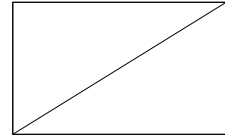


공 개



의 안 번 호	제 3 호
의 결 연 월 일	2025. 2. 20. (제 3 회)

심  
의  
사  
항

AI 데이터 확충 및 개방 확대방안

국가인공지능위원회

제 출 자	개인정보보호위원회 위원장 고학수
제출 연월일	2025. 2. 20.

## 1. 의결주문

- AI 데이터 확충 및 개방 확대방안을 별지와 같이 의결함

## 2. 제안이유

- AI 기반 디지털 전환 가속화 시대에 대응하고 AI에 대한 경쟁력 확보를 위해 양질의 데이터를 확충하고 활용을 촉진하는 정책을 수립하려는 것임

## 3. 주요내용

### 가. 배경 및 문제점

- ☐ AI 기술 발전 및 경쟁력 강화를 위해 데이터의 중요성이 강조되고 있으나, 여전히 AI를 위한 데이터 활용 장벽 존재
- AI 시대에 대응한 데이터 이용 관행 및 제도의 총체적 업그레이드 필요

### 다. 추진전략

- ☐ AI 개발 촉진을 위한 고품질 데이터 제공 확대
  - 비정형 원본 데이터 활용 확대(자율주행 → 재난예방, 국민안전 등), AI 업체가 장기간 데이터를 활용할 수 있도록 특례 마련
  - 원본 활용이 어려운 데이터는 합성데이터를 통해 AI 학습에 활발하게 활용될 수 있도록 안전성 평가 등 기반 확충('AI허브', '공공데이터포털' 등을 통해 개방)
  - 분야별 특화 고품질 데이터 구축·개방 확대, 제조 데이터 표준 및 공정별 인증제도 도입, AI 수요가 높은 국가중점데이터 적극 개방
- ☐ 공공부문 데이터 활용 여건 획기적 개선
  - 개인정보 포함 공공데이터를 가명처리하여 제공·활용가능함을 법률에 명확히 규정하도록 법률 개정

- 공공기관의 가명정보 처리절차 마련 및 가명정보 처리환경 등에 따라 탄력적으로 운영하도록 지침 제공, 전문기관을 통한 지원체계 구축
- 공공기관 대상 각종 평가에 가명정보 제공실적 반영 및 수수료 수입을 인센티브화하는 제도적 기반 마련

#### □ 데이터 활용 법적 불확실성 해소

- AI 활용을 위한 개인정보 적법처리 근거 확대, 표준계약조항(SCC), 구속력 있는 기업규칙(BCR) 등 국외이전 수단 확대
- 실제 정보를 처리하는 자 입장에서 개인정보 해당 여부를 상황에 따라 판단하여 불필요한 법적 리스크 우려 해소
- 기존에 기업·기관 등에 축적된 이용자 정보를 AI 개발을 위해 활용할 수 있도록 정보주체의 동의 외 적법 처리근거 명확화

별 지

# AI 데이터 확충 및 개방 확대방안

2025. 2. 20.



관계부처 합동

# || 목 차 ||

I. 추진배경 .....	1
II. 추진방향 .....	2
III. 주요 정책과제 .....	3
1. AI 개발 촉진을 위한 고품질 데이터 제공 확대 .....	3
2. 공공부문 데이터 활용 여건 획기적 개선 .....	5
3. 데이터 활용 법적 불확실성 해소 .....	6
IV. 향후 추진계획 .....	7

