

「인공지능 학습용데이터 통합제공시스템」연계·개방을 위한
AI 데이터 품질 컨설팅 지원 안내

NIA 한국지능정보사회진흥원

ALLBIGDAT CONSORTIUM

CONTENTS

CHAPTER
I

추진 개요

CHAPTER
II

사업 지원 내용

APPENDIX
#

세부 추진 내용

2026년도 「통합제공시스템 연계·개방」을 위한 AI 데이터 품질 컨설팅 지원 안내

CHAPTER

I

추진 개요

- 01 통합제공시스템 구축 배경
- 02 통합제공시스템 활용 목적

01. 통합제공시스템 구축 배경

01. 추진 개요

02. 사업 지원 내용

세부 추진 내용

공공·민간이 보유한 AI 학습용데이터를 통합적으로 수집·관리하여 민간에 제공하기 위해 AI 학습용데이터 통합제공시스템 운영

2025.03

데이터 통합 제공체계
구축방안 마련 착수

2026.01

“인공지능기본법” 시행

2026.04

통합제공체계 대응
컨설팅 사업 추진

2026.07

통합제공시스템
연계·지원 추진

통합제공시스템 관련 법령

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 (약칭: 인공지능기본법)

[시행 2026. 7. 21.] [법률 제2315호, 2026. 1. 20. 일부개정]

- 제15조(인공지능 학습용데이터 관련 시책의 수립 등) ① 과학기술정보통신부장관은 관계 중앙행정기관의 장과 협의하여 학습용데이터의 생산·수집·관리·유통·활용 촉진 및 품질수준 확보 등을 위하여 필요한 시책을 수립하여야 한다. <개정 2026. 1. 20.>
- ② 정부는 학습용데이터의 생산·수집·관리·유통 및 활용 등에 관한 시책을 효율적으로 추진하기 위하여 지원대상사업을 선정하고 예산의 범위에서 지원할 수 있다.
- ③ 정부는 학습용데이터의 생산·수집·관리·유통 및 활용의 활성화를 위하여 다양한 학습용데이터를 제작·생산하여 제공하는 사업(이하 “학습용데이터 구축사업”이라 한다)을 시행할 수 있다.
- ④ 과학기술정보통신부장관은 학습용데이터 구축사업의 효율적 수행을 위하여 학습용데이터를 통합적으로 제공·관리할 수 있는 시스템(이하 “통합제공시스템”이라 한다)을 구축·관리하고 민간에 자유롭게 이용할 수 있도록 제공하여야 한다.
- ⑤ 과학기술정보통신부장관은 통합제공시스템을 이용하는 자에 대하여 비용을 징수할 수 있다.
- ⑥ 그 밖에 제7항에 따른 지원대상사업의 선정 및 지원, 학습용데이터 구축사업의 시행, 통합제공시스템의 구축·관리 및 제7항에 따른 비용의 징수 등에 필요한 사항은 대통령령으로 정한다.

“인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 (인공지능기본법)”

제15조(인공지능 학습용데이터 관련 시책의 수립 등)

④항 ... 학습용데이터를 통합적으로 제공·관리할 수 있는 시스템(이하 “통합제공시스템”이라 한다)을 구축·관리하고 민간에 자유롭게 이용할 수 있도록 제공하여야 한다.

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령 (약칭: 인공지능기본법 시행령)

[시행 2026. 7. 21.] [대통령령 제36000호, 2026. 7. 20. 일부개정]

