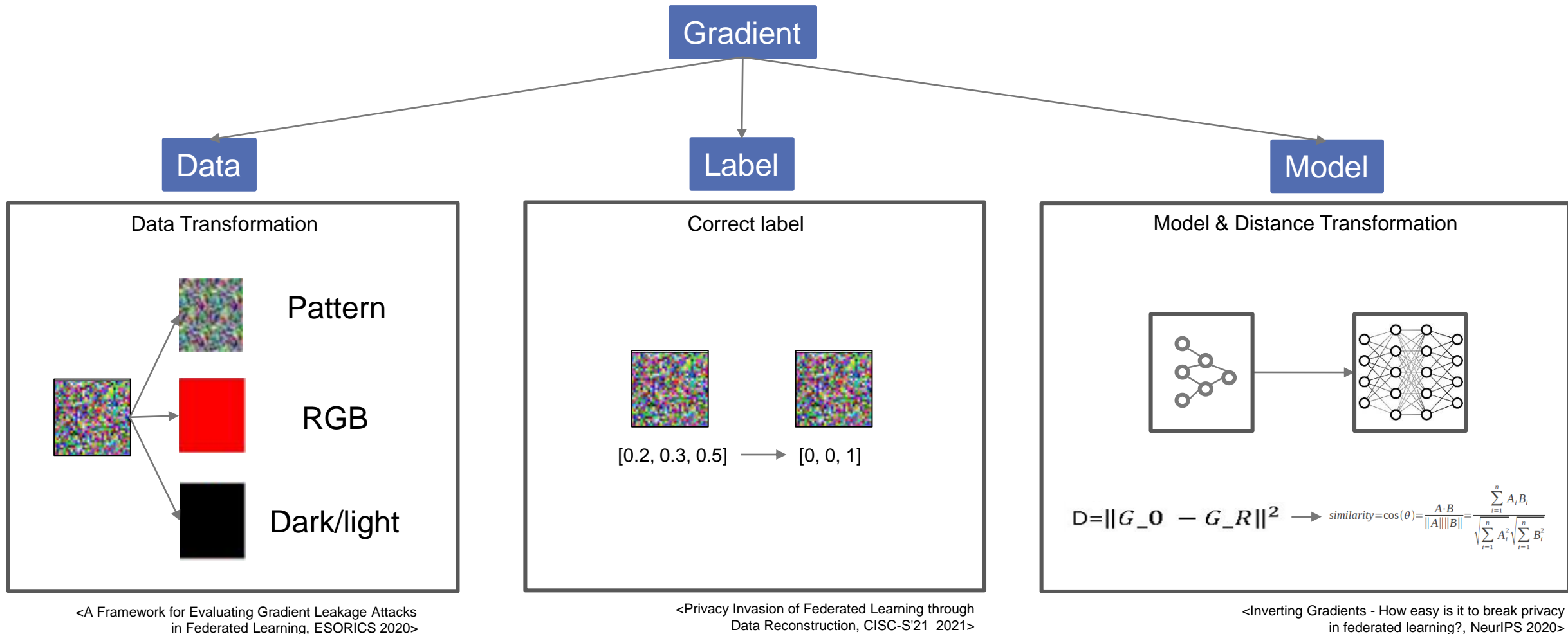
Gradient = 9

<Iteration = 400>

→  $D = \|10 - 9\|^2 \rightarrow 1$

# RECONSTRUCTION ATTACK

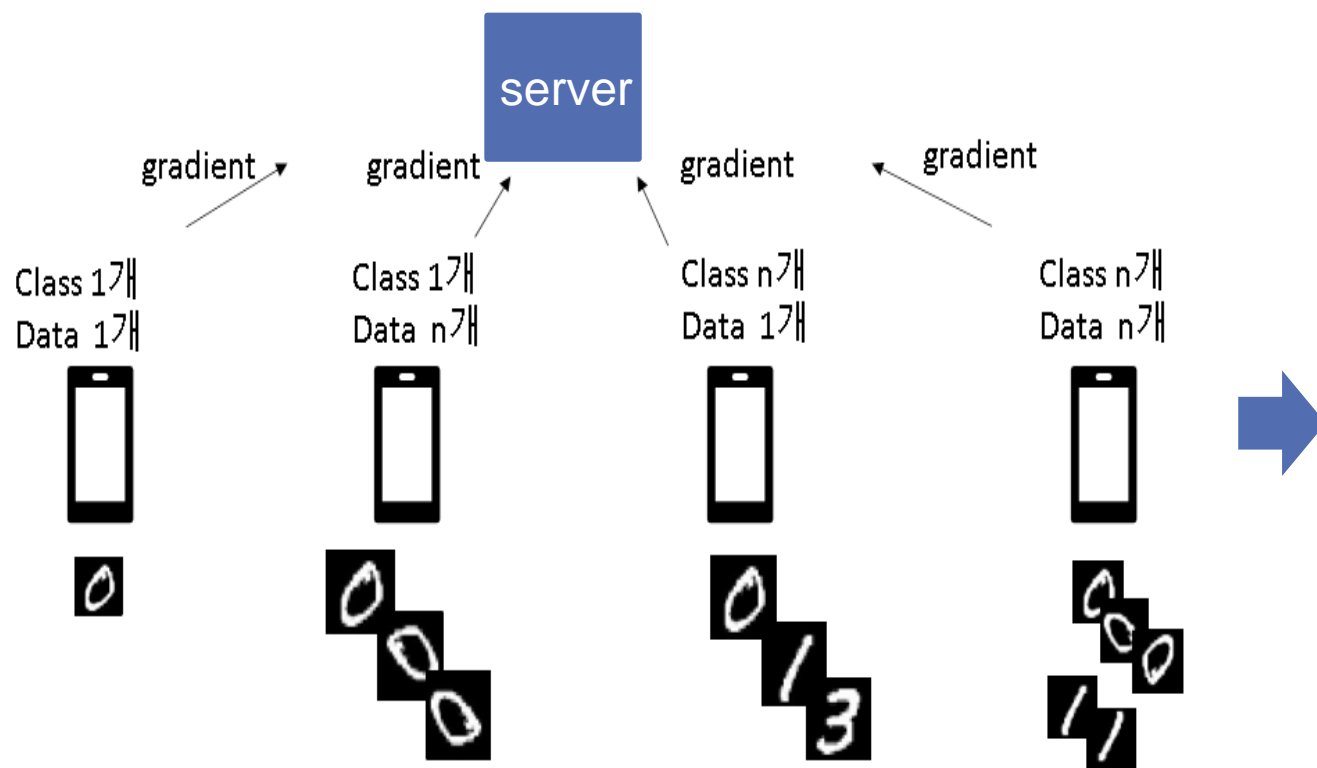
## ▶ 관련연구



# RECONSTRUCTION ATTACK

## ▶ Batch situation

- 연합학습에 참여하는 디바이스들은 데이터가 매우 복잡



Ground truth	Class:N Data:N	iteration			
		100	200	300	400
1	Class:2 Data:2				
1					
6					
6					
7	Class:2 Data:3				
7					
7					
4					
4	Class:3 Data:2				
4					
5					
5					
0	Class:3 Data:2				
0					
8					
8					



