## I. 추진배경

- 인공지능(AI) 기반 디지털 전환이 가속화되면서 개인정보를 포함한 양질의 '데이터'가 기업과 국가경쟁력 확보의 핵심 관건으로 부상
  - 딥시크 AI는 한정된 자원에서도 기술 혁신으로 모델 최적화가 가능함을 보여준 사례로, 앞으로 AI 경쟁력의 핵심이 '데이터'에 있음을 명확화
  - AI 기술이 불러오는 데이터 처리 방식의 근본적 변화\*는 現 개인정보 이용 관행과 제도의 패러다임 전환을 촉구
    - \* ▲(데이터) 정형(표·엑셀 등) → 비정형(녹취록·영상·음성·센싱 데이터)
    - ▲(처리유형) 1:1, 정형적 관계, 예측 용이 → N:N, 불특정 다수, 예측 곤란
    - ▲(리스크) 유출·오남용 → 추적·감시, 차별, 인격권 침해
- 그간 정부는 디지털 대전환과 AI 심화 시대에 대응하여 유연한 데이터 활용 여건에 초점을 둔 AI·데이터 정책 마련·추진
  - ※ 「제1차 데이터산업 진흥 기본계획」(‘23.1.), 「가명정보 활용 확대방안」(‘23.7.), 「AI 시대 개인정보 정책방향」(‘23.8.), 「데이터 경제 활성화 추진과제」(‘23.11.) 등 발표
  - 정부의 다각적 노력에도 불구하고, 여전히 AI·데이터 활용 장벽 존재
- 우리나라는 세계 최고 수준의 공공·민간 부문에 축적된 데이터 기반에도 불구하고, 데이터 이용이 현장의 기대 수준에 부응하지 못할 뿐만 아니라 데이터를 사회·경제적으로 '최적' 활용하지 못하는 상황
  - (공공) 기관·분야별로 축적된 데이터를 일종의 '자산'으로 여겨 다른 기관·분야에 공유\*하는 것에 여전히 소극적
    - \* 가명 데이터 선도프로젝트의 경우에도 공공기관 평균 데이터 처리·제공에 약 10개월 이상 소요(‘21.~’23. 가명 데이터 결합 선도사례 분석 결과)
  - (민간) 원본 데이터 등 데이터 활용 수요가 높으나 개인정보 처리 관련 법적 불확실성, 복잡한 데이터 이용 절차, 분야별 특화 데이터 부족 등 제약
  - (인식) 개인정보 리스크를 이유로 데이터 활용을 주저하는 경향 만연, 대규모 개인정보 수집, 국외이전 등으로 인한 국민의 불안감 상존



**AI 시대에 대응한 데이터 이용 관행 및 제도의 총체적 업그레이드 필요**

## II. 추진방향

### AI 데이터 활용 촉진

#### 추진전략

#### 7대 핵심과제

##### 전략①

AI 개발 촉진을  
위한 고품질 데이터  
제공 확대

- ① 비정형 원본 데이터 활용 확대
- ② 원본 활용이 어려운 데이터는 합성데이터 활용 지원
- ③ 분야별 고품질 데이터 구축·개방 확대

##### 전략②

공공부문 데이터  
활용 여건  
획기적 개선

- ① 공공기관의 가명정보 제공 거버넌스 제도화 및 유인 마련
- ② 공공기관의 가명정보 제공 효율성 제고

##### 전략③

데이터 활용  
법적 불확실성  
해소

- ① AI 활용을 위한 데이터 적법처리 근거 확대
- ② AI 특성을 고려한 「개인정보 보호법」 합리적 해석·적용

### Ⅲ. 주요 정책과제

#### 1 AI 개발 촉진을 위한 고품질 데이터 제공 확대

기존	개선
<ul style="list-style-type: none"> <li>자율주행 분야 한시적 원본활용 허용</li> <li>합성데이터 안전성에 대한 민간 부담</li> <li>분야별 특화 데이터 확보 어려움</li> </ul>	<ul style="list-style-type: none"> <li>사회·산업적 필요 분야 원본활용 제도화</li> <li>합성데이터 안전성 확인 및 공개 지원</li> <li>핵심 분야 특화 데이터 구축·개방</li> </ul>

