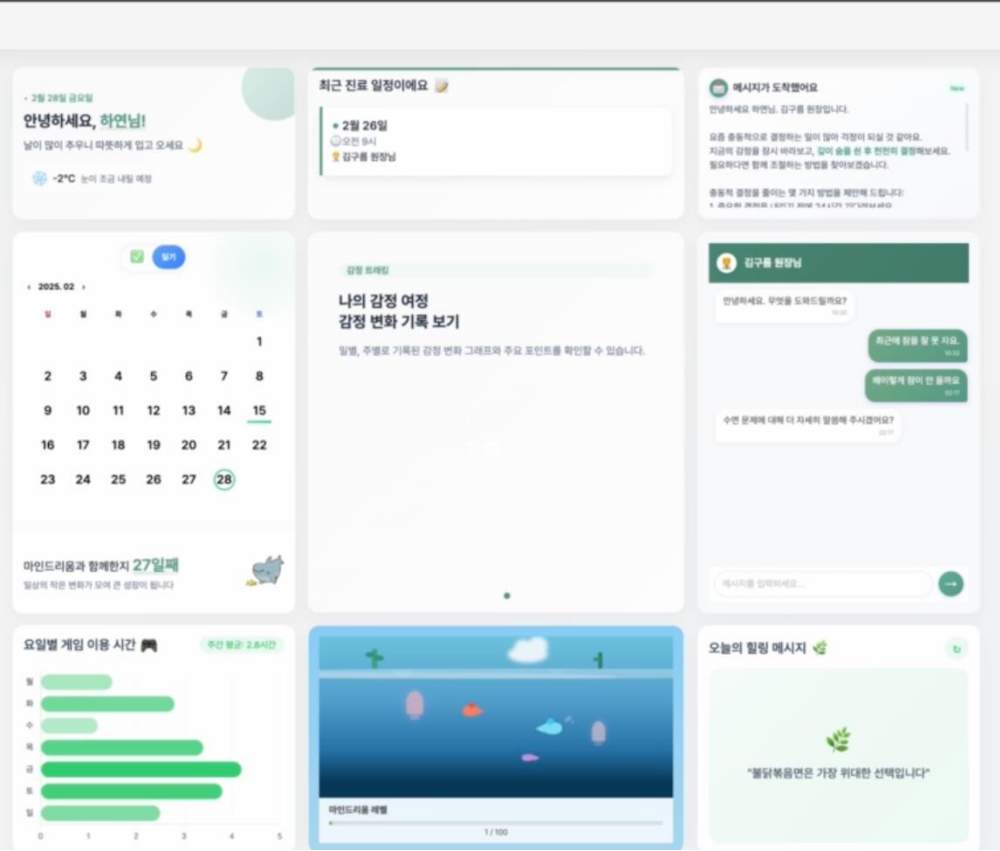
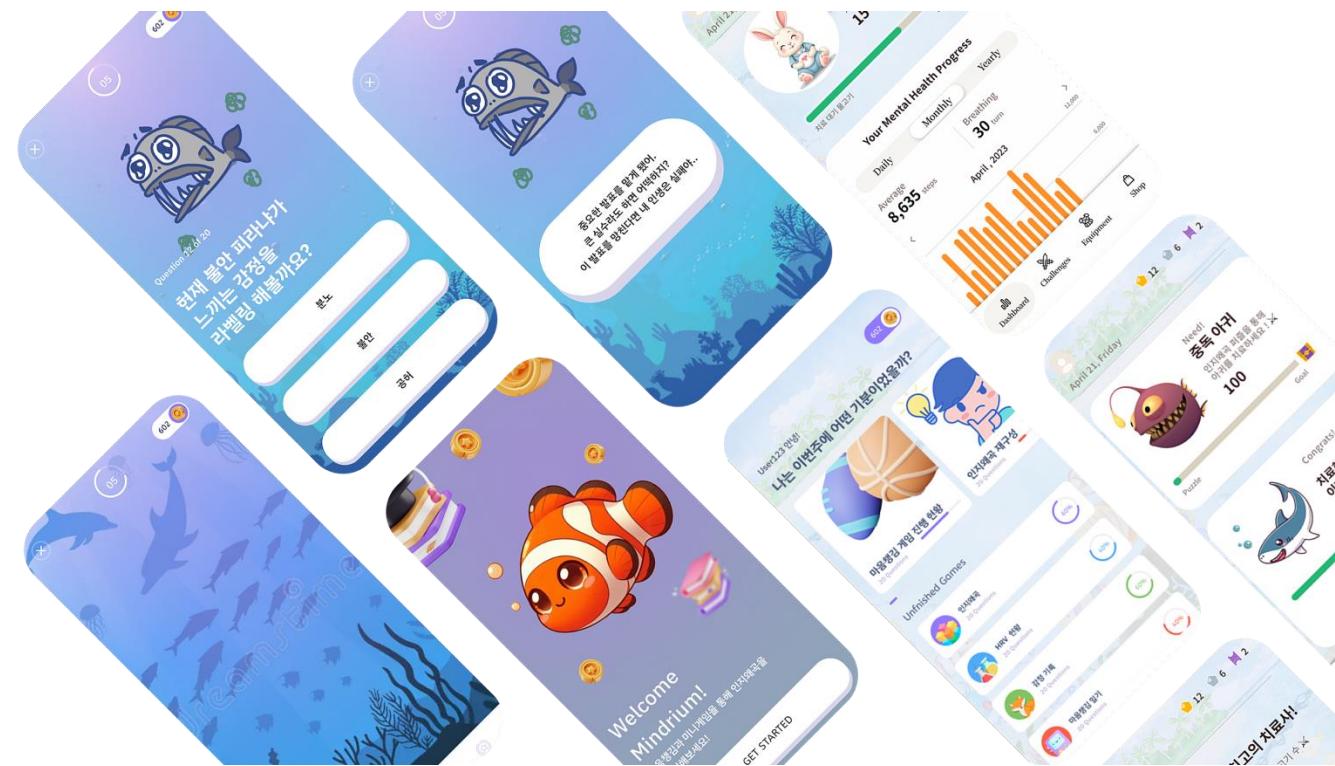