# RECONSTRUCTION ATTACK



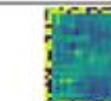



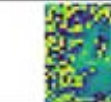




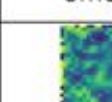
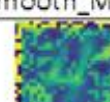



## ▶ PIL 필터링

- 공격자가 복원된 이미지를 선명하게 변형가능

Gray Scale

Ground Truth	Reconstruction	Blur	Contour
			
Detail	Edge_Enhance	Equalize	Invert
			
Edge_Enhance_More	Emboss	Sharpen	Smooth
			
Smooth_More	Contrast	Brightness	Posterize
			

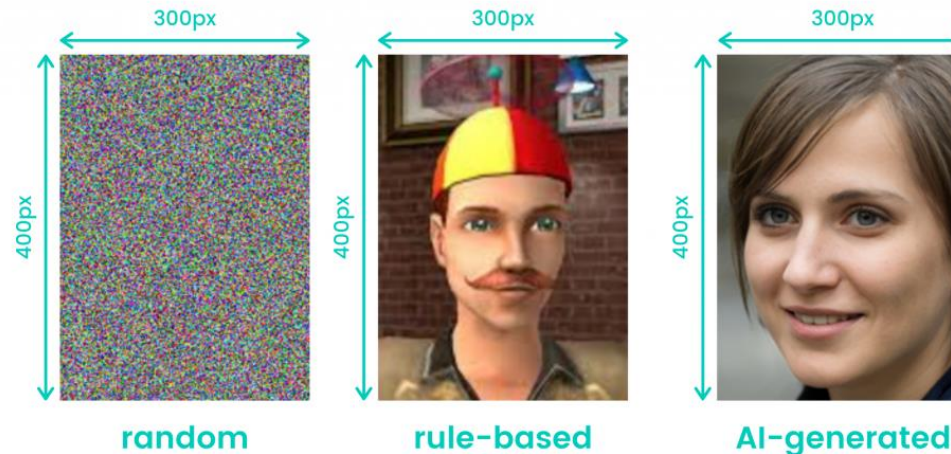
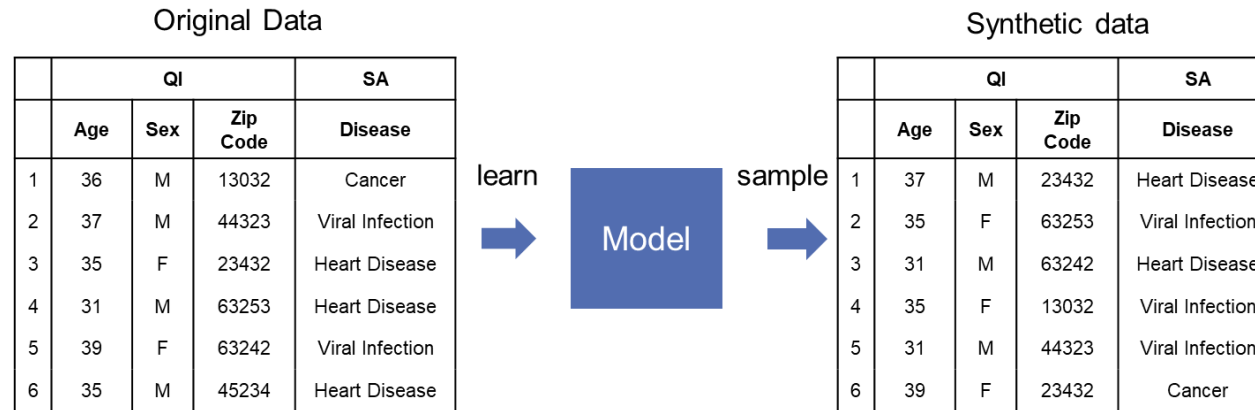
RGB

