

# AI 학습용 데이터 구축 비용산정 가이드

## 발간사



데이터의 시대입니다. 우리의 일상생활과 산업환경에서 생산된 다양한 데이터들이 유기적으로 연결되어 다시 생활 곳곳에서 활용되는 새로운 서비스를 창출하는 기반이 되고 있습니다. 이처럼 데이터 시대에는 활용할 수 있는 양질의 데이터가 필수적입니다. 국민들이 필요로 하는 데이터를 선제적으로 발굴·구축하고 개방하는 것은 한국지능정보사회진흥원(NIA)이 꾸준히 해 온 일이기도 합니다.

이번에 발간하는 「AI 학습용 데이터 구축 비용산정 가이드 v1.0」은 이러한 노력의 일환으로 준비하였습니다. AI 학습용 데이터에 관심 있는 지자체, 학계 등 공공 영역과 AI 시장 진입을 원하는 기업들이 데이터 구축 과정에 필요한 작업내용과 소요 예산을 사전에 점검하고, 적정 비용으로 사업을 추진할 수 있도록 방향성을 제공하고자 합니다.

본 보고서는 NIA가 다년간 수행한 정부 사업 자료를 바탕으로 구성하였습니다. AI 학습용 데이터의 대표적인 4가지 데이터 유형인 '텍스트, 이미지, 음성, 비디오' 각각의 데이터 구축 단계별 작업 내용을 분석하여 제시하였고, 각 작업 수행 시 예상되는 비용을 산정하기 위한 지침을 담았습니다. 이를 통해, 데이터 구축 과정의 시행착오를 줄이고, 나아가 관련 지식의 확산을 통해 기업 간에 투명하게 경쟁할 수 있는 AI 시장이 조성되기를 기대합니다.

아울러, 본 가이드 발간을 위해 진심 어린 의견을 보내주신 AI 분야 공공·민간의 각계 전문가 모든 분께 감사드립니다. 본 가이드가 디지털 플랫폼 경제 시대에 국민이 필요로 하는 공공 서비스를 창출하고, AI 산업 경쟁력을 높이는 데 귀중한 자원으로 활용되기를 희망합니다. 감사합니다.