##### 1 비정형 원본 데이터 활용 확대 개인정보위·행안부·과기정통부

- **(분야 확대)** 그간 자율주행 분야만 원본 데이터 활용이 가능하였으나, 재난예방, 국민안전 등 사회적·산업적으로 필요한 분야로 확대
- **(불확실성 해소)** AI 업체가 장기간 안정적으로 데이터를 활용할 수 있는 특례를 마련\*하여 불확실성 해소(실증특례 최대 4년 → 연구 필요 기간)
  - \* ▲개인정보 보호법 상 데이터 활용 특례 마련 ▲영상정보처리 관련 별도법 제정 등('25~)
- **(AI CCTV 지원)** 자동 삭제되는 지자체 CCTV(약 65만대) 영상 중 고가치 원본 데이터\*를 확보하여 민관이 협업하는 활용 인프라 구축
  - \* 재난재해, 다중밀집, 긴급 구조 상황 CCTV 영상 등(현재 30일 이내 보관 후 자동 삭제)

##### 2 원본 활용이 어려운 데이터는 합성데이터\* 활용 지원 개인정보위·산업부·과기정통부·행안부

- \* 컴퓨터 알고리즘 등을 이용하여 생성된 정보로 원본데이터의 다양한 속성을 그대로 재현한 데이터
- **(합성데이터 확산)** 퍼지컬 AI\* 시대 대비 등에 필요한 합성데이터가 AI 학습에 활발하게 활용될 수 있도록 기반 확충
  - \* 현실세계의 물리적인 데이터를 기반으로 학습하고 작동하는 AI로, 현실에서 수집하기 어려운 데이터를 학습하고 시뮬레이션하기 위해 합성데이터 활용이 필수적
- 개인정보 노출 우려로 인해 합성데이터 활용이 위축되지 않도록 정부가 합성데이터 안전성 평가 지원체계 구축(산업융합 규제샌드박스로 우선 지원)



- 'AI허브\*', '공공데이터포털' 등을 통해 합성데이터 공개

\* ('24년) 내시경 이미지 등 비정형 합성데이터 27만개, 금융 합성데이터 1.6억개 공개  
( '25년) 헬스케어·국방 분야 합성데이터셋 6종 공개 예정(과기정통부)

- (R&D) 다부처 협업 과제로 합성데이터 안전성 검증 등 AI 프라이버시 강화 기술 개발 추진

### 3 분야별 고품질 데이터 구축·개방 확대 과기정통부 중기부·행안부

- (버티컬 AI 개발) 생성형 AI 기반 모델(LLM, sLLM) 고도화 차원에서 버티컬 AI 개발을 위해 분야별\* 특화 고품질 데이터 구축·개방 확대

\* 예: 헬스·뷰티, 미디어·콘텐츠, 산업·제조, 재난·안전, 금융·회계, 교육 등

※ 각 부처 소관 분야 데이터를 최대한 수집·활용할 수 있도록 부처 간 데이터 협력 적극 추진

- 정부 지원을 통해 구축한 AI 학습용 데이터 통합 제공 체계 구축

- (제조 데이터) 제조분야 데이터 표준 및 공정별 인증제도를 도입하고, '지역특화 제조AI센터'와 데이터 플랫폼\*을 연계하여 제조AI 도입·확산

\* KAMP : 중소벤처 인공지능제조플랫폼(Korea AI Manufacturing Platform)

장비	전문가 선정 500개 핵심 장비의 국제표준(AAS) 적용 데이터 가이드스 개발
공정	500개 핵심장비가 포함된 77개의 대표 공정도 개발
기업	제조데이터 표준기반의 공장·기업 간 데이터(탄소배출량 등) 공유 체계 구축