# Preferences of the Public for Sharing Health Data : Discrete Choice Experiment

성균관대학교 인공지능융합학과 석사과정 양하연

# 기존 연구 소개



# 건강 데이터 공유에 대한 대중의 선호도: 이산 선택 실험

일반 대중이 건강 정보를 제3자와 공유할 때 어떤 조건을 가장 중요하게  
생각하는지를 이산 선택 실험을 통해 분석한 연구

## 논문 선정 이유

응답자들은 데이터 공유 시  
제공하겠다 or 거부하겠다는 가능성,  
데이터 사용 목적, 검토절차의 존재를 가장 중요하게 여김

⇒ 사용자에게 “내 데이터가 어떻게, 왜 사용되는가”를 명확히 설명하고,  
거절할 수 있는 기회를 주는 투명한 UX 구조 설계

⇒ 논문 구조를 따라, 플랫폼 내에서 수집하는 데이터 유형, 사용자 유형,  
개입 목적 등에 따라 속성-수준 설계 → DCE 검토 가능

# 연구 목적 및 방법



## 연구 목적

본 연구의 목적은 다양한 상황에서 건강 정보를 공유하는 것에 대해 북유럽 국가들 일반 대중의 선호를 밝히는 것



## 연구 방법

이산 선택 (Discrete Choice) 실험에서 응답자들은 여러 선택 과제를 수행했다. 각 과제는 가상의 상황을 기술하며, 해당 상황에서 데이터 공유가 수용 가능한지

여부를 묻는 형식. 잠재 클래스 로지스틱 회귀모델을 사용해 속성 수준 추정값과 선호의 이질성을 도출하였다. 추정값을 기반으로 각 속성의 상대적 중요성과 다양한 데이터 공유 상황에서의 예측 수용 가능성을 계산하였다.

# 연구 결과 요약

5199

총 설문 수

전체 설문 중 1967개(37.83%)가 최종 분석에 사용됨

4

잠재 클래스 수

클래스 1: 27%, 클래스 2: 32%, 클래스 3: 23%, 클래스 4: 18%

85%

최고 수용률

가장 선호되는 속성 조합의 수용 확률

14%

최저 수용률

가장 비선호되는 속성 조합의 수용 확률

자신의 데이터 사용에 대한 정보를 제공받고,  
공유를 거부할 수 있는 선택지를 제공하는 것을 가장 선호

가장 중요한 속성

'데이터가 공유된다는 사실을 응답자가 인지하는가' 여부  
데이터 공유에 대해 '동의(opt-in)' 기회를 제공받는 것보다  
'거부(opt-out)'할 수 있는 기회를 선호

# BACKGROUND: 디지털 기술 발전과 건강 데이터



## 디지털 기술 발전

지난 20년간의 디지털 기술 발전은 건강 데이터를 디지털 방식으로 수집, 활용, 공유하는 데 있어 상당한 성장을 가져왔다.



## 의료 분야 변화

이러한 변화는 전자 건강기록, 유전자형 분석, 바이오뱅크, 모바일 기기를 통한 자가 추적 애플리케이션의 개발 및 채택에 기인한 바가 크다.



## 데이터 의존도 증가

보건의료, 의학 연구, 기술 및 제약 기업과 같은 다양한 분야는 건강 관리와 새로운 의학·기술 제품 개발을 위해 디지털 방식의 데이터 수집 및 공유에 점점 더 의존하게 되었다.



## 개인 역할 변화

이러한 흐름은 개인들이 건강 정보를 스스로 탐색하고, 새로운 건강 관련 웹사이트 및 모바일 앱에 접근함으로써 자신의 건강을 관리·증진하는 데 보다 적극적인 역할을 수행하게 만들었다.

