

---

「독자 AI 파운데이션 모델 프로젝트」

**공동활용 데이터 지원사업(1차) 공고문(재공고)**

---

2026. 4.



과학기술정보통신부

**NIA**

지능정보원  
한국지능정보사회진흥원

**K data**  
한국데이터산업진흥원

## 【 목 차 】

1. 지원목적 및 배경 .....	1
2. 사업 주요내용 .....	2
3. 수행기관 선정 방안 및 평가 기준 등 .....	6
4. 사업 요구 사항 .....	9
5. 사업 추진체계 및 절차 등 .....	17
6. 제안서 접수 및 방법, 서류 등 .....	19
7. 주요 추진 일정(안) .....	22
<b>【붙임】</b> 공동활용 데이터 지원사업 세부 데이터별 요구사항	23

과학기술정보통신부 공고 제2026-0518호

「독자 AI 파운데이션 모델 프로젝트」

**공동활용 데이터 지원사업(1차) 공고(재공고)**

국민 누구나 AI를 활용하고 그 혜택을 누릴 수 있는 생태계를 구축하기 위해, 다양한 고품질 데이터를 확보하고 글로벌 수준의 독자 AI 파운데이션 모델 개발을 지원하고자 ‘공동활용 데이터 지원사업(1차)’을 재공고하오니 참여를 희망하는 기업·기관은 적극 신청하여 주시기 바랍니다.

2026년 4월 24일

과학기술정보통신부 장관

한국지능정보사회진흥원 원장

한국데이터산업진흥원 원장

# 1 지원 목적 및 배경

## □ 지원 배경

- AI가 산업·경제 전반의 혁신을 주도하는 핵심 전략기술로 부상하고, 주요 국가들이 대규모 투자와 제도 정비를 통해 글로벌 AI 주도권 경쟁을 가속화
- 국내에서도 AI 기술 자립과 산업 생태계 고도화 필요성에 대한 공감대가 확산되며, 데이터 등 핵심자원의 전략적 확보 필요성 증대
- 초거대·멀티모달 AI 확산에 따라 학습데이터의 품질·다양성·신뢰성이 모델 성능과 직결되는 데이터 중심 환경 심화

## □ 지원 목적

- 독자 AI 파운데이션 모델 개발을 가속화하기 위해 고품질 학습 데이터의 체계적 확보 및 공동활용 기반 구축
  - 정예팀에 데이터 자원을 집중 지원하여 세계적 수준의 AI 모델의 신속한 개발
- 공동활용 데이터 공급기업 참여 확대를 통해 AI 생태계 전반의 경쟁력 강화 및 지속 가능한 성장 기반 마련
  - 글로벌 AI 경쟁에 대응할 수 있는 국가 차원의 기술 주권 및 산업 경쟁력 확보

☞ 고품질 전문·특화 데이터의 공동활용 체계를 마련하여, 독자 AI 파운데이션 모델 개발을 가속화하고, 국내 AI 생태계의 지속 가능한 성장 기반 확보

## 2 사업 주요내용

- 공고 및 접수 기간 : '26. 4. 24(금) ~ 5. 4(월) 13시 까지
- 사업기간 : 협약체결일로부터 ~ '26. 9. 30.
- 지원대상 : 기업, 대학, 공공기관, 정부·지자체, 협회 등
- 규모 : 총 2,350백만원
  - (예산규모) 정부지원금 1,000백만원~1,350백만원 수준
    - ※ 정부지원금은 과제조정위원회 심의·조정에 따라 일부 조정될 수 있음
  - (공모분야) 2개 분야 데이터
    - ※ '26년도 예산 상황 등에 따라 선정규모 및 평가 일정 등은 변경될 수 있음
  - (지원대상) 기관(업) 단독 또는 컨소시엄
- 지원 조건
  - AI 학습용 데이터를 수요기관(정예팀 등)이 **공동으로 활용**할 수 있도록 **제공이 가능한 기관·기업\***(컨소시엄 가능)
    - \* AI 학습 등에 적합한 권원(저작권, 소유권 등)을 보유하거나, 확보·유통할 수 있고, 요구에 따라 데이터 정제·가명처리·가공 표준화 등이 가능한 기관 또는 기업
  - 주관기관 또는 참여기관 자격으로 최대 2개의 데이터 분야까지 지원할 수 있으며, 동일한 데이터 분야 내 주관기관 또는 참여기관으로 중복지원은 불가
    - ※ 1개 기관기업은 총 2개의 데이터 분야에 지원 가능, 예시) 주관기관 자격으로 2개, 또는 주관기관이나 참여기관 자격 합산 2개, 또는 참여기관으로만 2개까지 지원 가능
    - ※ 동일 기관이 동일한 데이터 분야(예: 전문도서)에 주관기관으로 지원한 후 다른 컨소시엄을 구성하여 주관기관 또는 참여기관으로 재차 지원 불가

- ※ '26년 독자AI 파운데이션 모델 프로젝트 주관 및 참여기관은 본 사업에 참여 불가
- ※ 주관기관 또는 참여기관 자격으로 다수 분야에 참여 시 참여 한도인 총 2개 분야 접수 완료 기준 순으로 인정하며, 그 이후 접수된 지원은 모두 불인정(접수 순서는 최종 제안서 제출 송신시간 기준)

## □ 유의사항

- 타 정부지원 사업\* 등을 통해 구축·활용한 데이터 제외
  - \* 초거대AI 확산 생태계 조성 사업(舊 AI 학습용 데이터 구축 사업, 한국지능정보사회진흥원) 등
- 수행기관이 제안한 데이터와 既제공 공동활용데이터('25년 하반기 제공)간 중복성 검토(과제조정위원회) 결과에 따라, 제안 데이터에 대한 교체 요구를 할 수 있으며, 수행기관은 해당 결과를 반영한 대체 데이터를 공급해야 함
  - ※ 과제조정위원회의 중복성 검토 결과 우선 지원대상 수행기관이 미수용시 차순위 수행기관이 우선 지원대상 수행기관으로 변경 될 수 있음
- 협약체결 후, 수행기관이 제공한 데이터 관련 법률적 분쟁 또는 이슈가 발생할 경우, 협약서 등 명시된 절차와 기준 등에 따라 처리
- 한국지능정보사회진흥원, 한국데이터산업진흥원의 사업 수행 중 개인 정보보호법 등을 위반하여 사업 참여 제재 처분 중인 경우 지원 불가
- ChatGPT 등의 생성형 모델을 이용한 증강 데이터는 제안 불가

## □ 민간부담금에 관한 사항

- 국가연구개발혁신법 시행령 제19조(연구개발비의 지원과 부담)에 따라 수행기관은 민간부담금 부담

**【정부지원금 지원기준 및 민간부담금 중 현금 부담 기준】**

▶ **정부지원금 지원기준**

중소기업인 경우	중견기업인 경우	공기업, 대기업인 경우
해당 수행기관 총 사업비의 75%이하	해당 수행기관 총 사업비의 70% 이하	해당 수행기관 총 사업비의 50% 이하