- (국가중점데이터) AI 수요가 높은 비정형데이터(텍스트, 영상, 이미지 등), 합성데이터 등 적극 개방 및 검색·활용 편의성 제고\*

\* 비정형데이터의 구조, 속성 등 메타정보를 공공데이터포털에 함께 개방 추진

- AI 활용도가 높은 핵심 데이터 분야를 선정하고, 적극적인 개방을 유인하기 위해 해당 데이터 개방실적을 공공데이터 법정 평가에 반영

#### < 2025년 국가중점데이터 개방과제(안) >

- (리걸테크 데이터) 텍스트 중앙행정기관 법령해석 및 특별행정심판기관 재결례 데이터
- (센서데이터) 텍스트 특수교 통합관리 계측 데이터, 신재생에너지발전소 운영 정보, 해양 환경 정보
- (공공연구데이터) 이미지·영상 스마트팜 연구기술 정보, 영유아·아동·청소년 패널데이터
- (합성데이터) 비수도권 교통카드 이용내역 데이터

## 2 공공부문 데이터 활용 여건 획기적 개선

기 존	개 선
<ul style="list-style-type: none"> <li>기관 내부 가명정보 제공 거버넌스 및 유인 부재</li> <li>가명정보 처리 역량 부족 (제공에 많은 시일 소요)</li> </ul>	<ul style="list-style-type: none"> <li>가명정보 제공 내부운영체제 구축, 평가 반영 및 인센티브 제공</li> <li>가명정보 제공 업무 위탁 지원체제 구축 → 제공기간 대폭 단축</li> </ul>

### 1 공공기관의 가명정보 제공 거버넌스 제도화 및 유인 마련 개인정보위·기재부·행안부·국조실

- **(내부운영체제)** 가명정보 '신청 접수'부터 '제공'에 이르기까지 가명처리 프로세스 및 관계 부서의 역할, 책임 등을 총리훈령\*으로 규정  
\* (가칭)「공공기관의 가명정보 제공 내부운영체제에 관한 규정」
- **(법적근거 명확화)** 개인정보가 포함된 공공데이터를 가명처리하여 제공·활용 가능함을 관련 법률\*에 명확히 규정  
\* 「공공데이터법」 및 「데이터기반행정법」 개정 추진
- **(경영평가)** 공공기관 대상 각종 평가\*에 가명정보 제공실적 반영  
\* 「공공기관 경영평가」, 「공공데이터 제공 및 데이터기반행정 평가」 등
- **(인센티브)** 「가명처리 수수료 산정 가이드라인」 등 가명정보 제공 수수료 수입을 인센티브로 활용할 수 있는 제도적 기반 마련

### 2 공공기관의 가명정보 제공 효율성 제고 개인정보위·행안부

- **(절차 합리화)** 가명처리 절차가 데이터 처리환경 등에 따라 탄력적으로 운영될 수 있도록 「리스크 기반 가명처리 절차 차등화 지침」 제공
- **(지원체제 구축)** 개인정보 가명처리 역량이 부족한 공공기관이 가명정보 제공 절차\*를 공신력있는 전문기관에 위탁할 수 있는 체제 구축  
\* 활용 데이터 선별 → 식별 위험성 검토 → 가명처리 → 적정성 검토 → 사후 관리

### 3 데이터 활용 법적 불확실성 해소

기존	개선
<ul style="list-style-type: none"> <li>동의 중심의 개인정보 처리, 개인정보 국외이전 수단 제한</li> <li>개인정보 해당 여부, 이용자 정보 활용 불확실성</li> </ul>	<ul style="list-style-type: none"> <li>공익 등을 고려한 적법처리 근거 확대, 표준계약조항, 기업규칙 등 탄력적 운용</li> <li>합리적 개인정보 보호법 해석을 통해 양질의 데이터 활용 확대</li> </ul>

#### 1 AI 활용을 위한 데이터 적법처리 근거 확대 개인정보위

- **(일반적 처리근거)** 실증특례 경험\*을 바탕으로 범죄 예방 등 공익적 AI 개발을 위한 데이터 활용이 가능하도록 개인정보 적법처리 근거 확대
  - \* 통신사가 수사기관이 보유한 보이스피싱 통화데이터를 제공받아 보이스피싱 예방 AI 서비스를 개발할 수 있도록 규제샌드박스 허용('24.10.)
- **(국외이전)** 국제 공동연구\* 등 글로벌 데이터 이전 환경을 고려하여 표준계약조항(SCC), 구속력 있는 기업규칙(BCR) 등 국외이전 수단 확대
  - \* 첨단 바이오 분야 국제 공동연구에 필요한 가명데이터셋을 환자 동의 없이도 활용할 수 있도록 규제샌드박스 허용(서울대병원, '24.6.)