# 최종 DCE에 포함된 모든 속성과 수준 목록

## 건강 정보 수집 주체

- 사용자가 자신의 휴대폰이나 컴퓨터를 통해 사용한 서비스, 프로그램 또는 애플리케이션을 제공하는 기술 회사.
- 사용자가 참여했던 학술 연구 프로젝트
- 병원이나 일반 개원의 등 사용자의 진료를 위해 건강 정보를 수집한 보건의료 제공자.

## 데이터 사용자

- 건강 앱을 개발하는 기술 회사.
- 신약을 개발 및 제조하는 제약 회사.
- 인간 건강에 대한 가설 및 이론을 검증하여 새로운 지식을 생산하는 학술 연구 프로젝트.

## 데이터 사용 목적

- 의료기기, 약물, 휴대폰 앱 또는 새로운 건강 서비스나 프로그램과 같은 신제품 또는 서비스 개발.
- 제품이나 서비스의 홍보, 광고 또는 마케팅.
- 정책 제안 조사.
- 데이터 사용자의 제품 또는 서비스 품질 평가 및 향후 자원 분배 계획 수립.

정보 제공 및 동의: 사용자의 거가 정보가 공유되는 경우, 사용자에게 정보가 제공되는지 여부



# 최종 DCE에 포함된 모든 속성과 수준 목록

## 정보 제공 및 동의

사용자의 건강 정보가 공유되는 경우, 사용자에게 정보가 제공되는지 여부.

- 건강 정보가 새로운 맥락에서 공유되고 사용되더라도 사용자에게는 통보되지 않는다.
- 건강 정보가 새로운 맥락에서 공유되고 사용된다는 사실을 사용자에게 통보한다.
- 건강 정보가 공유되고 사용된다는 사실을 사용자에게 통보하고, 거부(opt-out)할 수 있는 기회도 제공한다.
- 사용자에게 통보하며, 공유 및 사용에 대해 동의(consent)를 요청한다.

## 데이터 공유 검토 절차

- 데이터 공유에 대한 검토는 없음
- 위원회가 새로운 맥락에서 건강 정보를 공유하는 것에 대해 검토
- 위원회가 새로운 맥락에서 건강 정보를 공유하고 사용하는 것 모두에 대해 검토

# Method



## Bayesian D-efficient design

선택 실험에서는 신뢰할 수 있는 모수 추정을 위해 Bayesian D-efficient design를 사용  
설계는 NGene를 이용하여 개발



## Pilot test

추정값을 예측하여 prior로 설정  
이때 Halton draw 500회, 반복 1000회를 적용하였다. 설계에서는 속성 간 상호작용이 없다고 가정하였다. 또한, 각 수준이 동일한 횟수로 등장하도록 level balance이 최적화되었다. 파일럿 설계의 D-error는 0.31이었다.



## 선택 과제 구성

총 28개의 고유한 선택 과제가 생성되었고, 이는 4개의 블록으로 나뉘었다. 응답자들은 무작위로 하나의 블록에 배정되어, 7개의 선택 과제를 수행하였다.

# 설문지 구성

## 인구통계학적 특성

연령, 성별, 교육 수준, 자기보고 건강 상태, 만성질환 보유 여부 등 인구통계학적 특성에 관한 질문들로 구성

또한, e-Health Literacy Scale (eHEALS)을 사용하여 응답자의 건강 정보기술 활용 능력을 평가하였다

8개 항목으로 구성되어 있으며, 각각은 '전혀 동의하지 않음'부터 '매우 동의함'까지 5점 척도로 응답하게 되어 있다.

## DCE 항목

각 참가자에게는 특정 상황이 제시되었고, 이를 수용할지 거절할지를 선택

따라서 최종 DCE는 32개의 고유한 선택 과제로 구성되었으

며, 이를 4개 블록으로 나누어, 각 참가자는 두 가지 건강 정보 유형에 대해 8개 선택과제, 총 16개 과제에 응답하였다.

## 신뢰 및 태도 평가

다양한 기관 및 사람들에 대한 신뢰, 신기술에 대한 태도, e-헬스 리터러시 자기평가 항목으로 구성

# Method

## Random utility theory

응답자들은 매 과제마다 해당 상황을 accept 여부를 선택해야 했다.

이들의 반복된 선택을 통해 선호를 도출할 수 있었음.

DCE는 개인이 각 선택 과제에서 직면한 선택 항목으로부터 utility을 도출한다고 가정하는 utility theory에 기반한다.

## Attribute-level estimates

속성 수준 추정치를 비교함으로써 각 속성의 상대적 중요성을 판단하였으며,

실험에서 도출된 추정값을 바탕으로 다양한 데이터 공유 상황의 효용과 수용 가능성도 계산할 수 있다.