▶ **민간부담금 중 현금부담 기준**

중소기업인 경우	중견기업*인 경우	공기업, 대기업인 경우
해당 수행기관 민간부담금의 10% 이상	해당 수행기관 민간부담금의 13% 이상	해당 수행기관 민간부담금의 15% 이상

※ 비영리 기관(대학, 공공기관, 정부지자체, 협회 등)의 경우 정부지원금 100% 지원

\* 「중견기업 성장촉진 및 경쟁력 강화에 관한 특별법」 시행령 제7조에 따라 산정한 평균 매출액 등이 3천억원 미만인 중견기업은 중소기업 수준으로 민간부담금 중 현금부담 가능

□ **정부지원금 지급**

- 정부지원금은 수행기관으로 참여한 모든 주관·참여기관에 분배됨을 원칙으로 함
- 정부지원금은 과제수행에 지장이 없는 범위 내에서 분할 지급
- 예산 등 정책상의 변동이 발생할 경우 협약 금액이 감액 등 조정 (최대 25% 내외)될 수 있으며, 이 경우 과제 조정을 통해 변동된 협약 금액에 따라 구축량 등을 조정할 수 있음

□ **관련 법령·규정·지침 등**

- 개인정보보호법(법률 제20897호)
- 정보통신망 이용촉진 및 정보보호 등에 관한 법률(법률 제21066호)
- 저작권법(법률 제20841호)
- 국가연구개발혁신법(법률 제20354호)
- 정보통신산업 진흥법(법률 제21065호)
- 정보통신진흥기금 운용·관리규정(과학기술정보통신부 고시)

- 기금사업 협약체결 및 사업비 관리 등에 관한 지침(과학기술정보통신부 훈령)  
※ (유의) 9조 2항을 참고하여 평가배제(완전 자본 잠식 등) 대상 여부를 반드시 확인
- ICT 예산 정책 협의체 운영 등에 관한 지침(과학기술정보통신부 훈령)
- 기금 사업비 산정 및 정산 등에 관한 지침(과학기술정보통신부 훈령)
- 기금사업 결과 평가 등에 관한 지침(과학기술정보통신부 훈령)
- 기금사업 성과관리 및 활용 등에 관한 지침(과학기술정보통신부 훈령)
- 기금사업 수행상황 및 정산 보고 등에 관한 지침(과학기술정보통신부 훈령)
- 기금사업 점검계획 등에 관한 지침(과학기술정보통신부 훈령)  
※ 상기 규정 등은 과학기술정보통신부 정보통신진흥기금 운용·관리 규정 및 부속 지침의 제개정시 해당 지침으로 대체하여 적용
- 한국지능정보사회진흥원 ICT 기금사업 및 연구개발사업 관리지침  
※ (유의) ICT기금사업 및 연구개발사업 관리지침 제20조 감점 기준 확인  
※ 상기 지침 등은 지침의 제·개정 시 해당 지침으로 대체하여 적용
- 한국지능정보사회진흥원 참여기관 선정평가 세부운영지침
- 위 법령·규정·지침에 정의되지 않은 항목은 관계법령 및 규정 준수

□ 지원 데이터 분야

연번	데이터 분야	세부 요구 데이터	데이터(종)	예산(억원)	요구사항
1	자격증 문항	국가기술, 국가공인, 국가전문자격 등	1	13.5	25p
2	심화추론 문항	대학편입, 올림피아드, 각종 경시대회 등	1	10.0	26p
합계			2	23.5	-

※ 현재 제시된 예산은 최대 범위이며, 과제조정위원회 등을 통해 조정될 수 있음  
 ※ '26년도 예산 상황 등에 따라 선정 규모 및 평가 일정 등은 변경될 수 있음

### 3 수행기관 선정 방안 및 평가 기준 등

#### □ 수행기관 선정 및 사업수행계획서 평가

- (사전 적합성 검토) 사업수행계획서, 제출서류 및 자격요건 사전검토
  - 기한 내 접수 완료된 과제에 대해 사업수행계획서의 구비요건, 자격요건 등을 제출서류를 통해 사전검토하고 평가 대상 선정
  - ※ 사전검토는 「기금사업 협약체결 및 사업비 관리 등에 관한 지침」 제9조를 기준으로 실시
- (사업수행계획서 평가) 평가위원회의 구성 및 운영은 한국지능정보사회진흥원(이하 한국지능정보원) 관련 규정에 따라 시행
  - 평가위원은 수행기관의 사업수행계획서를 사전에 검토하고 피평가자의 프레젠테이션 발표 및 질의응답을 통해 평가(발표 : 15분 내외, 질의응답 : 15분 내외)
  - ※ 발표 시간은 수행기관 접수 결과에 따라 일부 조정될 수 있음
  - 수행기관 총괄책임자가 발표하는 것이 원칙이며, 총괄책임자 발표가 불가능할 경우 서면으로 평가
  - 평가점수 산출은 위원별 평가점수 중 최고·최저 점수를 제외한 나머지 점수를 평균하며 소수점 넷째 자리까지 산정(다섯째 자리에서 반올림)
  - ※ 동점자 발생 시 평가표의 '데이터 제공계획'의 배점 상위자, '품질관리 및 거버넌스'의 배점 상위자를 차례대로 우선 선정
  - 발표자료는 사업수행계획서 접수 시 제출한 자료 범위 내에서 활용 가능
  - ※ 제안서 접수 마감 후, 자료 수정제출, 추가 제출, 동영상 사용 불가
- (수행기관 선정) 평가결과 적합(70점 이상)인 과제에 대해 평가점수 순으로 순위를 부여하고 우수 수행기관을 선정 후 심의·조정 실시
  - ※ 사업수행계획서에 제안 분야 번호(연번) 및 분야명 반드시 기재
  - ※ 평가점수 70점 미만인 수행기관은 선정될 수 없음
  - ※ 공모참가자가 1개 업체일 경우에도 평가 및 심의를 통해 수행기관으로 선정될 수 있음