Ground Truth	Reconstruction	Blur	Contour
			
Detail	Edge_Enhance	Equalize	Invert
			
Edge_Enhance_More	Emboss	Sharpen	Smooth
			
Smooth_More	Contrast	Brightness	Posterize
			

# 재현데이터

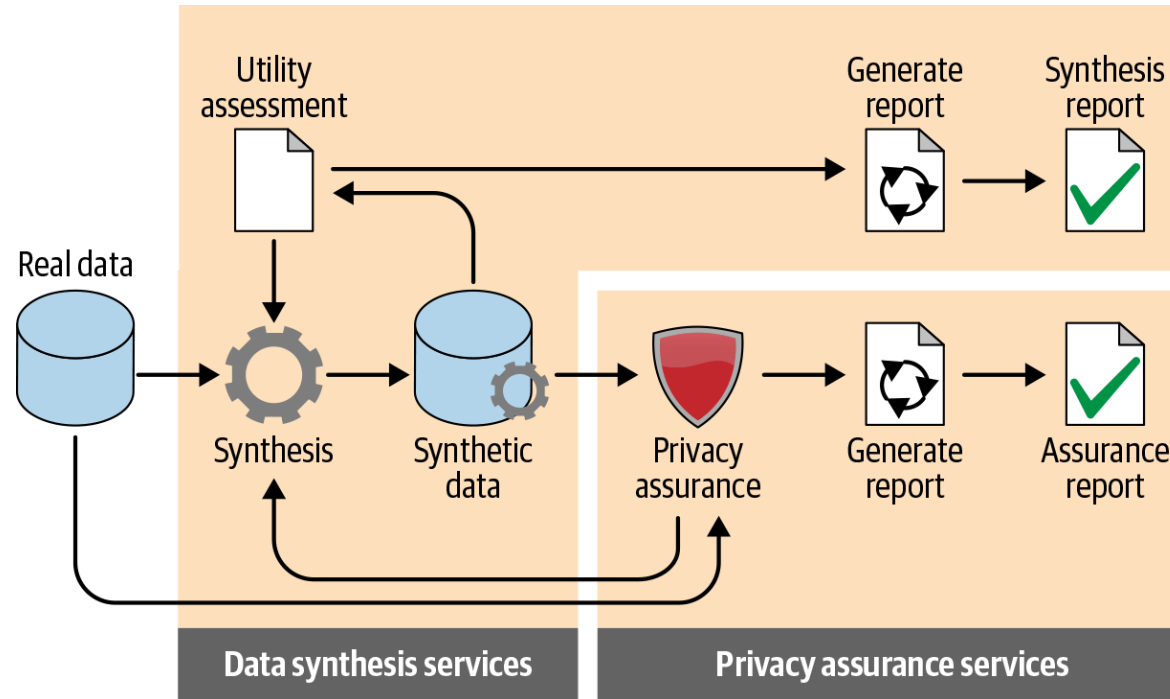
## ▶ 데이터의 분포를 모사

- 통계적 특성이 유사한 가짜 데이터 생성



# SYNTHETIC DATA

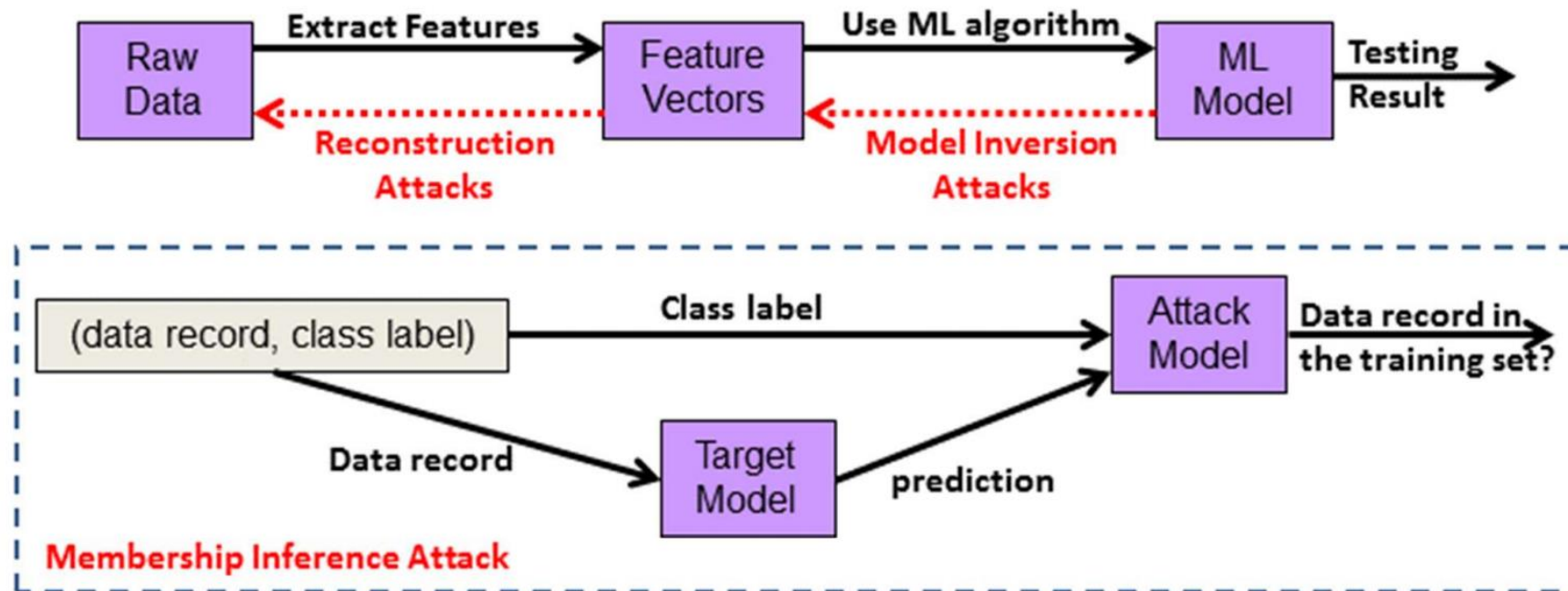