	Situation 1 of 8	
<p>The organisation collecting my information is</p> <p>They will share it with</p> <p>The reason they want to use my information is to</p> <p>When they share my information, I will</p> <p>There will be</p>	<p>an <b>academic research project</b>.</p> <p>a <b>technological company</b>.</p> <p>develop a <b>new product or service</b>.</p> <p>be offered an <b>opt-out and information</b>.</p> <p>a committee that <b>reviews the sharing</b> of information.</p>	
Do you think this situation is acceptable?	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No

일반적인 DCE와는 달리 서로 교환 가능한 상품 간 비교가 아닌,  
하나의 상황을 수용할지 말지를 판단하는 것이므로,  
응답자는 각 선택 과제에서 주어진 상황을  
accept 또는 reject하는 선택지

# Statistical Analysis

## 기술 통계

관심 변수들을 요약하기 위해 평균값 및 빈도수와 같은 기술 통계를 사용하였다.

- 문항 중 하나라도 '전혀 동의하지 않음' 또는 '동의하지 않음'에 응답한 사람은 부족한 eHL로 분류하였다.
- 문항 중 하나라도 '동의도 반대도 아님'으로 응답한 사람은 문제 있는 eHL로
- 분류하였다. '동의' 또는 '매우 동의'로 응답한 사람은 충분한 eHL로 분류하였다.

## 잠재 클래스 모델

관측되지 않은, 잠재적 선호 이질성에 따라 응답자를 서로 다른 클래스로 분류하는 분석 방법 사용하여 속성 수준 추정치와 각 속성의 중요도 가중치를 도출



## Binary logit model

설문을 통해 수집된 이산 선택 자료는 각 국가별로 binary logit model을 사용하여 분석  
Swait-Louviere 검정을 통해 국가 간 표본 간의 스케일 차이가 통계적으로 유의한지를 확인하였다.



# 속성의 상대적 중요도 및 수용률 추정

## 속성의 상대적 중요도

DCE를 통해 도출된 속성 수준 추정값을 바탕으로, 각 속성의 상대적 중요도를 계산하였다.

## 수용률 계산식

속성 수준을 기반으로 특정 시나리오를 구성하고, 각 시나리오의 효용 값은 다음과 같이 계산

$$V_{\text{Scenario 1}} = \beta_A + \beta_B + \beta_C$$



## 계산 방법

각 속성마다 효용 값이 가장 큰 수준과 가장 작은 수준의 차이를 계산하여 해당 속성이 전체 효용에 끼치는 영향 범위를 구하였다.

## 수용률 추정

수용률은 특정 데이터 공유 시나리오에 대해 참가자가 선택할 확률을 의미.

# Result

## 통계적 유의성

binary logit model 분석 결과, 모든 속성의 계수는 통계적으로 유의미하였고, 그 부호는 사전 기대와 일치  
응답자들은 다음과 같은 상황을 더 수용 가능하다고 판단하였다

- 1) 데이터 수집 주체가 자신의 의료 제공자인 경우
- 2) 새로운 데이터 사용자가 국가 기관인 경우
- 3) 목적이 의료 서비스의 질 평가인 경우