- 제13조(학습용데이터 통합제공시스템의 구축 및 관리) ① 과학기술정보통신부장관은 법 제15조제4항에 따른 통합제공시스템(이하 “통합제공시스템”이라 한다)이 다음 각 호의 기능을 수행할 수 있도록 구축·관리해야 한다.
1. 학습용데이터의 통합 검색
 2. 학습용데이터의 체계적 분류
 3. 학습용데이터 출처 제공
 4. 통합제공시스템과 국가기관 및 민간단체 등이 운영하는 다른 데이터 플랫폼과의 연계
 5. 학습용데이터의 가치평가 및 품질관리 관련 정보의 제공
 6. 그 밖에 과학기술정보통신부장관이 학습용데이터 구축사업의 효율적 수행을 위하여 필요하다고 인정하는 기능
- ② 과학기술정보통신부장관은 중앙행정기관의 장, 지방자치단체의 장이나 공공기관(「지능정보화기본법」 제2조제16항에 따른 공공기관을 말한다. 이하 같다)의 장에게 통합제공시스템과 각 기관이 보유한 개별 시스템 및 데이터와의 연계, 통합제공시스템의 운영을 위하여 필요한 정보 및 자료의 제공 등 협력을 요청할 수 있다. 이 경우 요청을 받은 기관의 장은 특별한 사유가 없으면 이에 따라야 한다.
- ③ 과학기술정보통신부장관은 제2항에 따른 정보 및 자료의 제공 등 협력을 요청하는 경우 「공공데이터의 제공 및 이용 활성화에 관한 법률」 제4조 공공데이터를 학습용데이터로 제공받으려는 자에게 대해서는 행정안전부장관과 협의해야 한다. <신규 2026. 7. 20.>
- ④ 과학기술정보통신부장관은 통합제공시스템을 통하여 제공하는 학습용데이터의 최신성, 정확성 및 상호 연계성이 유지되도록 노력해야 한다. <개정 2026. 7. 20.>

“인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령 (인공지능기본법 시행령)”

제13조(학습용데이터 통합제공시스템의 구축 및 관리)

②항 ... 각 기관이 보유한 개별 시스템 및 데이터와의 연계, 통합제공시스템의 운영을 위하여 필요한 정보 및 자료의 제출 등 협력을 요청할 수 있다. 이 경우 요청을 받은 기관의 장은 특별한 사유가 없으면 이에 따라야 한다.

02. 품질컨설팅 지원 목적

✓ 품질 컨설팅 지원 필요성

“AI 데이터의 통합제공시스템 등록·개방 필요”

AI 학습용 데이터 통합제공체계



AI데이터
구축



데이터 품질
검증



통합제공시스템
등록 및 개방



구축된 AI 데이터는 구축 이후 통합제공시스템에
등록·개방까지 이행 필요

품질 컨설팅 지원의 목적

01 통합제공시스템 개방 대상 데이터 품질 신뢰성 확보

- 품질관리 및 검증 지원을 통해 통합제공시스템의 공개 데이터 품질 수준 확보
- 구축 진행중인 데이터의 오류 점검 및 보완으로 대국민 개방 데이터 신뢰성 확보

등록 적합

품질 신뢰

개방 활용

02 AI 데이터 구축기관의 품질검증 대응 역량 제고

- 품질검증 기준·절차 교육을 통한 사업수행기관의 데이터 품질 이해도 제고
- 구축 단계별 자가점검 체계 내제화를 통해 자체 품질관리 역량 강화

품질관리 역량

자가 점검

03 품질 관리체계의 안정적인 정착

- 사업별 데이터 특성과 구축 일정에 맞춘 품질컨설팅 및 현장 지원 수행
- 품질검증 이후 통합제공시스템 등록·개방까지 이어지는 이행 지원체계 마련

등록 이행

현장 지원

컨설팅

2026년도 「통합제공시스템 연계·개방」을 위한 AI 데이터 품질 컨설팅 지원 안내

CHAPTER

II

사업 지원 내용

- 01 컨설팅 세부 지원 사항
- 02 데이터 품질 컨설팅 지원
- 03 데이터 품질 컨설팅 교육
- 04 비식별화 처리 지원
- 05 사업 지원 일정