2022년 9월

한국지능정보사회진흥원장 **황종성**

# Contents

## I 개요

1. 목적 및 배경 .....	3
2. 가이드 구성 체계 .....	4
3. 적용 범위 .....	5
4. 향후 계획 .....	6

## II 주요 개념 및 용어 해설

1. 데이터 특성 및 구조 .....	9
2. 비용산정 요소 .....	11
3. 비용 구성 요소 .....	13

## III AI 학습용 데이터 구축 비용산정

1. 비용산정 절차 .....	17
2. 데이터 4개 유형별 비용산정 절차 및 상세 내용 .....	24

AI 학습용 데이터 구축  
비용산정 가이드

# I 개 요

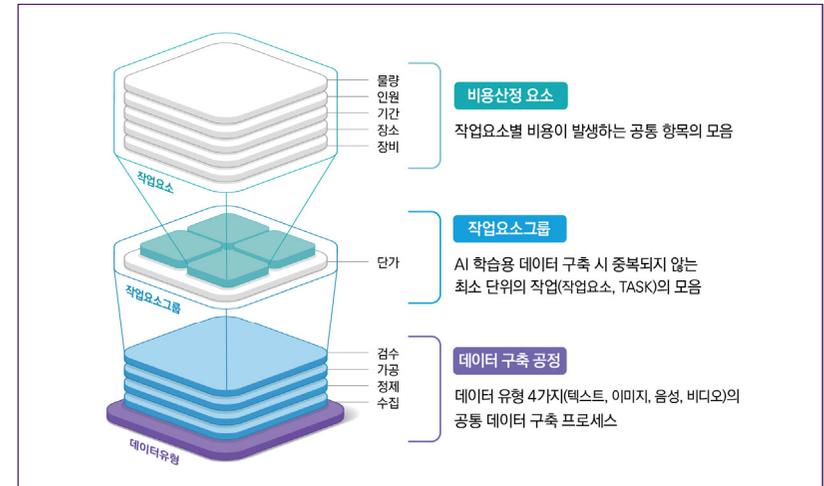
## 1 목적 및 배경

- ▶ AI 학습용 데이터 구축 과정을 사전에 점검하여 필요한 작업 내용과 소요예산을 확인하고 적정 비용을 산정할 수 있도록 지원
  - AI 학습용 데이터 구축은 많은 시간과 비용이 소요되는 작업으로, 추진시 시행착오를 줄이고 적정 비용을 산정할 수 있는 표준화된 체계 필요
  - 한국지능정보사회진흥원(이하 NIA)는 대규모 「AI 학습용 데이터 구축 사업」의 수행기관(’20년-)으로, 데이터 구축 노하우를 대국민에게 공개하여 국내 AI 산업 발전에 기여
    - ※ ’20년 170종(3,315억), ’21년 190종(3,725억), ’22년 310종(5,700억) 구축 추진
- ▶ AI 기술에 관심 있는 연구자, 산업계 등에서 참고할 수 있는 가이드를 제공하여 원활한 시장 진입을 유도하고 건전한 시장 환경 조성
  - 국내 AI 산업은 초기단계로, 기업 등에서 필요로 하는 AI 학습용 데이터의 구축 노하우는 기업 자산으로 인식되어 시장에서 관련 정보 확산에 한계
  - 전통산업을 비롯하여 신규 AI 시장 진입을 고려중인 기업에게는 높은 진입장벽으로 작용하는 등 다양한 시장 환경 조성을 위한 정보 공유 필요
- ▶ 국내 AI 학습용 데이터는 정부 주도로 대규모 데이터 구축·개방 사업이 추진되고 있으며, 축적된 자원을 활용하여 표준 가이드 제시
  - NIA는 「AI 학습용 데이터 구축 사업」의 실제 수행 내용을 바탕으로 고품질 데이터 구축을 위한 안내서 2종을 제작하여 대국민 공개
    - ※ 「AI 학습용 데이터 품질관리 가이드라인 v2.0(’22.1)」 2종(1권 품질관리 안내서, 2권 데이터구축 안내서)을 제작하여 대국민 공개 및 매년 업데이트 진행
  - 본 ‘비용산정 가이드’는 안내서의 후속작업으로, 데이터 구축 과정 중 예산이 투입되는 작업요소를 분석하여 참고 기준을 제시

본 가이드는 정부 사업인 「AI 학습용 데이터 구축 사업(정보통신진흥기금)」 자료를 바탕으로 구성되어, 일부 내용은 본 사업 특성이 반영되어 있으나, **보편적인 ‘AI 학습용 데이터’ 구축 비용산정 과정**을 제시하고자 함

## 2 가이드 구성 체계

- ▶ 본 가이드는 AI 학습용 데이터 구축에 관심 있는 기업·기관·학계 등을 대상으로 구축 과정의 비용 요소를 총체적으로 제시
  - I 장은 본 「AI 학습용 데이터 구축 비용산정 가이드」의 총괄적인 내용을 소개하는 장으로, 추진 목적, 추진 배경, 구성 체계, 적용 범위 및 향후 계획을 정리하여 구성
  - II 장은 주요 용어들을 정의하고 해설하는 장으로, AI 학습용 데이터의 특징과 구조, 비용산정 시 고려되는 요소에 대한 내용으로 구성
  - III 장은 데이터 구축 비용산정을 위한 프로세스를 제시한 장으로, 비용산정 절차, 데이터 4개 유형별 작업요소와 작업요소별 설명으로 구성