또한

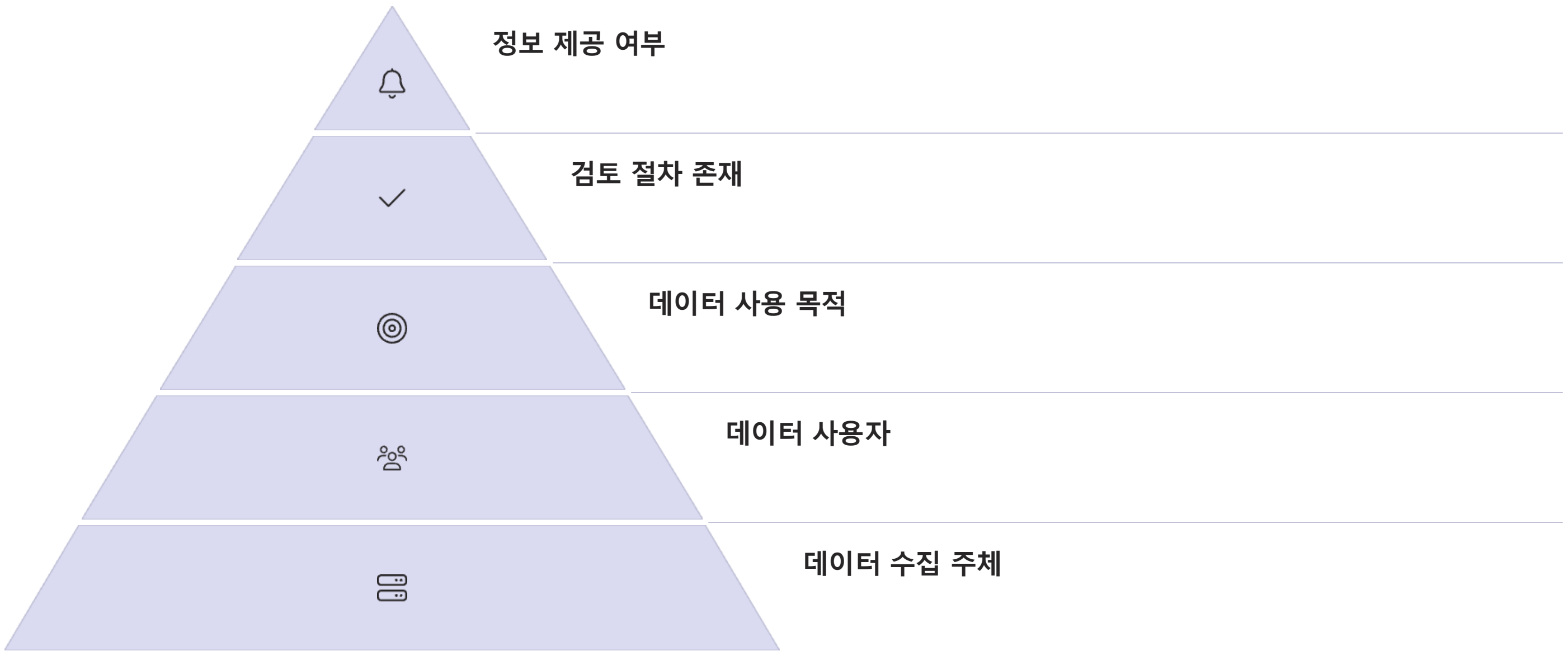
데이터 공유 사실에 대해 통보받는 것을 중요하게 생각하였고,

**거부할 수 있는 기회가 주어지는 상황을 선호하였다.**

정보 공유와 사용에 대해 검토(review) 절차가 있는 상황도 선호되었다.