## □ 평가 기준

평가항목	세부항목	평가기준	배점
사업이해도 및 전략 (15)	사업목적 및 이해도	· 사업목적 및 독자 AI 파운데이션 모델과의 연계성 이해도 · 본 사업의 추진 배경 및 핵심 요구 사항에 대한 명확한 이해	5
	특화 데이터 추진 전략	· 전문·특화 데이터 전략의 차별성 및 적정성 · 도메인별 특성을 반영한 현실적인 목표 달성 전략	10
데이터 제공계획 (40)	데이터 권리 확보 사전 준비도	· 원천데이터 확보를 위한 계약, 협약, 제공동의서 등 증빙자료의 확실성 · AI 학습, 2차적 저작물 작성, 공동활용 개방 등에 대한 권리 해소 수준	15
	원천데이터 확보 방안	· 원천데이터 확보 방안의 타당성 및 실현가능성 · 최신성 및 도메인 다양성 확보 계획 · 데이터 확보 규모에 대한 적정성	15
	데이터 정제·가공·구조화 및 특화 기술	· 데이터 수집·정제·가공·구조화 체계의 구체성 · 멀티모달 확장 가능성 및 활용성 · 데이터 가공 특화 기술(CoT 데이터 등) 및 방법론의 우수성	10
품질관리 및 거버넌스 (20)	품질목표 및 검증 체계	· 품질목표 및 관리 지표의 적정성 · 품질관리 조직·절차·도구의 체계성 · 중복/유사도 필터링 및 도메인 전문가 교차 검수 체계	10
	법적 안정성 및 윤리 대응	· 개인정보·저작권·AI윤리 등 대응체계 · 문맥 보존형 가명처리 등 고도화된 비식별 처리 기술	10
수행역량 (15)	조직 및 인력 구성	· 전담조직 및 전문인력 구성	10
	추진체계 및 수행 이력	· 주관·참여기관 역할 분담의 적정성 · 유사사업 수행경험 및 전문성	5
성과 및 확산 (10)	성과 창출 및 파급효과	· 공동활용 전략의 구체성 · 산업적 파급효과 및 상용화 기여도	5
	유지관리 및 사후지원	· 유지관리 및 고도화 계획 · 데이터 오류 및 하자 발생 시 보완·수정 대책의 타당성	5
<b>합계</b>			<b>100</b>
<b>부정당업자 여부 (-30)</b>		· 국가계약법에 따른 부정당제재 조치 중인 업체	(-30)

## □ 과제조정위원회 심의·조정 및 사업수행계획서 확정

- 우선 지원 대상 수행기관은 한국지능정보원 관리지침에 따라 과제 조정위원회에서 수행내용, 예산의 적정성 등을 종합적으로 검토 받으며 그 결과에 따라 사업계획 조정 가능
  - 우선 지원 대상 수행기관은 과제조정위원회에서 검토한 사항을 사업수행계획서에 반영하고, 한국지능정보원은 과제조정위원회에서 검토·조정된 사항의 반영 여부를 확인하고 협약체결
- 우선 지원 대상 수행기관은 심의·조정 결과를 합리적인 이유 없이 거부하는 경우 협약을 포기하는 것으로 간주하고 후보 과제 중 평가점수의 차순위 수행기관에 지원 가능
  - 우선 지원 대상 수행기관은 통보받은 과제조정위원회의 조정 결과에 대해 이의 신청할 수 있으며, ICT기금사업 및 연구개발사업관리 지침에 따라 처리
- 사업수행계획서의 중복 또는 표절이 의심되는 경우(표절 검증 도구 등 활용), 과제조정위원회의 심의 결과에 따라 협약체결이 거부될 수 있음
  - 다만, 과제조정위원회에서 내용의 중복 등이 경미 하다고 판단하는 경우 사업수행계획서의 조정을 요청할 수 있음
- 과제조정위원회에서 표절로 판단하였거나, 의견을 제시하여 조정을 요청하였음에도 불구하고 합리적인 이유 없이 이를 거부하는 경우 후보 과제 중 평가점수의 차순위 수행기관에 지원할 수 있음

## 4 사업 요구 사항

### □ 공동활용 데이터에 관한 사항

- 수행기관은 제공데이터의 세부사항 및 법률검토 확인서를 사업 수행계획서와 함께 제출
  - 수행기관의 공급 권원 확인을 위해 법률검토확인서를 사업수행 계획서에 포함시켜 제출하되 검토내용이 미비하여 추가검토가 필요한 경우 보완요청할 수 있음
    - ※ 사업수행계획서 사전 적합성 검토 시 법률검토확인서에 대한 추가 확인 예정
  - 旣 제출한 사업수행계획서는 과제조정위원회와 한국지능정보원이 검토한 추가사항 등을 반영하여 최종적으로 확정된 후 사업 수행의 기준 문서로 활용
- 수행기관은 본 공고 내 세부 데이터별 요구사항을 만족하는 데이터셋 100% 물량을 1차 제출(~5월 4주)
  - 해당 데이터셋은 인공지능 모델 학습에 적합한 형태로 제출해야 함
    - ※ 세부 데이터 요구사항은 '붙임' 참조이며 제출시기는 사업 상황에 따라 변동될 수 있음
- 데이터셋 1차 제출 이후, 수요기관(정예팀)에서 요구하는 데이터셋 정제·가공에 대한 요구사항을 반영하여 최종 산출물 제출(~9월말)
  - ※ 가공에 대한 범위 및 규격은 수행기관-수요기관 간 협의에 따라 진행
- 1차 데이터셋 제출 이후, 세부 데이터별 요구사항에 대한 부합 여부를 검토하여 선금을 지급하며, 최종 품질검증 및 보완조치 완료 이후, 잔금지급
  - 데이터 최종 품질검증을 통해 도출된 결과와 내역에 따라 최종 품질 개선 또는 보완 조치를 하여야 함
  - 최종 품질검증 미달성 시, 수행기관 비용으로 자격을 갖춘 제3자 품질검증업체를 통한 재검사를 필수 시행
    - ※ 최종 데이터 제출 시기는 필요에 따라 조정될 수 있음

## □ 데이터 검증에 관한 사항

- 수행기관별 데이터 확보 시점 등이 유동적임을 고려하여 수시로 발생하는 공동활용 데이터에 대한 유연한 검증 시행
  - ※ 검증 대상, 검증 기준 및 범위 등은 데이터별 특성을 반영하여 검증기관과 협의하여 최종 확정
- 검증기관은 제공 데이터의 구성 및 품질의 검증 기준 등에 따라 수행기관별 데이터 검증을 수행하고, 필요 시 수행기관에 보완 요청
  - 검증기관은 보완 요청 시, 보완 필요 항목 및 범위, 기한 등을 명확하게 안내하고 재검증을 통해 이행 여부를 확인
- 한국지능정보원, 한국데이터산업진흥원 및 검증기관의 중간점검 및 최종 산출물의 품질검증 활동에 적극적으로 협조
  - 수행기관은 제공 데이터 관련 책임자 지정하여, 데이터 누락 및 오류 등에 대한 보완 요청 시 신속하게 대응할 수 있는 체계 마련

## □ 참여인력에 관한 사항

- 동 사업에 참여하는 작업자의 권익 보호 방안, 적정 임금 제공 방안 등을 사업수행계획서에 제시
  - ※ 직접 고용 인력(정규직, 계약직)은 4대 사회보험 가입 필수

## □ 공개 및 성과 확산에 관한 사항

- 수행기관이 제공한 데이터로 학습시킨 독자 AI 파운데이션 모델 및 이를 활용하여 생성된 모든 결과물은 정예팀이 상업적(산업적) 이용이 가능해야 함(필수)
- 수행기관은 데이터 공급에 따른 사업비를 제3자와의 거래조건, 홍보 등에 활용 불가

## □ 법적권리에 관한 사항

○ 수행기관은 데이터가 적법하게 구축되고 취득되었으며, 데이터 공급하는데 필요한 권리를 보유하고 있음을 보증

※ 타인의 지식재산권(저작권격권 포함) 및 기타 일체의 권리를 침해하지 않음을 보증

※ 대상데이터가 개인정보보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률 등 관련 법령 또는 제3자와의 계약상 어떠한 제한도 존재하지 않는다는 것을 보증