| 그림 1-1 | 가이드의 AI 학습용 데이터 구축 비용산정 구조

### 3 적용 범위

- ▶ 본 가이드는 활용 가능한 AI 학습용 데이터를 도출하는 전 과정 중 예산 투입 비중이 높은 데이터 구축 과정으로 한정하여 산정 기준 제시
  - 구축 과정 중 **데이터 수집**부터 AI 알고리즘 학습에 필요한 **데이터 한 쌍**(원천 데이터+라벨링 데이터)을 생성\*하는 단계까지의 작업요소와 소요비용을 분석
    - \* 구축 절차에 따라 품질에 결격 사유가 없는 데이터 한 쌍을 도출했을 경우, 현장에서 활용가능한 수준의 데이터라고 할 수 있음
    - ※ AI 학습용 데이터를 구축하기에 앞서, 과업에 대한 임무정의 및 구축 기획·설계 과정이 선행되어야 하며, 해당 과정에 대한 비용은 본 가이드에 미반영
  - '작업요소'는 데이터 구축 시 중복되지 않는 최소 단위의 작업(TASK)으로, 데이터 유형별 필요한 작업과 각 소요비용 산정 기준을 제시



| 그림 1-2 | AI 학습용 데이터 생애주기 중 구축과정

본 가이드는 '21년 AI 학습용 데이터 구축 사업 공모시 과제별 사업 예산을 산출하기 위해 작성된 자료 (총 42개 과제)를 바탕으로 산정기준을 제시함

### 4 향후 계획

- ▶ 본 가이드는 현재 활용 가능한 자원에 기초하여 제작한 것으로, AI 학습용 데이터 구축 시장에 표준화된 비용산정 기준이 전무한 상황에서 방향 제시를 할 것으로 기대
- ▶ AI 시장에서 통용 가능한 수준의 가이드 제작을 위해, AI 학습용 데이터 구축 현장 상황을 반영하여 가이드를 고도화할 예정

AI 학습용 데이터 구축  
비용산정 가이드

## II

### 주요 개념 및 용어 해설

## 1 데이터 특성 및 구조

- ▶ AI 학습용 데이터는 AI 알고리즘 학습을 위한 데이터로, 학습 목적에 따라 구축 방식의 차이가 있어, 이를 고려한 비용 산정 필요
  - 본 가이드는 AI 학습용 데이터의 특성을 반영하여, AI 데이터의 유형, 구축 절차 및 작업 요소 단위로 세분화하여 비용 산정 기준 제시
- ▶ AI 학습용 데이터는 구축 데이터의 특성에 따라 대표적으로 4가지 유형으로 구분되며 유형별 세부적인 구축 과정의 차이 존재
  - 기계학습을 목적으로 자료 수집(획득) 또는 생성한 원시데이터는 일반적으로 4가지 데이터 유형(텍스트, 이미지, 음성, 비디오(영상))으로 구분
  - 데이터 유형에 따라 적합한 라벨링 유형 선택이 필요하며, 이에 따라 세부적인 구축 과정의 차이 발생
    - ※ 멀티모달(multimodal) 데이터는 두 가지 이상의 데이터 유형이 결합된 것으로 본 가이드에는 미반영
- ▶ 데이터 유형별로 차별화된 구축 과정의 수행 업무를 세분화하고, 각 작업별 비용 발생 요소를 분석하여 비용 산정 방향 제시
  - 기계학습에 활용하기 위해 모든 AI 학습용 데이터는 공통적으로 '수집 및 정제, 가공, 검수' 단계의 구축 과정(작업요소그룹)을 통과
    - ※ 각 과정은 기본적으로 순차적이나, 데이터 특성에 따라 동시적으로 진행되기도 하며 본 가이드에서는 개별 단계가 구분되지 않는 경우 같이 기술
  - 작업요소그룹의 하위 작업요소는 한 쌍의 AI 학습용 데이터 파일을 구축하기 위한 수행 업무(Task)를 의미하며, 각 업무는 비용과 관계된 활동을 포함
    - ※ 작업요소 예 : 영상촬영, 장비 임차, 콘텐츠 구매 등
  - 작업요소(Task)에서 비용이 발생하는 공통 항목을 도출하여 '비용산정 요소'로 명명하고, 요소 간 조합을 통해 비용 산정 산식 제시

표 2-1 | AI 학습용 데이터 공통 구축 과정(작업그룹)