# 데이터 공유 선호도의 중요 속성

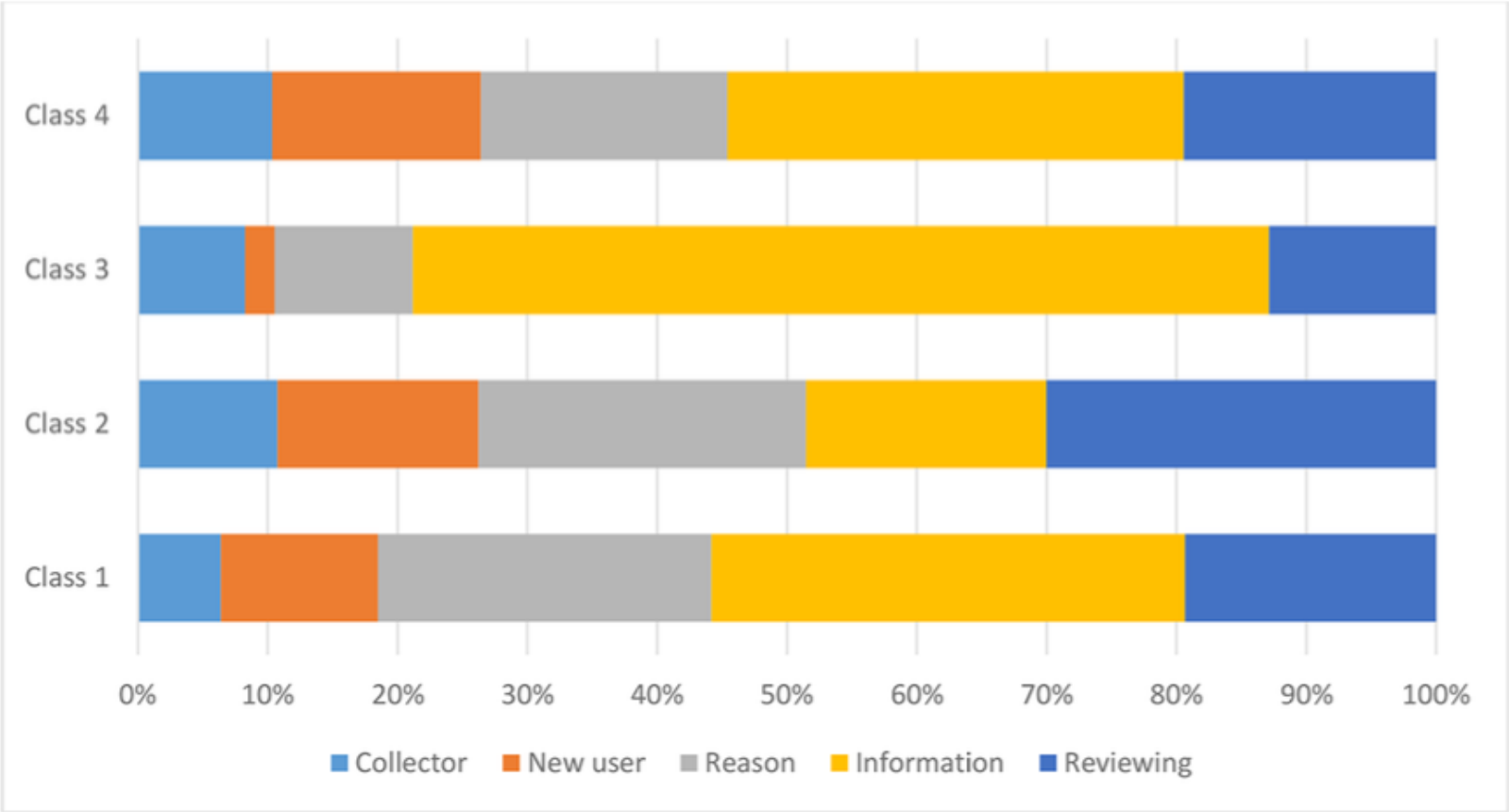


건강 정보가 새로운 사용자에게 이전되는 상황에서, 응답자들은 '정보 제공'을 가장 중요한 요소로 판단하였다. 그 다음으로 중요한 요소는 검토 절차의 존재였다. 스웨덴 응답자는 건강 정보 공유 목적을 다른 나라보다 더 중요하게 여겼으며, 아이슬란드 응답자는 검토 절차를 가장 중요하게 여겼다.

# 잠재 클래스 분석 - 건강 정보 공유에서 중요하게 여긴 속성의 상대적 중요도

4개의 클래스가 가장 적합한 모델로 식별

**Figure 2.** Relative importance score for respondents' preferences stratified using the four-class model. The reason and having a review process in place were most important for class 2. Being informed was most important for classes 1, 3, and 4.



클래스	주요 특징	정보공유태도
Class 1	정보 제공 중시	정보 제공 + 거부 가능성이 있을 때만 수용
Class 2	검토 절차, 목적 중시	공적 목적 + 철저한 검토일 때 수용
Class 3	정보 제공 선호	대부분 수용
Class 4	공유에 비교적 개방적	정보 제공 없어도 수용 가능

# 국가별 선호도 차이

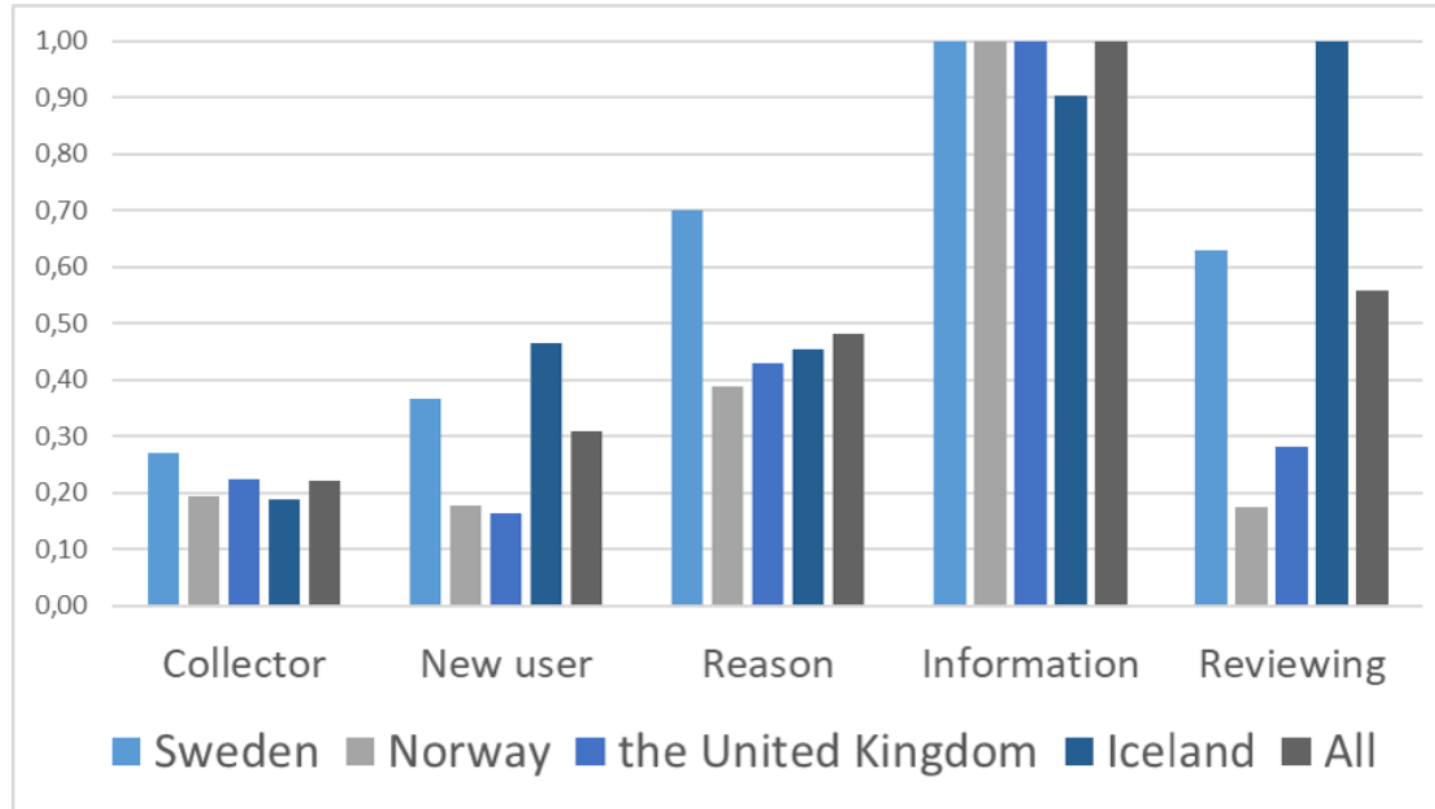
노르웨이	아이슬란드	스웨덴	영국
노르웨이 응답자들은 정보 제공 여부와 사용 목적이 중요하다고 여기는 클래스 1에 속할 가능성이 높았다.	아이슬란드 응답자들은 데이터 공유 검토 및 사용 목적을 중시하는 클래스 2에 속할 가능성이 높았다.	스웨덴 응답자들은 클래스 1 및 클래스 2에 속할 가능성이 높았다.	4개 클래스에 비교적 고르게 분포되었다.