－ 인공지능(AI) 학습(Training)에 이용하는데 제한이 되거나 될 수 있는 데이터가 포함되지 않아야 함

※ 해당 데이터의 확보에 대한 개인정보수집·활용 동의서, 초상권 이용 동의서, 저작권 계약서 등 제출 필요

－ 구매 및 협약 등을 통해 원시 데이터 등을 확보하는 경우 사업 수행계획서 접수 시 증빙자료(계약서, 협약서, MOU, 데이터 제공 동의서 등) 제출

\* “원시데이터”란, 기계학습을 목적으로 최초 단계에서 수집 또는 생성한 음성, 이미지, 영상, 텍스트 등의 데이터로 전처리를 거치지 않은 데이터를 말함

－ 데이터의 개인정보(얼굴, 번호판 등), 국가보안사항(공간정보, 위치정보 등) 등이 포함된 경우 개인정보·민감정보 비식별화 조치

－ 협약 전에 데이터 수집, 가공, 공개와 관련한 법적사항에 대한 법률 검토 결과(법률검토확인서) 등의 제출 및 사전 심의 진행

－ 수행기관이 제공하는 데이터와 관련된 법률상 분쟁에 대해서는 수행기관에서 일체의 책임을 부담

※ 수행기관이 공급한 데이터에 대한 제3자의 지식재산권(저작권격권 포함), 초상권, 개인정보 등 권리 주장, 이의제기 또는 소제기 등을 하는 경우 수행기관의 비용과 책임으로 상대방이 입은 손해를 배상

## □ 개인정보보호 및 관리에 관한 사항

- 데이터 구축에 개인정보가 포함되는지 여부와 이를 포함하는 경우 반드시 개인정보의 구체적인 사항을 사업수행계획서에 포함
- 개인정보를 활용하는 경우, 개인정보보호법 상의 의무·권고사항에 대해 사전점검 및 법률 자문을 실시하고 그 결과를 사업수행계획서에 포함
  - ※ 인공지능 개인정보보호 자율점검표(개인정보위, '21.5.)에 따른 자가 점검을 실시하고 그 결과를 과제조정위원회에서 제출 및 검토
- 개인정보보호 및 관리를 위한 이행·점검·조치 방안을 사업수행 계획서에 구체적으로 제시
  - ※ 개인정보 비식별화를 시행하는 경우 개인정보보호위원회의 “가명정보 처리 가이드라인” 및 “비정형데이터 가명처리 기준('24.2.)” 준수
  - ※ 사업 착수 이전에는 “비정형데이터 개인식별 위험성 검토 체크리스트”를 활용하여 사전 검토 후 해당 위험을 낮추기 위한 적절한 조치 필요(가명정보 처리 가이드라인 참고)
  - ※ 과제 종료 후 최종 평가시 “비정형데이터 가명처리 결과에 대한 자체 검증 결과서”를 최종 보고서에 포함하여 제출 필요(해당 과제만 작성)
- 수행기관(주관기관, 참여기관) 담당자는 사업착수단계에서 한국지능정보원에서 시행하는 개인정보보호 교육 필수 이수
- 개인정보보호법의 심각한 위반사항 발생시 관련 법령 및 규정에 따라 협약 해지, 사업참여제한 등의 불이익을 받을 수 있음

## □ 사업관리 요구사항

- 일반적 사업관리 추진현황은 주간(격주)/월간 보고와 필요에 따라 수시보고로 진행하고 착수보고회, 중간점검, 최종결과평가를 실시
  - 사업 수행 중 중간 결과물(중간보고서), 최종 결과물(최종결과 보고서)을 확인하기 위해 현장실사 등을 통해 진행상황 점검
  - 최종 결과보고서, 정산보고서 및 증빙자료는 사업종료 후 30일 이내 제출
  - ※ 각종 보고서의 제출 시기는 사업관리 상 필요에 따라 조정될 수 있음

- 데이터 수집, 정제·가공 등은 수행기관(주관기관, 참여기관)에서 직접 수행하는 것이 원칙
  - 다만, 수행기관이 아닌 외부 기관의 용역을 통해 데이터 구축 등을 할 경우에는 제안 단계에서부터 사업수행계획서에 용역 계획을 사전에 수립하여 제출하고 과제조정위원회에서 조정한 범위 내에서 가능
  - 위탁용역비는 정부지원금에서 위탁용역비를 뺀 사업비의 40% 이내로 산정
- ※ 위탁용역·장비 구매 등 발주 시 ‘국가를 당사자로 하는 계약에 관한 법률’을 준용하여 사업비 집행

**【 위탁용역비 산정 기준 】**

- ▶ 위탁용역비 = 정부지원금 X 약 28.57%(7분의2) 이내
  - 정부지원금이 1억원일 경우 0.2857억원 이내 편성
  - ※ 주관/참여 기관별 산정(컨소시엄 전체 정부지원금의 28.57%를 의미하지 않음)
  - ※ 위탁용역비는 최소로 편성해야 하며 총사업비의 10% 이상인 경우 과제조정위원회에서 조정될 수 있음

- 과제 추진기간 중 과제목표 등 추진 상의 중대한 계획 변경이 있을 경우 정해진 절차에 따라 한국데이터산업진흥원에 사전 신청 및 한국지능정보원의 승인 후 시행 가능
- 참여인력 변경에 관한 사항은 한국데이터산업진흥원과 사전 협의 필수

**□ 사업비 산정 및 정산에 관한 사항**

- 사업비는 정보통신진흥기금운용관리 규정(과학기술정보통신부 고시)과 부속 지침 등에 따라 산정
  - ※ 동 사업은 비R&D 사업으로 간접비를 책정할 수 없음
  - 사업비는 과제조정위원회에서 예산산출 적정성 검토 후 최종 확정
  - 사업비(정부지원금 및 민간부담금 현금)는 다른 용도의 자금과 분리하여 전용 계좌 관리·운영하며 해당 계좌와 연결된 사업비 카드 또는 계좌이체 등을 통해 투명하게 집행

