

AI데이터 확충 및 개방 확대방안

AI

AI 발전 추세



한정된 자원으로 모델 최적화 가능
➔ 비용효과적 AI 개발·운영



텍스트, 멀티모달, 피지컬 AI
➔ 응용분야의 무한 확장

“데이터”

핵심경쟁요소로 부각

세계 최고 수준의 데이터 기반에도 불구하고, 활용 미흡

공공

- 공공데이터법은 개인정보 미포함
- 보유 데이터 제3자 제공·활용에 소극적

민간

- 고품질 데이터 부족
- 개인정보 이용 시 법적 리스크

인식

- 프라이버시에 대한 국민 불안감 여전

AI 데이터 활용 촉진

추진전략

전략 ①

AI 개발 촉진을 위한 고품질 데이터 제공 확대

비정형 원본 데이터 활용 확대

원본 활용이 어려운 데이터는
합성데이터 활용 지원

분야별 고품질 데이터
구축·개방 확대

전략 ②

공공부문 데이터 활용 여건 획기적 개선

공공기관의
가명정보 제공 거버넌스
제도화 및 유인 마련

공공기관의 가명정보
제공 효율성 제고

전략 ③

데이터 활용 법적 불확실성 해소

AI 활용을 위한
데이터 적법처리 근거 확대

AI 특성을 고려한
「개인정보 보호법」 합리적
해석·적용

7대 핵심과제

전략 ① 고품질 데이터 제공 확대 - 비정형 원본 데이터

➤ 원본 데이터 활용 환경 조성

- 자율주행 분야 한정 (5개사, '24.2~)
- 실증특례를 통해 한시 허용중 (2+2년)



- 사회·산업적 필요 분야 적극 확대
※ 재난 예방, 국민 안전 등
- 사업·연구 필요기간 원본 활용 제도화

➤ 부처 협업을 통한 인프라 구축

- 전국 216개 지자체 통합관제센터 CCTV (65만 대) 촬영영상 중 AI 응용·개발에 반드시 필요한 고가치 원본 데이터* DB 구축

* 재난재해, 다중밀집, 긴급구조, 특수상황 등 CCTV원본 영상 (현재 30일 이내 보관 후 자동 삭제)



행정안전부

재난안전법에 근거
학습데이터 수집체계 구축



과학기술정보통신부

데이터 활용
기술·설비 지원



개인정보보호위원회

원본 활용 근거 마련
안전조치 점검

전략 ① 고품질 데이터 제공 확대 - 합성데이터*

* 원본데이터의 다양한 속성을 그대로 재현한 가상의 데이터

학습데이터 부족

- 학습 가능 데이터는 빠르게 고갈 중
 - * 학습에 필요한 공개데이터는 빠르면 '26년 한계 도달 예측 ('24,美 Epoch AI)
- AI 품질 확보를 위해 특수상황 데이터 학습, 결측데이터 보완 필요

다가오는 피지컬 AI 시대

- 피지컬 AI*는 공간·상황 인식 및 학습을 위해 다양한 데이터 필요
 - * 물리적 세계에서 사고하고 움직이는 AI
- 원본 활용이 어렵거나 학습데이터가 부족할 시 AI 성능 저하 초래

더 많은 고품질 데이터, 안전한 데이터를 위한 해결책

합성데이터 생성·활용 지원

안전성 평가 지원체계 구축

개인정보 노출, 재식별 우려가 없도록
합성데이터 안전성 평가 지원체계 구축

분야별 합성데이터 공개

AI 허브, 공공데이터포털
등을 통해 대국민 공개

연구개발 추진

합성데이터 생성, 안전성 검증 등을 위한
기술·도구 개발



산업통상자원부



과학기술정보통신부



행정안전부



개인정보보호위원회

협업

전략 ① 고품질 데이터 제공 확대 - 분야별 데이터



버티컬 AI 데이터

- 생성형 AI 모델 고도화를 위한 분야별 특화 고품질 데이터 구축·개방
- AI 학습용 데이터 통합 제공 체계 구축

미디어	Text to Video AI 서비스 구현 지원 등 K-콘텐츠 경쟁력 강화
금융·회계	회계처리·재무제표 작성 기준서 데이터 등 AI 서비스화
교육	단계적 사고 데이터 구축으로 추론형 문제해결 AI 개발 지원