서로 다른 국가의 응답자들은 데이터의 새로운 사용 목적과 검토 절차의 가치에 대해 상이한 선호를 보였다. 자신의 데이터 사용에 대한 정보를 제공받고, 공유를 거부할 수 있는 선택지를 제공하는 것이 가장 선호되는 거버넌스 방식이었다.

# 국가별 응답자들이 건강 정보 공유에서 중요하게 생각하는 속성

was the most important attribute for respondents in Iceland.

**Figure 3.** Relative importance score for all respondents' preferences, stratified by country. Receiving information and having the opportunity to opt out was the most important attribute, on average, followed by review process and the reason for sharing the information.



## 1.Information(정보 제공 및 거부 기회)

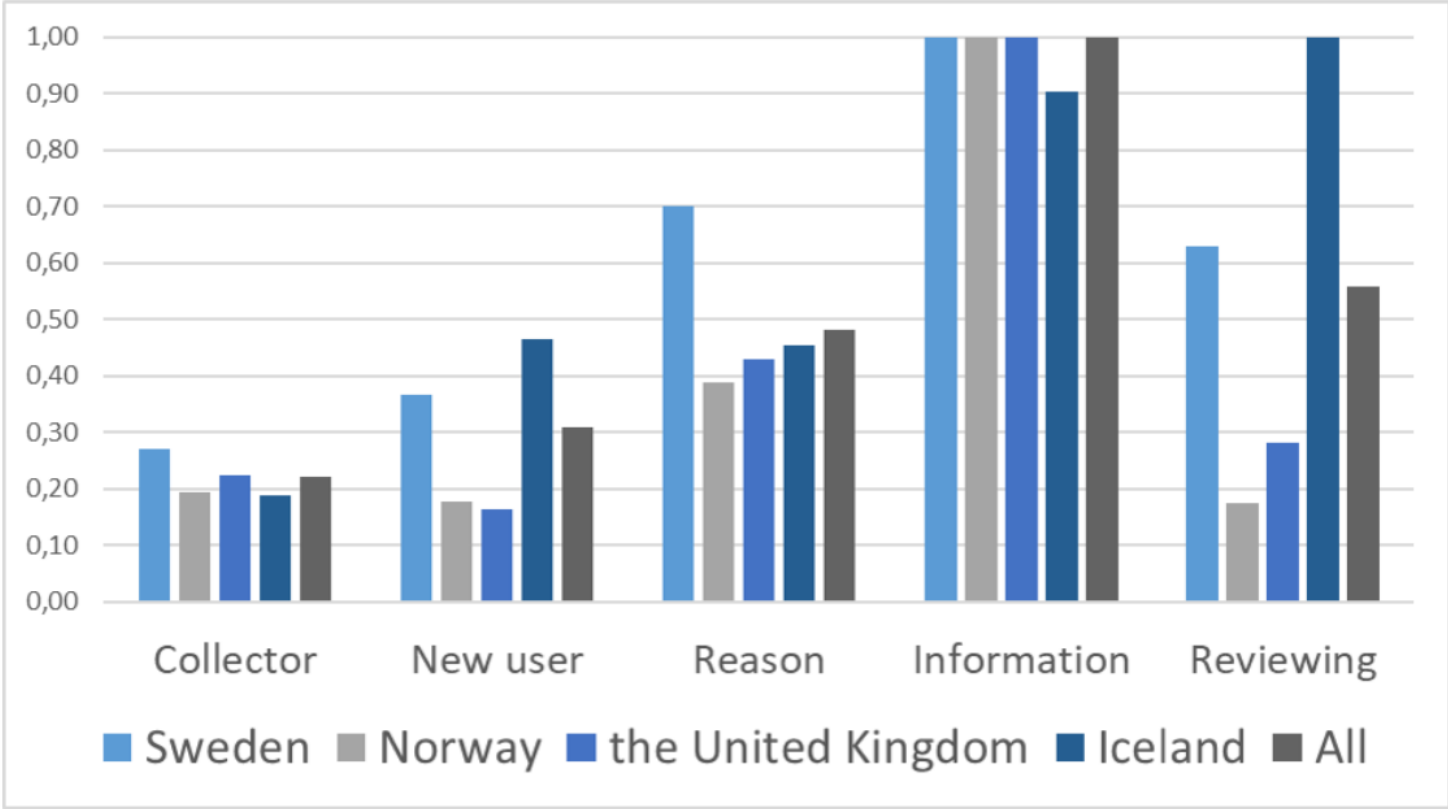
1. 모든 국가에서 가장 중요한 속성으로 나타남
2. 특히 스웨덴, 영국, 아이슬란드에서는 상대적 중요도가 1.0으로 최고 수준
3. 이는 정보 제공 + 거부할 수 있는 권한이 응답자에게 절대적으로 중요함을 의미