- 사업비 집행내역은 관련 증빙자료와 함께 관리하고 점검·정산 시 제출
  - ※ 사업비 사용 증빙은 5년간 보관하고 한국지능정보원이 요구 시 제출
- 투입인력은 정부지원금으로 인건비를 전액 반영할 수 있음
  - ※ 중앙정부, 지방자치단체 등으로부터 인건비 및 경상운영비를 100% 지원 받는 기관 등은 현금 산정불가
  - ※ 신규 인력의 인건비는 경력 등을 고려하여 수행기관 급여기준에 준하여 산정하되 통계청 기준 임금근로자 월평균 소득의 3배를 넘지 않아야 함. 단, 초과자는 이에 대한 사유 및 증빙 제출 필수
- 클라우드소싱 작업자 인건비는 반드시 예산항목 중 ‘일용임금’에 산정
  - 클라우드소싱 작업자의 대가는 과제종료 전에 집행된 부분만 인정
- 참여인력(클라우드 소싱 포함)은 원칙상 국내거주 대한민국 국적자로 제한
  - 단, 외국어를 모국어로 하는 해외 거주 외국인이 언어(음성, 자연어 등)와 관련된 데이터 구축 등을 위해 참여가 불가피한 경우, 과제 조정위원회가 조정한 범위 내에서 사전 승인을 받아 참여 가능
  - ※ 해당 참여 외국인의 신원, 계좌, 입금현황 등 예산집행의 투명한 증빙이 가능한 외국인에 한함
  - 국내 거주 외국인(유학생 연구원 등)의 참여가 불가피한 경우 과제 조정위원회에서 조정한 범위 내에서 참여 가능
- 고가의 장비 구매 및 자산취득은 최소화하고 임차 권고
  - ※ GPU 장비 구입 불인정, 클라우드 컴퓨팅 자원 임차를 원칙으로 함
  - ※ 구매가 필요한 경우 과제조정위원회를 통해 조정한 범위 내에서 구매 가능
- 자사 개발 솔루션, 자사 보유 데이터 등의 현물 출자, 다수 과제 참여시 자산의 중복 현물 출자는 불인정
- 정보통신기금 등 정부지원사업을 통해 획득한 자산은 현물 불인정
- 컨소시엄(주관, 참여기관) 수행기관 간의 내부 거래 불인정

- 한국지능정보원의 승인이 필요한 사업비 변경 또는 사용에 대해 사전 승인 절차를 거치지 않고 집행한 경우 사업비 불인정
- 사업비의 집행내역 검토 및 정산은 정보통신진흥기금 운용·관리 규정(과학기술정보통신부고시) 및 부속지침, 2026년 예산안 편성 및 기금운용계획안 작성 및 세부지침을 준용
  - ※ 상기 규정 등은 과학기술정보통신부 정보통신진흥기금 운용관리 규정 및 부속 지침의 제 개정 시 해당 지침으로 대체하여 적용
- 수행기관은 사업비 사용실적을 사업종료일로부터 30일 이내에 별도로 정하는 서식에 따라 한국지능정보원에 제출
  - 사업비 정산결과 사용잔액이 있거나 사업비를 부당하게 집행한 경우 해당 금액 중 출연금 지분에 해당하는 금액은 환수(사업비 사용 증빙은 5년간 보관)
  - 사업비 정산은 한국지능정보원이 지정한 전문회계법인을 통하여 위탁 진행하며 사업비 정산 비용은 주관기관에서 부담(사업비에 산정 필수)
    - ※ 「기금 사업비 산정 및 정산 등에 관한 지침」 참조하여 주관기관에서 일괄 산정
    - ※ 주관기관의 예산 편성이 어려운 경우 과제조정위원회 승인 필요
- 사업비의 부정청구 또는 편취한 경우 차년도 사업의 선정 평가대상에서 배제할 수 있음

## □ 기타

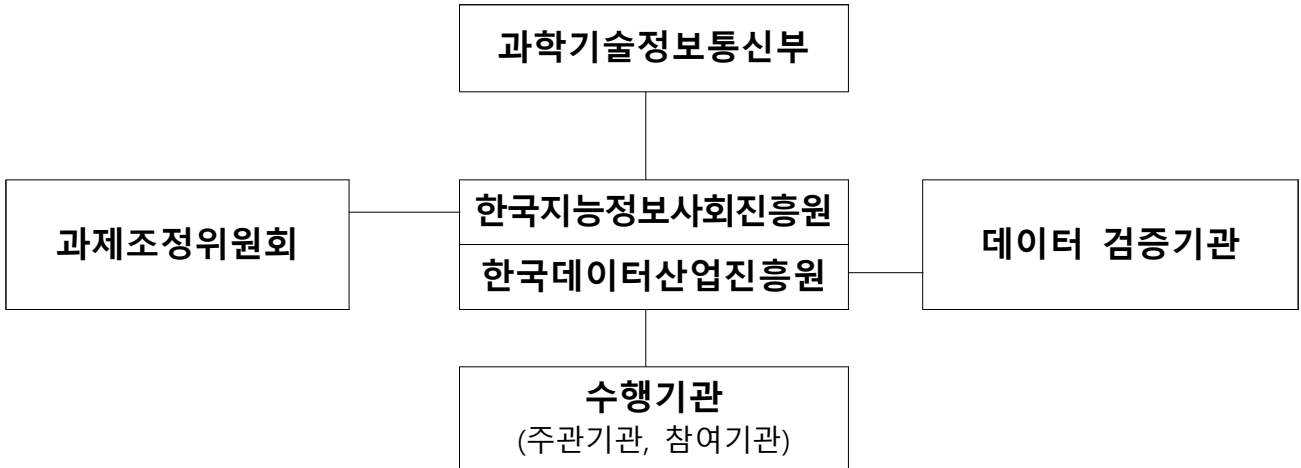
- 과제 참여 인원은 과제 수행 중 취득한 모든 결과물을 한국지능정보원의 승인 없이는 외부에 제공 또는 다른 용도로 활용 금지
  - 수행과정에서 발생할 수 있는 개인정보 노출, 자료 유출 등의 보안사고를 방지를 위한 보안정책을 수립
- 수행과정에서 천재지변, 전염병 등 불가항력적인 상황에 따라 과업 범위를 전부 또는 일부를 변경하거나 협약 해약 가능

○ 안전조치 및 보건 조치

- 분임 안전보건 책임자는 사업장의 환경을 고려하여 소속 근로자와 관계 수급인 근로자의 산업재해를 예방하는 데 필요한 안전조치 및 보건 조치에 대한 사업주의 이행을 확인하여야 함