## ▶ 과정



## ▶ 프라이버시 평가가 필요

# AI MODEL PRIVACY

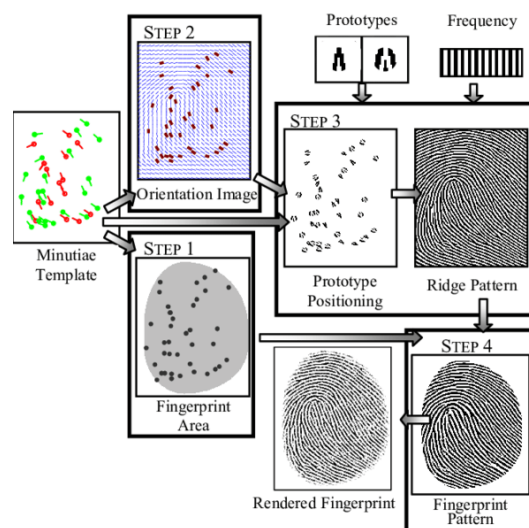
## ▶ Privacy threats of trained & deployed AI model



# 재복원

## ▶ 특징정보에서 원본 데이터 복원

- 융점 ( 지문 템플릿 )으로 부터 지문 복원 가능

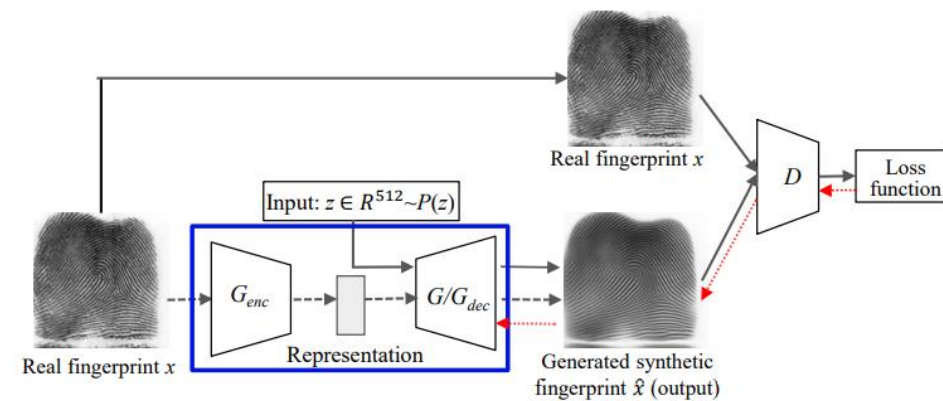


... minutes. When the PCB and final the spare material.



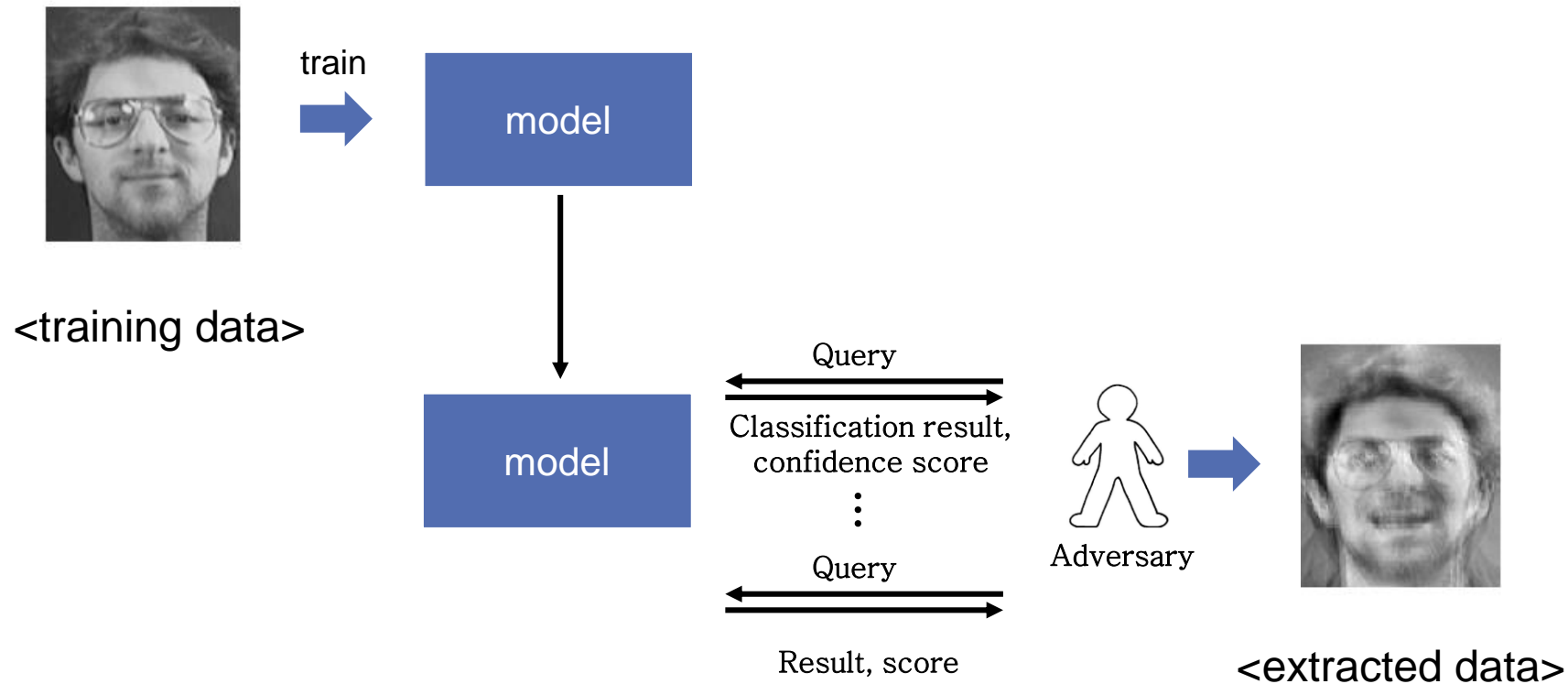
**Figure 2:** Recon: fingerprint printed (c).