## 2.Reviewing(검토 절차)

1. 아이슬란드 응답자에게 두 번째로 중요한 요소 (0.6 이상)
2. 반면, 노르웨이와 영국 응답자는 이 요소의 중요도를 상대적으로 낮게 평가함
3. 전체 응답자 평균은 약 0.45 수준으로, 꽤 중요한 속성으로 인식

# 국가별 응답자들이 건강 정보 공유에서 중요하게 생각하는 속성

Figure 3. Relative importance score for all respondents' preferences, stratified by country. Receiving information and having the opportunity to opt out was the most important attribute, on average, followed by review process and the reason for sharing the information.



### 3. Reason(정보 공유 목적)

- 1. 스웨덴 응답자에게는 두 번째로 중요한 속성
- 2. 정보가 공익적 목적(예: 정책 설계, 건강 연구 등)으로 사용되는지 여부가 선택에 큰 영향을 미침

### 4. Collector / New user

- 1. 전체적으로 덜 중요한 속성으로 평가됨 (0.2~0.4 수준)
- 2. 즉, 누가 데이터를 수집했는가/누가 사용하는가보다는, 왜 어떻게 사용하는가가 더 중요하게 작용

# 연구의 시사점

## 투명성 원칙 지지

본 연구 결과는 EU 일반 개인정보보호법(GDPR)의 기본 원칙 중 하나인 투명성을 뒷받침하는 근거를 제공한다. GDPR은 데이터 수집 행위에 대한 투명성 요건을 강화함으로써 개인이 정보적 자기결정권을 가질 수 있도록 하고 있다.

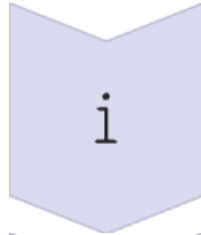
## 국가별 차이 고려

참여 국가 모두 동일한 규제(GDPR)를 따르고 있음에도 불구하고, 본 연구 결과는 국가별로 건강 정보 공유 시 중요하게 여기는 요소가 다르다는 점을 보여준다.

## 거부(opt-out) 선호

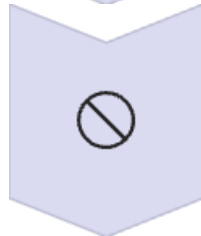
예상과 달리, 응답자들은 동의(opt-in) 기회를 제공받는 것보다 거부(opt-out) 기회를 더 선호하였다. 이는 다른 거버넌스 메커니즘이 병행되는 경우, 단순히 거부할 수 있는 구조만으로도 충분하다고 인식할 수 있음을 시사한다.

# Conclusion



## 정보 제공

데이터 사용에 대한 정보를 제공



## 거부 기회

원하지 않을 경우 거부(opt-out) 기회 제공



## 검토 시스템

데이터 공유 및 사용에 대한 검토 절차 마련



## 국가별 맞춤 접근

국가별 선호 차이를 고려한 정책 수립

디지털 시대의 정책 결정에서, 대중의 데이터 공유 수용 여부는 점점 더 중요한 요소가 되고 있다. 본 연구는 대중이 건강 정보를 공유할 때 어떤 요소를 중요하게 여기는지에 대해 사이버 보안 및 개인 정보보호 분야에 실질적 통찰을 제공한다. 사람들에게 데이터 사용에 대한 정보를 제공하고, 거부(opt-out)할 기회를 주는 것이 응답자들의 데이터 공유 수용 의향을 가장 높이는 요소였다. 또한, 검토 시스템의 존재도 매우 중요했다. 참여 국가 간에는 선호에 차이가 존재하였으며, 동의 또는 거부 구조에 대한 향후 심층 연구가 필요하다.

# 인사이트

DCE 기반으로 “사용자들이 어떤 데이터 공유 구조를 선호하는가?”

어떤 형태의 데이터를 사용하는 것이 사용자에게 분석 결과 전달력이 높은가?

- 수면 기록, 감정 일기, 음성 데이터 등

그래프 형태, 영상, 텍스트 형태 중 어떤 방식이 사용자에게 높은 전달력을 가질 수 있는가?