## 5 사업 추진체계 및 절차 등

### □ 사업 추진체계

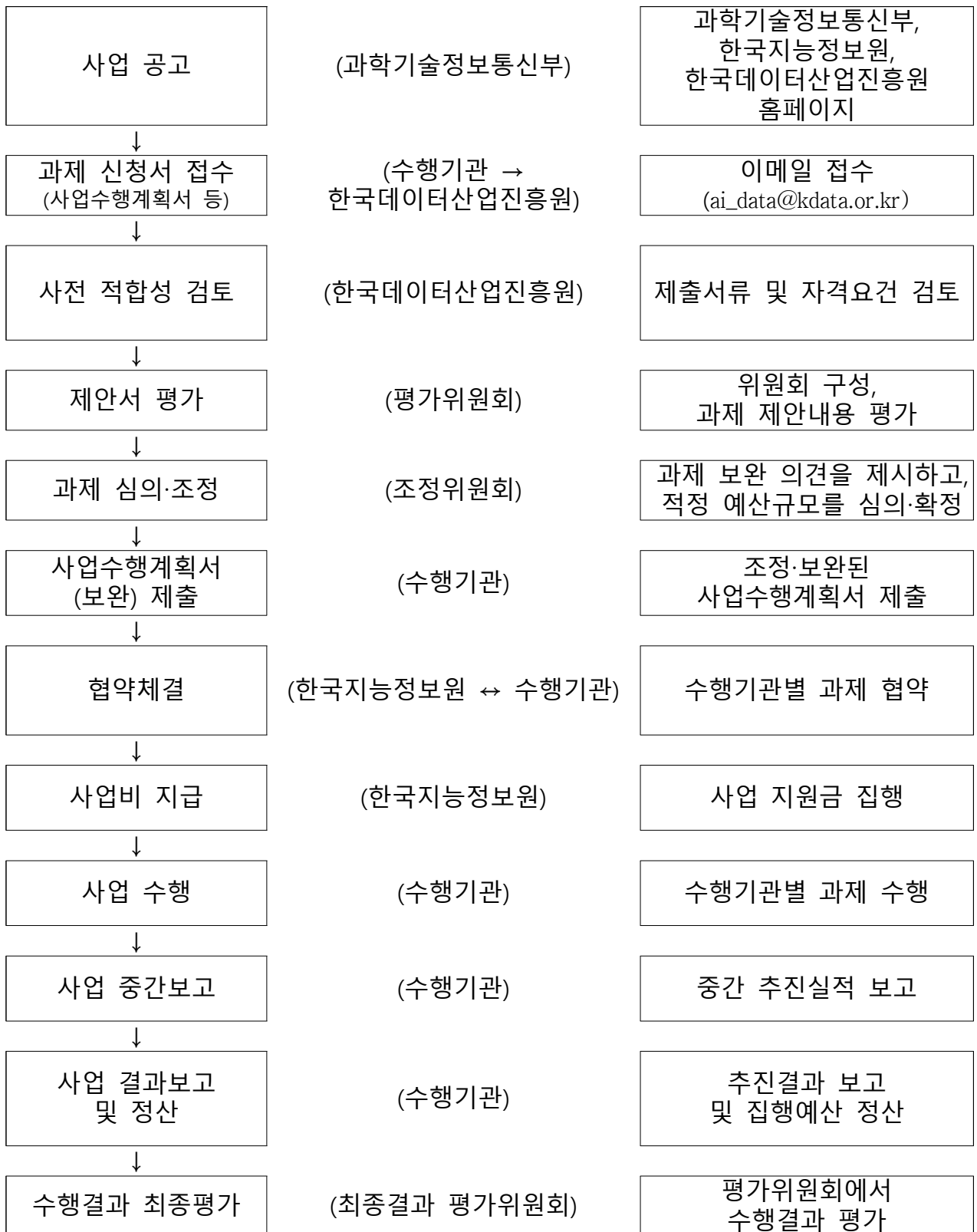


### □ 주요 역할

구분	주요 역할
과학기술 정보통신부	<ul style="list-style-type: none"> <li>· 사업 기본계획 및 추진전략 수립</li> <li>· 사업예산 확보 및 전담기관 배정</li> </ul>
한국지능정보 사회진흥원	<ul style="list-style-type: none"> <li>· 사업 세부계획 수립 및 시행 지원</li> <li>· 세부 사업 예산 내역 수립 및 운영</li> <li>· 사업 성과관리 및 홍보 지원</li> </ul>
한국데이터 산업진흥원	<ul style="list-style-type: none"> <li>· 공동활용 데이터 확보 및 제공 지원</li> <li>· 공동활용 수요 분석 및 조정</li> <li>· 공동활용 데이터 과제 운영 지원</li> <li>· 공동활용 데이터 품질 및 중복 검증 운영 관리</li> <li>· 수행기관 선정 계획 수립 및 선정위원회 운영</li> </ul>
과제조정 위원회	<ul style="list-style-type: none"> <li>· 사업수행계획 검토, 과제비 내역, 규모, 참여인력, 장비, 위탁계획 등의 적정성 검토 및 과제조정 의견 작성 및 제출</li> <li>※ AI 및 데이터 관련 기관·기업, 민간 전문가 등으로 구성</li> </ul>
데이터 검증기관	<ul style="list-style-type: none"> <li>· 공동활용 데이터에 대한 다양성, 정확성 등 품질 검증 수행</li> <li>· 공동활용 데이터 오류, 누락 등에 대한 보완 조치 요구 및 확인</li> </ul>
수행기관 (주관, 참여)	<ul style="list-style-type: none"> <li>· AI 학습용 공동활용 데이터 제공</li> <li>· 데이터 정제·가명처리·가공·표준화 등의 수행</li> </ul>

※ 데이터 검증기관은 한국데이터산업진흥원에서 지정

□ 추진 절차



## 6 제안서 접수 및 방법, 서류 등

☞ 규격 착오 또는 규정의 미숙지 등으로 수행기관이 협약을 체결하지 않거나 협약을 체결하고 불이행하는 경우 향후 일정기간 동안 “한국지능정보사회진흥원 ICT 기금 사업 및 연구개발사업 관리지침”에 의하여 공모사업에 참여 제재를 받을 수 있음

### □ 접수 기간

분야 (과제)	접수기간	제안서 접수 (온라인)
1~2번	'26.4.24(금)~5.4(월)	5.4.(월) 13:00시한

※ 다수 분야에 참여 시 참여 한도인 2개 데이터 분야는 접수완료 기준 순으로 인정하며, 그 이후 접수는 불인정(접수처의 최종 제안서 제출 송신시간 기준)이 원칙이나 접수시스템 기능 오류 등 기타 K-DATA가 인정하는 부득이한 사정으로 기한 내 접수하지 못했을 경우, 별도 안내에 따라 신청서류를 다시 제출하여 접수할 수 있음(해당 시 별도 안내)

### □ 접수방법

- 사업수행계획서 접수 : 전자우편을 통한 온라인 접수(ai\_data@kdata.or.kr)
  - PC 및 네트워크 등 사용자 환경에 따라 관련 서류 제출의 어려움이 발생할 수 있으므로 사업수행계획서는 최소 3일 전 사전 제출 권장
  - 사업수행계획서 접수는 온라인 제출만 인정

### □ 제출서류 : 별도 첨부 서식 활용

- 사업수행계획서 및 발표자료 각 1부
- 사업자등록증 사본(주관기관, 참여기관 포함) 각 1부
- 법인등기부등본 및 인감증명서(주관기관, 참여기관 포함) 각 1부

- 주관기관, 참여기관 참여의사 확인서 각 1부
- 자격요건 자가점검표(주관기관, 참여기관 포함) 1부(서식 파일 참조)
- 최근 1년 재무제표, 국세 및 지방세 납입증명서(주관기관, 참여기관 포함) 각 1부
  - ※ (유의) 직전년도 재무제표 미제출 또는 완전자본잠식 등에 해당하는 경우 평가배제 등 불이익을 받을 수 있음(컨소시엄 구성 및 공고 접수 이전에 확인 필요)
  - ※ 추후 협약체결 확정시 필요한 서류를 제출하여야 하며 기타 필요한 서류를 함께 요청할 수 있음

#### □ 선정평가 장소 및 일정(안)

- 장소 : 한국데이터산업진흥원(서울 중구)
- 일정 : 5. 8.(금) 실시 예정(기관 개별 통보)