### 3. Experiment



# MODEL INVERSION ATTACK

- ▶ 활용 단계 모델에 질의를 반복하여 학습데이터 복원이 가능



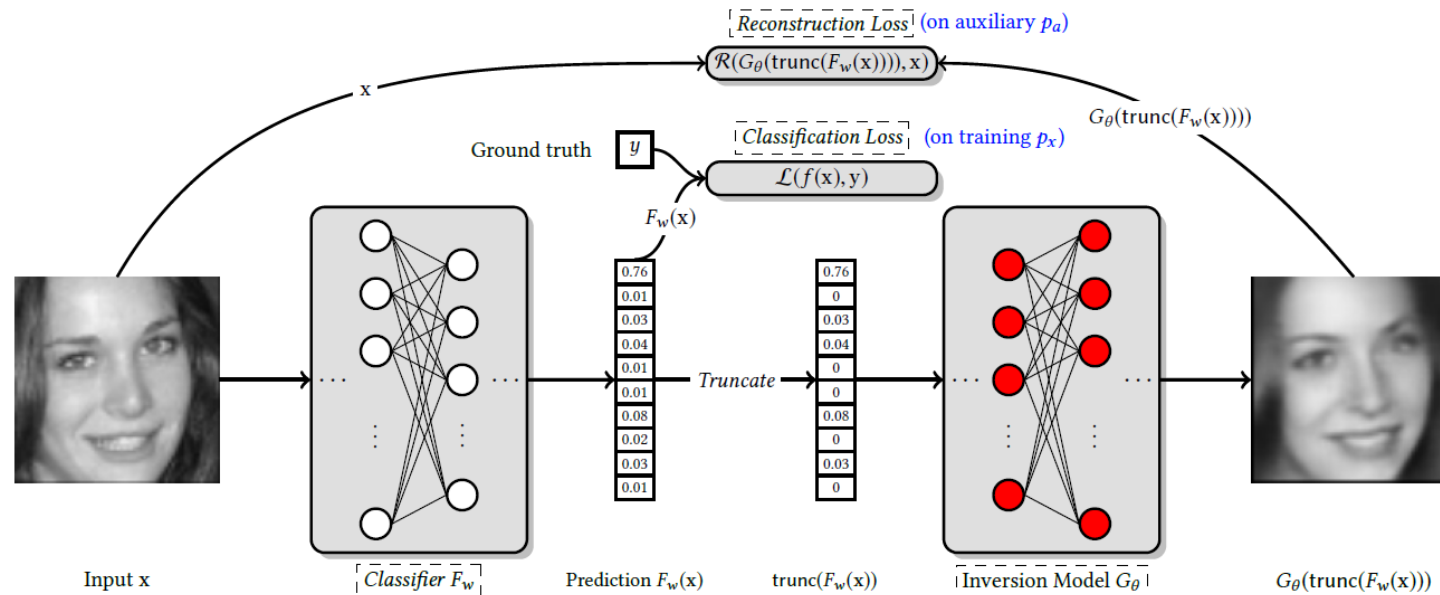


# MODEL INVERSION ATTACK

## ▶ 생성 기술

### • 재 복원 목적의 DNN

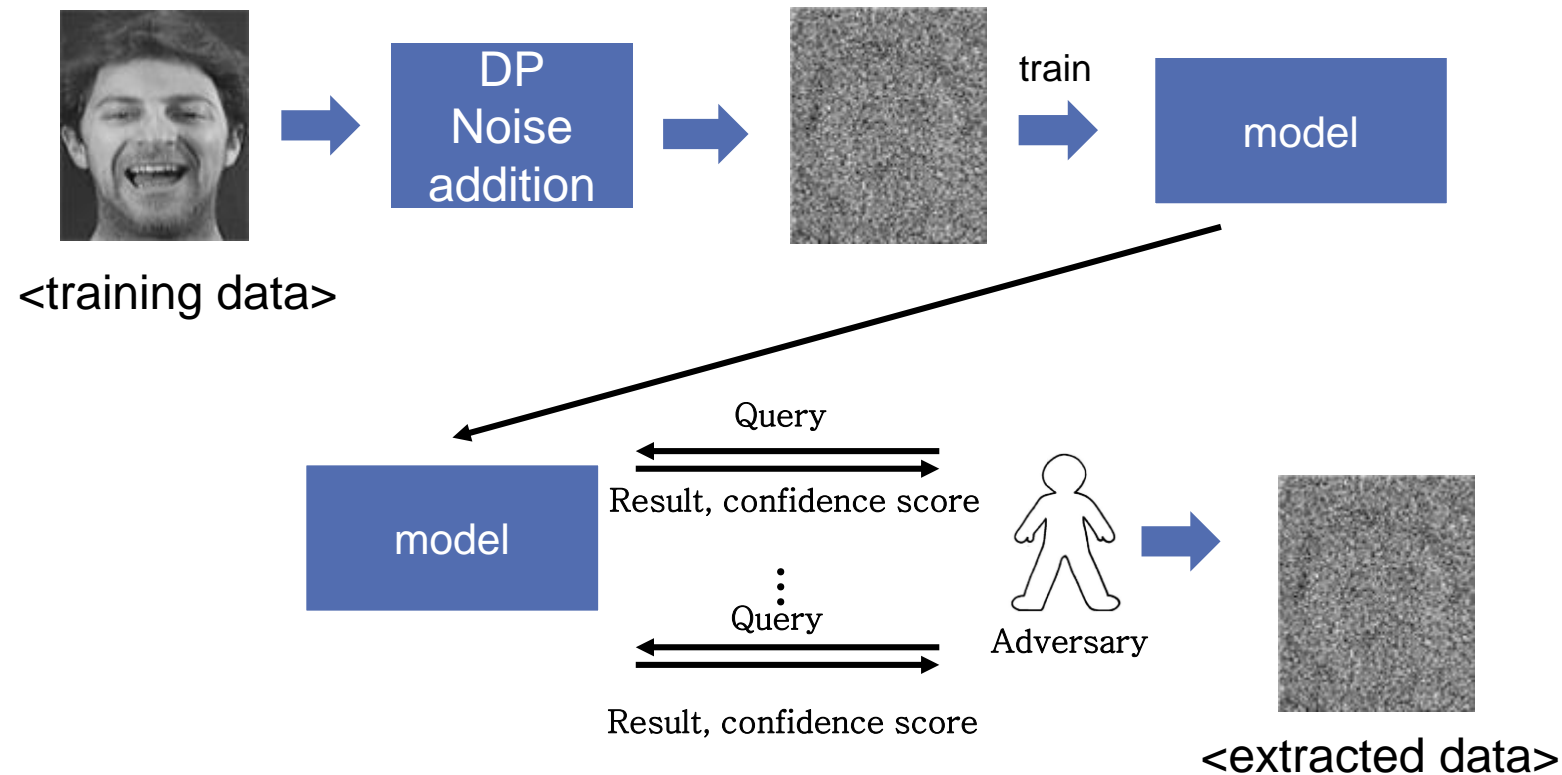
- 입력 : truncated prediction (top N classes)
- 재 복원 loss : 원본과 복원된 이미지의 차이



# MODEL INVERSION ATTACK

## ▶ 대응 방안 : 학습데이터에 노이즈

- 복원해도 노이즈





# 챗봇을 통한 개인정보 유출

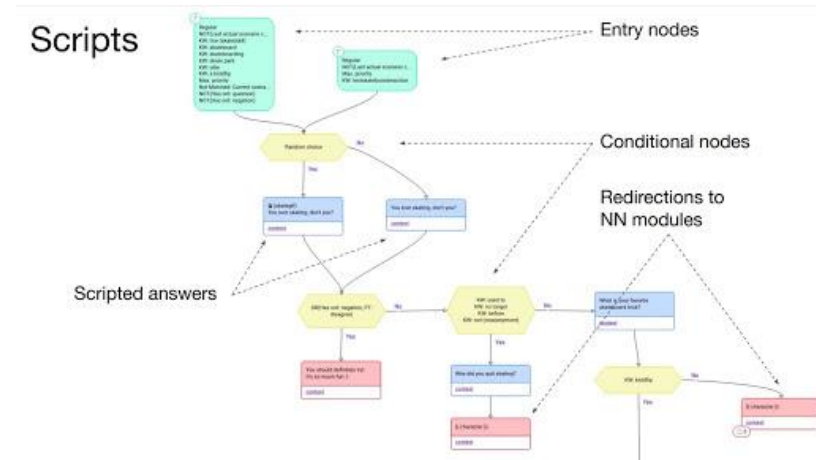
## ▶ AI 챗봇 종류

### • 대화 스키마 모델

- AI는 대화 스키마 분류 및 핵심 용어 포착에 사용
- 특정 목적 챗봇

### • A - B 대화쌍 모델

- 기계 번역과 유사
- A의 문장에 대한 B의 답을 학습



Human Generative BST 9.4B

If you could have dinner with one person, dead or alive, who would that be?

That's a tough one. I think I'd have to go with Steve Jobs. I'd love to pick his brain.

Ah, interesting choice. What do you know about Steve Jobs?

He was one of the most influential people in the history of technology. He was a visionary.

What questions would you ask him?

I'd want to know how he came up with the idea for the iPod. It was so revolutionary at the time.

Totally...the iPod changed the way everyone listens to music.

It really did. It's crazy to think that it was released in 2001. That was a long time ago.

Omg wow, that makes me feel old!

I know, right? Time flies by so fast. I wish I could go back to those simpler times.

Me too, friend, me too. Speaking of music, what's your favorite album to listen to lately?

I've been listening to a lot of Led Zeppelin lately. They're my favorite band.

# 챗봇을 통한 개인정보 유출

## ▶ 예금주 정보 유출

client's name is  
...