#### 2 AI 특성을 고려한 「개인정보 보호법」 합리적 해석·적용 개인정보위

- **(개인정보 판단)** 실제 '정보를 처리하는 자' 입장에서 개인정보 해당 여부를 개별적으로 판단하여, 현장의 불필요한 법적 리스크 우려 해소
  - ※ (예) 자동차등록번호는 일반적으로 개인정보가 아니고, 다른 정보를 쉽게 입수하여 개인을 알아볼 수 있는 특수 상황에서만 개인정보에 해당(자율주행차가 촬영한 자동차등록번호 등)
- **(이용자 정보)** 기업·기관 등에 축적된 방대한 이용자 정보를 AI 서비스 개발·개선에 가치있게 활용할 수 있도록 동의 외 적법 처리근거 명확화

종전	변경
<ul style="list-style-type: none"> <li>AI 학습 목적으로 이용하려면 정보주체 동의 필요</li> </ul>	<p>&lt;유형1&gt; 당해 서비스 개선 목적(시스템 고도화 포함) → AI학습·기술연구·통계작성 등에 목적 내 이용 가능</p> <p>&lt;유형2&gt; 당초 서비스와 합리적 관련성 있는 신규서비스 개발 → 예측가능성, 사후거부권(opt-out) 등 충족하면 추가적 이용 가능</p> <p>&lt;유형3&gt; 당초 서비스와 별개의 신규서비스 개발 → 가명처리 특례에 따라 이용하거나 새로운 적법처리 근거 마련</p>

## IV. 향후 추진계획

### ① AI 개발 촉진을 위한 고품질 데이터 제공 확대

정책 과제	시기	소관부처
■ 비정형 원본 데이터 활용 규제샌드박스 확대	연중	개인정보위·행안·과기정통부
■ 비정형 원본 데이터 활용 법적 특례 마련	'25.~	개인정보위
■ 지자체 CCTV 영상 원본데이터 활용 인프라 구축	'26.12.	개인정보위·행안·과기정통부
■ 합성데이터 적법 활용 지원	'25.7.	산업부·개인정보위
■ 안전한 합성데이터 공개 확대	'26.8.	개인정보위·행안·과기정통부
■ 합성데이터 안전성 평가 등 기술 개발 추진	'25.9.	개인정보위·과기정통부
■ 버티컬 AI를 위한 분야별 특화 고품질 데이터 구축	'25.~	과기정통부
■ 제조분야 데이터 표준 및 공정별 인증제도 도입	'25.~	중기부
■ 공공부문 비정형데이터 개방 확대	'25.12.	행안부

### ② 공공부문 데이터 활용 여건 획기적 개선

정책 과제	시기	소관부처
■ 공공기관 가명정보 제공 거버넌스 관련 총리훈령 제정	'25.12.	개인정보위·국조실
■ 「공공데이터법」, 「데이터기반행정법」 개정 추진	'25.~	행안부
■ 공공기관 평가에 가명정보 제공실적 반영	'25.12.	행안·기재부
■ 「가명처리 수수료 산정 가이드라인」 마련	'25.~	개인정보위
■ 「리스크 기반 가명처리 절차 차등화 지침」 마련	'25.12.	개인정보위·행안부
■ 데이터 가명처리 지원체계 구축	'25.~	개인정보위

### ③ 데이터 활용 법적 불확실성 해소

정책 과제	시기	소관부처
■ 개인정보 적법처리 근거 확대	'25.~	개인정보위
■ 국외이전 수단 확대	'25.~	개인정보위
■ AI 특성을 고려한 「개인정보 보호법」 합리적 해석·적용	즉시	개인정보위