#### □ 유의사항

- 사업수행계획서 제출 마감 이후 사업수행계획서 및 발표자료 변경 불가
- 공모에 필요한 모든 파일은 PDF 형식 단일파일로 각각 만들어 한 개의 파일로 제출하고 제안서 마감 이후 발표자료 및 수행계획서 변경, 추가 제출은 불가능
- 공모에 필요한 모든 사항(공고서, 관련 규정 및 지침 등)에 대하여 공모 전에 확인·숙지하고 확인·숙지하지 못한 책임은 수행기관에게 있음
- 마감일시까지 온라인 창구(ai\_data@kdata.or.kr)에 사업수행계획서 미제출 또는 첨부파일의 하자(오류)가있을 경우 제출하지 않은 것으로 간주  
- 단, 발표자료 및 제안요약서 등 미제출시에는 수행계획서로 평가
- 중복참여에 따른 공모 무효에 관하여는 공고서, 국가를 당사자로하는 계약에 관한 법률 시행규칙 제44조(입찰무효) 준용하여 처리

※ 중복구성의 판단 기준은 법인번호, 법인대표자의 동일성이며, 개인사업자의 경우 사업자등록증상 대표자의 동일성, 고유번호증의 경우 고유번호증의 대표자의 동일성

○ 제출된 사업수행계획서 등 제출물 일체는 반환하지 않음

□ 선정평가 결과 확인

○ (결과발표) 개별 통보 예정(5월 11일(월) 예정)

○ 단, 최종 적격과제 중 협약 전 불가피한 사유로 과제 참여를 포기하거나, 허위사실 등 중대한 결격사유가 발생했을 시 후순위자에게 기회를 부여함

□ 문의사항

○ 한국데이터산업진흥원 AI데이터지원팀 담당자 :  
02-3708-5441, 02-3708-5442

## 7 주요 추진 일정[안]

- '26. 5월 4일 13시 : 과제 신청서 접수 마감
- '26. 5월 2주 : 사전 적합성 검토 및 제안서 평가
- '26. 5월 3주 : 과제 심의·조정
- '26. 5월 4주 : 협약체결, 데이터셋 1차 제출 및 사업 착수
- '26. 7월 초 : 정예팀 요구사항 취합 및 반영
- '26. 8월 초 : 데이터 품질지표 수립 및 검증
- '26. 9월 중 : 최종 산출물(데이터) 제출
- '26. 9월 30일 : 과제 종료

※ 추진 일정은 일부 변경될 수 있음

□ **공통사항**

항목		요구사항
데이터 권리획득 및 법제도적 문제 해결	권리확보	<ul style="list-style-type: none"> <li>· 원천데이터 확보 시 원저작자 동의 필수 확인</li> <li>· 2차적 저작물 작성 허용 포함 및 AI 허브 등 공개 시 이슈가 없도록 사전 조치</li> <li>· 제안 시 관련 증빙자료(계약서, 협약서, 제공동의서 등) 필수 제출</li> </ul>
	합성데이터 관리	<ul style="list-style-type: none"> <li>· 원천데이터 주관기관의 데이터 공유 허락 및 별도 제공을 명시한 공식 문서 제시 필수</li> </ul>
	법제도 준수	<ul style="list-style-type: none"> <li>· 지식재산권, 개인정보, 특허권, 초상권 등 공개 제약 해소 방안 제시</li> <li>· 비윤리적 내용 배제 방안 마련</li> </ul>
	생명윤리 심의	<ul style="list-style-type: none"> <li>· 인간 대상 과제(보건의료 등)는 IRB 및 DRB 승인 등 적법 절차에 따른 확보 방안 제시</li> </ul>
비식별화 및 가명처리 방안	가이드라인 준수	<ul style="list-style-type: none"> <li>· 개인정보(얼굴, 번호판 등) 및 국가보안사항(위치정보 등) 포함 시 비식별화 조치 방안 제시</li> <li>· 개인정보보호위원회 등 소관 부처의 최신 가명정보·보건의료 데이터 처리 가이드라인 엄격 준수</li> </ul>
	문맥 보전 가명처리	<ul style="list-style-type: none"> <li>· 단순 삭제 및 특수기호 마스킹(예: 000) 지양(문맥 단절 방지)</li> <li>· 원본 정보의 속성에 부합하는 자연스러운 가상 식별자로 치환하여 논리적 구조 보존</li> </ul>
데이터 정제 및 유사도 관리	중복방지	<ul style="list-style-type: none"> <li>· 데이터 간 중복 및 유사 내용 방지 대책 마련</li> <li>· 유사도 측정 등을 활용하여 중복 필터링 필수 수행</li> </ul>
원천데이터 적정성 및 목표 제공량 산정	부적합 데이터 필터링	<ul style="list-style-type: none"> <li>· 주제와 관련된 적절한 데이터 선별 필수</li> <li>· 문장 수집 전 문서·도서 단위 분류를 선행하여 부적합 데이터 유입 원천 차단</li> </ul>
	제공수량 기준	<ul style="list-style-type: none"> <li>· 수행계획서 내 세부 데이터별 제공 수량 및 측정 단위를 명확히 제시</li> </ul>
	생성형 AI 제약	<ul style="list-style-type: none"> <li>· 본 과업은 실제 데이터 기반의 원본 확보가 원칙</li> <li>· 생성형 AI로 인위 생산한 데이터는 제공 수량으로 인정 불가</li> </ul>

□ 데이터 특성별 요구사항

① **추론형AI 학습 데이터** : 초거대 모델의 추론(Reasoning) 능력 학습을 위해 논리적인 사고와 구조화된 단계를 통해 답변 및 행동을 도출하는 CoT\* 데이터 구축

\* CoT(Chain of Thought) : 문제 해결 또는 질문에 답할 때 단계별 추론 과정을 설명하며 답변을 생성하는 방식

※ 단답식이 아닌 복합적인 추론을 요구하는 문제를 만들고, 각 추론 단계별로 최적화된 사고 과정을 선별 후 작성하여 답변까지 도출하는 적절한 방안과 이를 검증하는 방안을 제시

※ CoT 데이터는 각 데이터별 특성에 맞추어 질문, 답변을 구성하고 전문적인 지식이나 고도의 추론 과정이 포함되도록 사람이 작성한 CoT 기반의 데이터로 구축할 것

② **초거대 언어모델(LLM) 학습 데이터** : 초거대 언어모델 학습을 위한 대규모 말뭉치 및 미세조정을 위한 Instruction 데이터 구축

③ **초거대 멀티모달모델(LMM) 학습 데이터** : 여러 Modality를 함께 활용한 학습 또는 Cross-Modality 등을 위한 학습 데이터 구축

※ 초거대 멀티모달모델(LMM) 학습 데이터는 각 모달리티의 데이터를 함께 활용하여 AI모델 학습할 수 있도록 데이터셋을 구성하는 적절한 방안과 이를 검증하는 방안을 제시