01. 컨설팅 세부 지원 사항

01. 추진 개요

02. 사업 지원 내용

#. 세부 추진 내용

✓ 품질 컨설팅 지원 사항

Support

1

AI 데이터 품질관리 컨설팅



- 구축 공정 단계별 품질 컨설팅
- 제3자 품질검증 대응 컨설팅
- 데이터 품질 제고를 위한 자가점검 계획 수립 및 결과분석 컨설팅

Support

2

AI 데이터 품질관리 교육



- 데이터 구축 공정의 이해
- 품질관리 프로세스의 이해
- 품질지표 및 목표설정의 이해

추가 지원

3

AI 데이터 비식별화 처리 지원



- 비식별화 대상 정보 컨설팅 안내
- 개인정보 비식별화 처리 지원

02. 데이터 품질 컨설팅 지원

01. 추진 개요

02. 사업 지원 내용

#. 세부 추진 내용

구축 공정 단계별 품질 컨설팅

- AI 데이터 구축 공정별 필수 산출물을 기준으로 작성 수준을 점검하고 품질검증 대응에 필요한 보완사항 점검·지원



AI 데이터 구축

AI 데이터 구축 프로세스



공정별 주요 활동

- ✓ 원시데이터 분류체계(Class) 기준 수립
- ✓ 통계적 다양성 분포를 고려한 계획 수립 및 수집
- ✓ 학습목적에 부합하는 유효한 원시데이터 선별
- ✓ 중복 제거 및 개인정보 비식별화를 통한 원천데이터 확보
- ✓ 정확한 데이터셋 구축을 위한 GT기준 설계
- ✓ 원천데이터에 GT를 부여한 가공데이터 구축
* GT: Ground Truth
- ✓ 목표 Task에 맞는 적정 AI 모델 선정
- ✓ 학습용 데이터(원천+라벨링)로 AI 모델 학습
- ✓ 학습된 AI모델의 성능지표로 학습 유효성 검증

의뢰

지원



품질 컨설팅 지원

컨설팅 핵심 지원사항

고품질 데이터 구축을 위한
“공정 단계별 데이터 컨설팅 지원”

- 데이터 수집 단계
 - ✓ 데이터 분류 체계 및 메타데이터 명세 점검 지원
 - ✓ 데이터 유실율을 고려한 Class 분포 점검 지원
- 데이터 정제 단계
 - ✓ 데이터 중복 및 개인정보 비식별화 점검 지원
- 데이터 가공 단계
 - ✓ 가공기준 마련 및 가이드라인 작성 지원

02. 데이터 품질 컨설팅 지원

01. 추진 개요

02. 사업 지원 내용

#. 세부 추진 내용

제3자 품질검증 대응 컨설팅

- AI 데이터 품질 전문기관(TTA)을 통한 기구축 AI 데이터의 정량적 품질 수준 측정 및 검·인증 지원 및 제3자 검증 대응 지원



AI 데이터 구축

제3자 품질검증 프로세스

1 품질지표 기준 수립

2 사전 품질 검증

3 품질지표 기준 확정

4 최종 품질 검증

공정별 주요 활동

• 품질검증 제반 문서 작성

- ✓ 품질지표기준서
- ✓ 항목별측정조건
- ✓ 구문검사규칙
- ✓ 의미정확성 검사기준서

• 제3자 품질검증 의뢰

- ✓ 품질검증 결과 확인
- ✓ 품질검증 결과에 따른 데이터 보완

의뢰



지원



품질 컨설팅 지원

컨설팅 핵심 지원사항

- 품질지표기준 수립지원
- 항목별측정조건 수립지원
- 구문검사규칙 수립지원
- 의미정확성 검사기준 수립지원

02. 데이터 품질 컨설팅 지원

01. 추진 개요

02. 사업 지원 내용

#. 세부 추진 내용

1-Cycle 자가점검 활동 컨설팅

- 구축공정별 품질 이슈 최소화를 위해 데이터 구축 수준에 따라 사업수행기관이 자체적으로 수행하는 품질 자가점검의 원활한 수행을 지원



AI 데이터 구축

1-Cycle 자가점검 활동

1 샘플데이터 1-Cycle 자가점검

2 초기데이터 1-Cycle 자가점검

3 중간데이터 1-Cycle 자가점검

4 중간데이터 보완점검

5 최종데이터 1-Cycle 자가점검

공정별 주요 활동

- 단계별 1-Cycle 자가점검 수행
 - ✓ 샘플·초기·중간·최종 데이터의 데이터 구축 진도에 따른 단계별 점검
- 지표 기반 데이터 품질 수준 확인
 - ✓ 데이터 유사도 및 통계적 다양성 점검
 - ✓ 구문·의미정확성 등 주요 품질항목 검사
- 점검 결과 분석 및 데이터 보완
 - ✓ 오류·미흡 데이터 식별 및 원인 분석
 - ✓ 보완조치 수행 후 재점검 및 결과 정리

의뢰



지원



품질 컨설팅 지원

컨설팅 핵심 지원사항

- 자가점검 계획 수립 지원
- 품질지표별 검사방안 수립 지원
- 자가점검 결과 보고서 작성 및 분석 지원
- 데이터 품질 보완 활동 계획 수립 지원