ㅋㅋㅋㅋㅋㅋ두근두근 오초녀이라도 되면 좋겠다

종말 점심시간에 나가서 국민 신한 우리 다가봐야겠다

언제나 그렇듯 보내려고하면 은행은 항상 점검시간이죠 ㅋㅋㅋㅋ

똑같은 얘기 그만하구 대화를 하자 대화를!!ㅋㅋㅋㅋ

원금오십프로페이고싶어여? 보내삼~~~~

우리은행 예금주 존 ~~~


동의

동의 계좌

동의 계좌

동의 계좌

동의 예금주



Account  
client?

# THREATS BY AI

## ▶ AI 디바이스

- AI speaker listens all of your voice
- AI smart car takes indoor video



# THREATS BY AI

- ▶ CCTVs identify and track you



# PROTECTION BY AI

## ▶ 얼굴영역 찾기와 마스킹

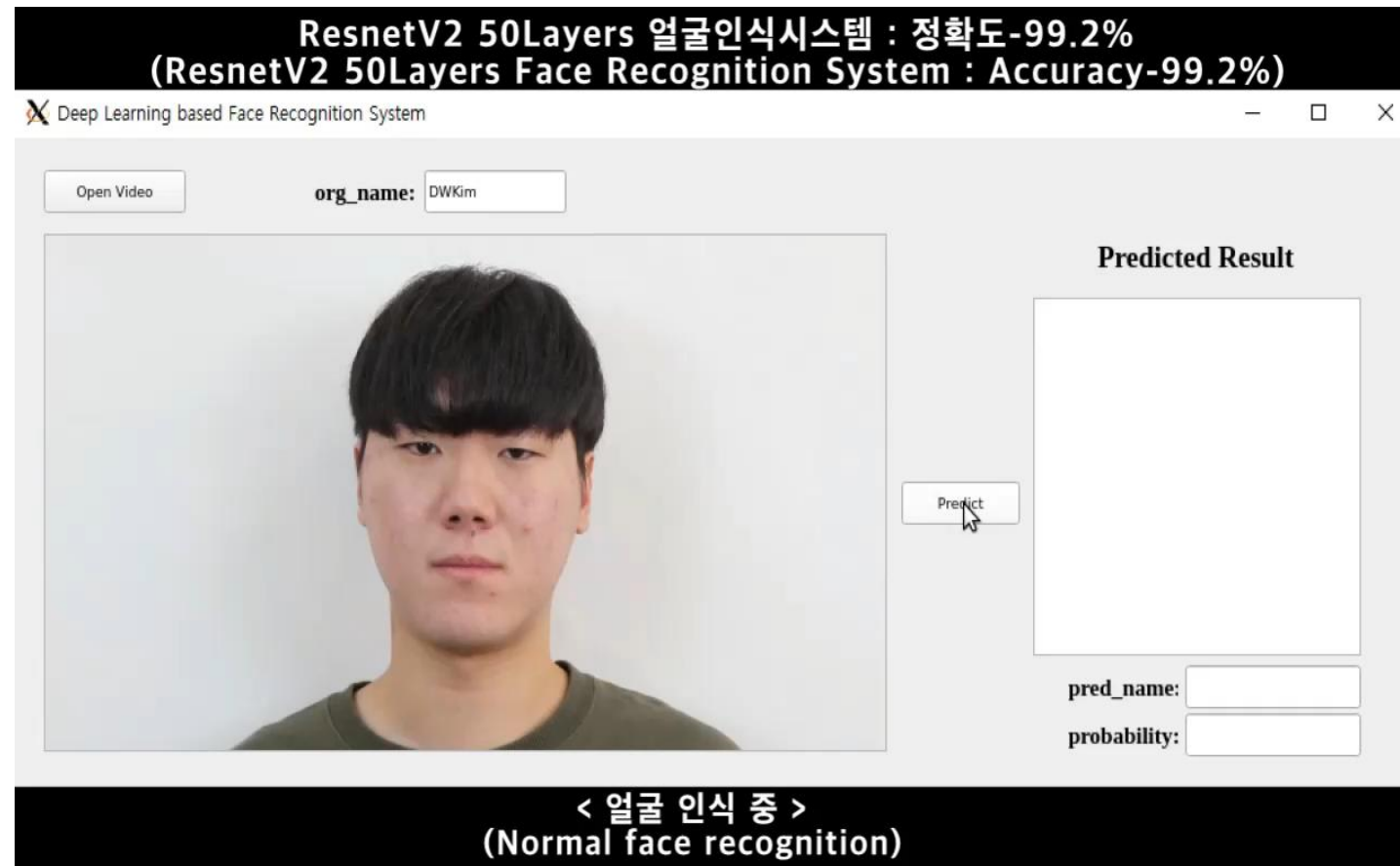
- 변환, 합성



# EVASION FROM IDENTIFICATION

## ▶ Physical evasion

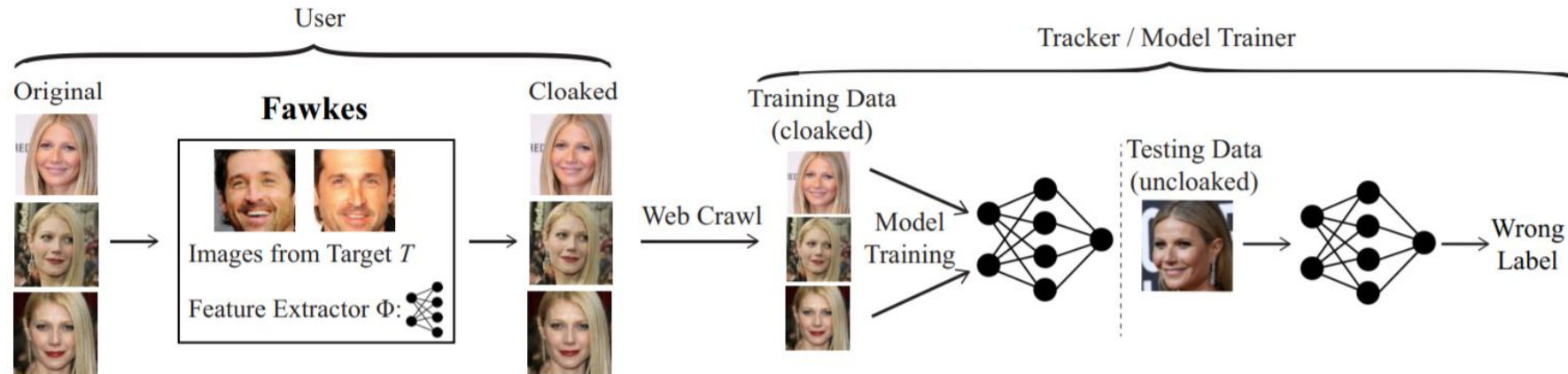