연번	데이터명	추론형 학습데이터	LLM 학습데이터	LMM 학습데이터
1	자격증 문항	○	○	○
2	심화추론 문항	○	○	○

## □ 데이터① : 자격증 문항

<b>목적</b>	AI 모델의 추론 및 전문 지식 해결능력 향상을 위해 고품질 자격시험 문항 데이터 확보			
<b>사업기간</b>	계약 체결일로부터 '26.9.30.까지	<b>예산</b>	1,350백만원 (※ 문항 분류, 난이도 라벨링, 비식별화 등 비용 포함)	
<b>요 구 사 항</b>	<b>대상 범위</b>	1) 국가기술자격 : 기사·산업기사(정보처리, 전기, 산업안전 등) 2) 국가전문자격 : 전문직 시험(회계사/세무사 1차, 공인중개사 등) 3) 국가공인자격 : 경제/경영(TESAT 등), 어문학, 실무(유통관리사 등) 4) 기타(민간) : 등록/공인 민간자격, 기업/기관 주관 직무능력 인증시험 등		
	<b>생성 기간</b>	2021년 이후에 생성된 데이터	<b>제공 형식</b>	JSON, CSV, xml 등
	<b>제공 수량</b>	최소 13.5만 세트(문제+정답+해설) 이상	<b>문항 구성</b>	대상범위(1~4) 중 1개 이상의 분야만으로 전체 수량 구성 허용 ※ 단, 데이터의 다양성을 위해 단일 자격 종목(예: 특정 기사 1과목)의 문항 수가 전체 납품 수량의 50%를 초과할 수 없음
	<b>문항 유형</b>	1) 객관식 60% 이상 2) 서술형(단답형 포함) 40% 이내	<b>난이도 분포</b>	쉬움 10%, 보통 50%, 어려움 40% 수준 권장
	<b>구성항목</b>	문항별 ① 자격종목 및 과목 분류 ② 지문(Context) ③ 문항(Question) ④ 정답(Answer) ⑤ 논리적 추론 과정(CoT 및 소거 논리) ⑥ 난이도 메타데이터(판정 기준 및 근거자료 포함)		
	<b>처리 요건</b>	1) 추론형(CoT) 해설 강화 : 단순 정답 나열을 엄격히 금지하며, '단계별 풀이 과정 (Step-by-step)'과 '오답 보기에 대한 배제 근거'를 반드시 명시 2) 수식/도표의 마크다운화: 고난도 수학, 물리, 화학 등의 수식 및 기호, 회로도 내 수치 등은 반드시 LaTeX 및 Markdown 포맷으로 텍스트화하여 철저히 검증 3) 계층적 매핑 : 하나의 복합 지문에 여러 문항이 딸려 있는 경우, 지문과 문항 간의 매핑 데이터를 파싱 가능한 구조로 명확히 연결 4) 비식별화 : 출제위원 이름, 소속 등 개인 식별 정보 철저히 제거 및 문맥 보존형 가명처리		
	<b>보유 권원</b>	수행기관은 제공되는 문항 데이터의 AI 학습 활용을 위한 복제권, 배포권, 공중송신권 (전송권), 2차적 저작물 작성권 일체를 보유하고 있어야 함		
	<b>이용 기간</b>	계약 체결일로부터 2027.12.31.까지 ※ 단, 데이터가 학습된 상태로 모델에 남아있는 기간은 영구	<b>활용 범위</b>	상업적·비상업적
	<b>기타 사항</b>	1) 문항 데이터에 대한 스키마 정의서, ERD, 메타데이터 등 별도 제공 2) 그래프, 도표, 삽화 등 모든 이미지 자료는 원본 파일과 함께 제공 3) 중복 문항은 제외하고 납품해야 하며, 발견 시 대체 문항 재납품		

## □ 데이터② : 심화추론 문항

<b>목적</b>	AI 모델의 복합 추론 능력 향상을 위한 고난이도 문항 데이터 공급			
<b>사업기간</b>	계약 체결일로부터 '26.9.30.까지	<b>예산</b>	1,000백만원 (※ 문항 분류, 난이도 라벨링, 비식별화 등 비용 포함)	
<b>요구사항</b>	<b>대상 범위</b>	1) 대학편입 : 상위권 대학 편입 수학, 편입 영어(독해/논리) 2) 경시/올림피아드: 수학, 물리, 화학, 정보 분야 등 경시대회 문항 3) 심화추론평가 : LEET, PSAT 등 복합 언어/논리 추론 문항		
	<b>생성기간</b>	2021년 이후에 생성된 데이터	<b>제공 형식</b> JSON, CSV, xml 등	
	<b>제공 수량</b>	최소 2만 세트(문제+정답+해설) 이상	<b>문항 구성</b>	분야의 제한 없이 수행기관이 전문성을 가진 과목(예: 수학 및 과학 집중)으로 전체 수량 구성 가능 ※ 단 편입 경시 추론평가 등 출처에 상관없이 최소 3개 이상의 상이한 과목(예 수학 언어 물리)이 포함되어야 함
	<b>문항 유형</b>	1) 객관식 50% 이상 2) 서술형(단답형 포함) 50% 이내	<b>난이도 분포</b>	어려움 100% (고난이도 특성 반영)
	<b>구성항목</b>	문항별 ① 출처 및 과목 분류 ② 지문(Context) ③ 문항(Question) ④ 정답(Answer) ⑤ 논리적 추론 과정(CoT 및 소거 논리) ⑥ 난이도 메타데이터(판정 기준 및 근거자료 포함)		
	<b>처리 요건</b>	1) 추론형(CoT) 해설 강화 : 단순 정답 나열을 엄격히 금지하며, '단계별 풀이 과정 (Step-by-step)'과 '오답 보기에 대한 배제 근거'를 반드시 명시 2) 수식/도표의 마크다운화: 고난도 수학, 물리, 화학 등의 수식 및 기호, 회로도 내 수치 등은 반드시 LaTeX 및 Markdown 포맷으로 텍스트화하여 철저히 검증 3) 계층적 매핑 : 하나의 복합 지문에 여러 문항이 포함된 경우, 지문과 문항 간의 매핑 데이터를 파싱 가능한 구조로 명확히 연결 4) 비식별화 : 출제위원 이름, 소속 등 개인 식별 정보 철저히 제거 및 문맥 보존형 가명처리		
	<b>보유 권원</b>	수행기관은 제공되는 고난이도 문항 데이터의 AI 학습 활용을 위한 복제권, 배포권, 공중송신권(전송권), 2차적 저작물 작성권 일체를 보유하고 있어야 함		
	<b>이용 기간</b>	계약 체결일로부터 2027.12.31.까지 ※ 단, 데이터가 학습된 상태로 모델에 남아있는 기간은 영구	<b>활용 범위</b>	상업적·비상업적
	<b>기타 사항</b>	1) 문항 데이터에 대한 스키마 정의서, ERD, 메타데이터 등 별도 제공 2) 그래프, 도표, 삽화 등 모든 이미지 자료는 원본 파일과 함께 제공 3) 중복 문항은 제외하고 납품해야 하며, 발견 시 대체 문항 재납품		