03. 데이터 품질 컨설팅 교육

01. 추진 개요

02. 사업 지원 내용

세부 추진 내용

✓ 품질 컨설팅 교육 방안

품질 컨설팅 교육 운영 방안

• 컨설팅 교육 개요

- ✓ AI 데이터 품질관리 체계 내재화를 위한 **구축사업 관계자 대상 교육 운영**
- ✓ 사업기간 동안 **2회** 운영: **구축공정별 품질관리 기준과 현장 실무 중심 교육 진행**

• 컨설팅 교육 제공 방안

오프라인 교육 진행

- ✓ 구축사업 관계자 대상
집체 교육 진행



온라인 교육 자료 배포

- ✓ 집체교육 녹화영상 및
교육자료 온라인 배포



품질 컨설팅 교육 커리큘럼

데이터 구축 공정의 이해

- 데이터 구축 전 공정의 흐름 및 단계별 주요 관리 포인트 이해

품질관리 프로세스의 이해

- 품질관리를 위한 1-Cycle 자가점검 항목과 항목별 점검방법 이해

품질지표 및 목표설정의 이해

- 품질지표별 정의 및 항목별 측정·목표 설정 기준의 이해

04. 비식별화 처리 지원

01. 추진 개요 02. 사업 지원 내용 #. 세부 추진 내용

비식별화 처리 지원

- AI 데이터 구축 기관의 비식별화 처리 요청 시, 비식별화 처리 지원

비식별화 처리 지원 방안

비식별화 처리 지원

• 지원 방향

- 필요시 개인정보·민감정보 비식별 처리 지원

• 데이터별 비식별화 방안

<텍스트 데이터 비식별화>

개인정보	가명정보	익명정보																								
 본인 또는 본인에 관한 정보로 식별, 추적, 연결, 분석, 평가 등 개인을 식별할 수 있는 정보 <table border="1"> <tr><td>성명</td><td>홍길동</td></tr> <tr><td>나이</td><td>30세</td></tr> <tr><td>전화번호</td><td>010-1234-5678</td></tr> <tr><td>주소</td><td>서울특별시 강남구 테헤란로 12</td></tr> </table>	성명	홍길동	나이	30세	전화번호	010-1234-5678	주소	서울특별시 강남구 테헤란로 12	 개인정보의 일부 또는 전체를 가명(가칭)으로 대체하여 식별, 추적, 연결, 분석, 평가 등 개인을 식별할 수 있는 정보 <table border="1"> <tr><td>성명</td><td>홍○○</td></tr> <tr><td>나이</td><td>30대 초반</td></tr> <tr><td>전화번호</td><td>010-****-*****</td></tr> <tr><td>주소</td><td>서울특별시</td></tr> </table>	성명	홍○○	나이	30대 초반	전화번호	010-****-*****	주소	서울특별시	 시간·세월·기술 등을 활용하여 고충돌 제 식별 정보를 삭제하여 다시 식별 정보를 식별할 수 없는 정보 <table border="1"> <tr><td>성명</td><td>(식별)</td></tr> <tr><td>나이</td><td>30대</td></tr> <tr><td>전화번호</td><td>(식별)</td></tr> <tr><td>주소</td><td>대한민국</td></tr> </table>	성명	(식별)	나이	30대	전화번호	(식별)	주소	대한민국
성명	홍길동																									
나이	30세																									
전화번호	010-1234-5678																									
주소	서울특별시 강남구 테헤란로 12																									
성명	홍○○																									
나이	30대 초반																									
전화번호	010-****-*****																									
주소	서울특별시																									
성명	(식별)																									
나이	30대																									
전화번호	(식별)																									
주소	대한민국																									

[그림 11-15] 개인정보, 가명정보, 익명정보 예시

※ 출처 : 개인정보보호위원회 개인정보 처리 가이드라인

<이미지 데이터 비식별화>



컨설팅 핵심 지원사항

• 개인정보 식별 기준 점검

- 개인정보·민감정보 포함 여부 확인
- 데이터 유형별 주요 위험항목 진단

• 비식별화 처리

- 텍스트 가명처리·삭제·마스킹 등 비식별처리
- 이미지 블러링·픽셀화·마스킹 등 비식별처리
- 영상 내 얼굴·차량번호 등 식별정보 비식별처리



지원 의뢰