※기타 헬스·뷰티, 산업·제조, 재난·안전 등 지원



제조 데이터

- 제조분야 데이터 표준 및 공정별 인증제도 도입
- 지역특화 제도AI 센터와 데이터 플랫폼(KAMP) 연계

장비	500개 핵심 장비 데이터 가이드스 개발
공정	77개 대표 공정도 개발
기업	공장·기업 간 데이터 공유 체계 구축



국가중점 데이터

- AI 활용도가 높은 비정형데이터 등 핵심 데이터 분야 선정
- 데이터 개방 실적을 공공데이터 법정 평가에 반영

리걸테크	중앙행정기관 법령해석, 심결례 데이터
센서	특수교 통합관리 계측, 신재생에너지·발전소 운영 정보 등
공공연구	스마트팜, 영유아·아동·청소년 패널데이터
합성	비수도권 교통카드 이용내역

전략 ② 공공부문 데이터 활용 여건 획기적 개선

▶ 공공부문 대규모 데이터 보유

 817개 기관  1만 7902개 시스템  35만개 개인정보DB

▶ 과학적 연구, AI 학습에 활용 저조

- 최근 1년 가명정보 제3자 제공한 공공기관 2.9% 불과
- 데이터협의(130일) - 가명처리(126일) - 결합·반출(54일) 까지 평균 310일 소요

가명데이터 선순환체계 구축

저조한 원인

담당자·담당부서·가명처리 프로세스 부재

가명정보 제공 인센티브 부족

가명처리 전문성, 경험 부족

개선대책

가명정보 제공 내부 운영체계 마련 (총리훈령)

공공기관 평가 반영*, 인센티브(가명처리 수수료) 제도 기반 마련

* 「공공기관 경영평가」, 「공공데이터 제공 및 데이터기반행정 평가」 등

가명정보 제공 관련 업무 위탁 지원체계 구축



협업

전략 ③ 법적 불확실성 해소 - 데이터 적법처리 근거 확대

실증특례 경험

- 보이스피싱 예방 AI 서비스 ('24.10.)

금감원·국과수 보유 “그놈 목소리”
통화 데이터 2만 5천건

- 첨단바이오 분야 국제공동연구 ('24.6.)

서울대병원·하버드·MIT 등과
의료 가명정보 공동연구
데이터 플랫폼 구축



법제화 추진

- AI 데이터 적법처리 근거 확대

정보주체
동의 위주 → 사업자의 (공익 목적
정당한 이익) 인정

- 국외 이전 수단 확대

동의·인증·
동등성 인정 → (표준계약(SCC)*
기업규칙(BCR)**) 추가

*Standard Contract Clauses

**Binding Corporate Rules

전략 ③ 법적 불확실성 해소 - 혁신 친화적 해석·적용

- 개인정보성 판단기준 명확화

현장 불확실성

개인 식별이 가능한
모든 상황 고려



개선

실제 '정보를 처리하는 자' 입장에서
개별적으로 판단

- 이용자 정보 AI 학습 시, 동의 외 적법 처리 근거 다양화·명확화

현장 불확실성

정보주체 동의 필요



개선

- | | | |
|-----|------------------------------------|---------------------------|
| 유형1 | 당해 서비스 개선·고도화 목적 | ▶ 목적 내 이용 가능 |
| 유형2 | 당초 서비스와 합리적 관련성 있는
신규 서비스 개발 목적 | ▶ 추가적 이용 가능(예측가능성, 옵트아웃) |
| 유형3 | 별개의 신규 서비스 개발 목적 | ▶ 가명처리 특례 또는 새로운 처리 근거 마련 |

국민 신뢰 속에서
세계 최고의 데이터 기반
AI 혁신 생태계를 조성하겠습니다.



관계부처 합동