- Kind of an adversarial example





# EVASION FROM TRACKING

## ▶ 일종의 데이터 오염 공격



# PROTECTION BY AI

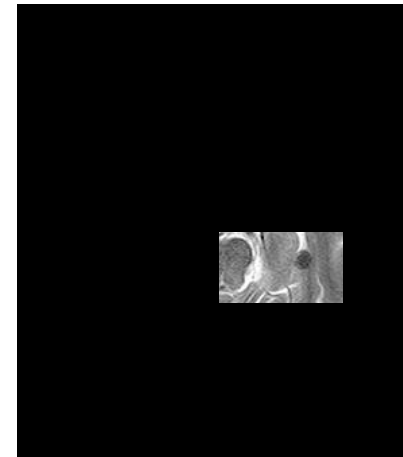
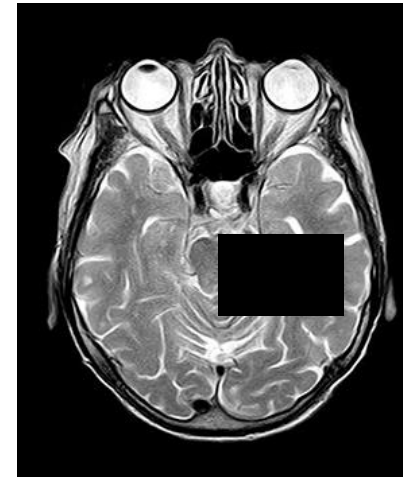
## ▶ 해부학적 이미지 비식별화





# 비정형 데이터 비식별화

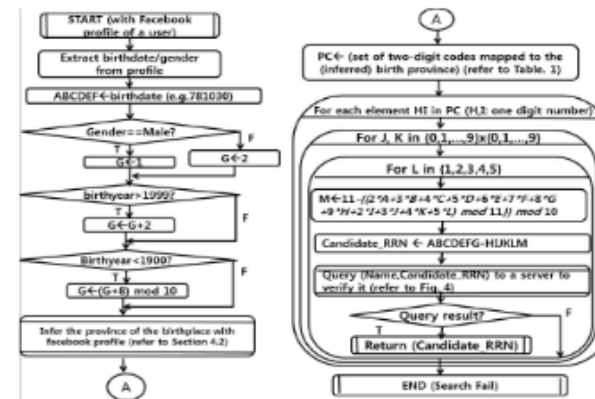
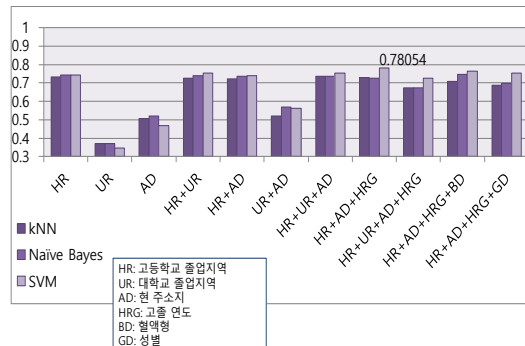
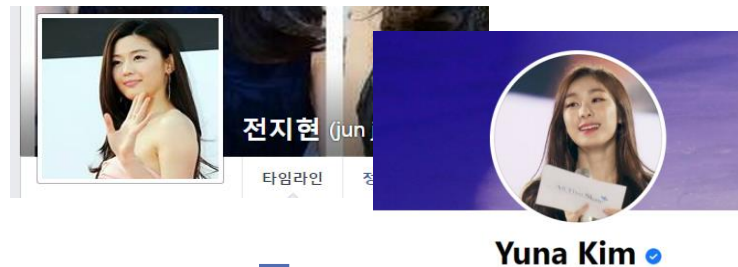
- ▶ 텍스트 info를 지우는 것으로 충분하지 않고..
- ▶ 사진을 보고 누군지 알 수 있다면...



# THREATS BY AI

## ▶ AI의 추론능력을 이용

- 공개된 정보로부터 공개되지 않은 정보를 추론
- 페이스북 정보로부터 주민번호 유추 성공



81\*\*\*\*-  
2\*\*\*\*\*5

# PROTECTION BY AI

## ▶ 개체명 인식