작업그룹	내용	용어 정의
데이터 수집 (획득)	원시데이터 수집	AI의 기계학습에 필요한 데이터를 현실 세계에서 직접 수집 또는 생성하거나, 이미 보유하고 있는 조직이나 시스템 등으로부터 법률적 제약이 없도록 '원시데이터'를 확보하는 활동
데이터 정제	원천데이터 생성	획득한 원시데이터를 기계학습에 필요한 형식으로 맞추거나 불필요한 중복을 제거하며, 개인정보를 비식별화하여 처리하는 등 일련의 전처리 과정을 통해 '원천데이터'를 확보하는 활동
데이터 가공 (라벨링)	라벨링데이터 생성	AI가 기계학습에 활용할 수 있도록 기능이나 목적에 부합하는 정보를 원천 데이터에 부착(라벨링)하는 활동
데이터 검수	데이터 검사	사전에 설정한 데이터 작업물(결과물)의 통과 기준(참값), 반려 기준 등을 바탕으로 데이터를 검사하고 최종 파일을 확정하는 활동 ※ 데이터 검수(품질검사)는 최종 파일을 도출하는 전수검사 뿐만 아니라 단계별 데이터가 생성될 때마다 교차 검사 병행

※ 구축 작업그룹에 대한 자세한 내용은 「인공지능 학습용 데이터 품질관리 가이드라인 V2.0」 제1권 품질관리 안내서, 제2권 데이터구축 안내서 참고

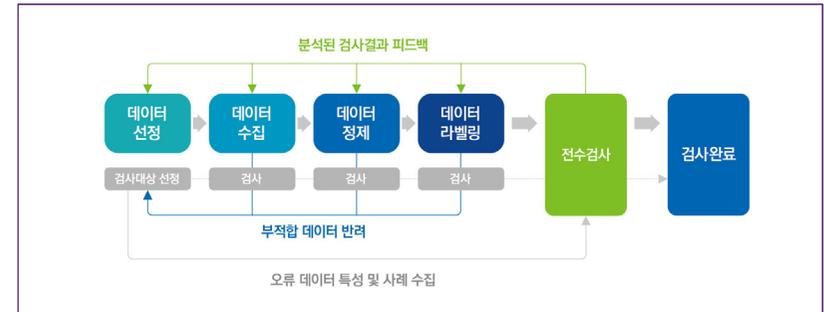


그림 2-1 | 단계별 구축 데이터 검수 절차

## 2 비용산정 요소

### 1) 비용산정 요소

▶ 비용산정 요소란, 구축 작업별 비용을 구성하는 항목 중 공통내용을 정의한 것으로, 작업 요소별 소요비용 산식의 단위 항목으로 활용

- 비용산정 요소는 '물량', '인원', '기간', '장비', '장소'의 5가지 항목으로 구성되며, 각각의 세부 정의는 아래 내용 참조

- 각 비용산정 요소는 요소별 단가 등과 결합하여 소요비용 산식을 구성하며, 산식의 단위가 '투입공수'인 경우, '투입공수'의 항목으로 활용

- ① 물량 : 데이터 구축 목표량, 작업량 등 작업의 규모를 의미
- ② 인원 : 구축 작업에 참여하는 작업자를 의미
- ③ 기간 : 구축 작업에 소요되는 시간을 의미
- ④ 장비 : 구축 작업에 활용되는 기기를 의미
- ⑤ 장소 : 구축 작업이 진행되는 공간을 의미



| 그림 2-2 | 작업요소의 비용산정 요소

### 2) 구축 대상 자료의 기준 단위

▶ AI 학습용 데이터의 구축량을 정확하게 측정하고, 목표 달성을 위한 작업 요소와 비용을 산출하기 위해 구축 대상 자료의 단위 통일 필요

- 본 가이드에 등장하는 4가지 유형 데이터의 구축 작업(수집, 정제 등) 시 작업량 측정의 기준이 되는 단위를 제시

- ① 건 사건, 서류, 안건 따위를 세는 단위
- ② 개 낱으로 된 물건을 세는 단위
- ③ 곳 장소를 세는 단위
- ④ 문장 생각이나 감정을 말과 글로 표현할 때 완결된 내용을 나타내는 최소의 단위  
※ 주어와 서술어를 갖추고 있는 것이 원칙이나 생략 가능
- ⑤ 쌍 둘을 하나로 묶어서 세는 단위로서 질의응답의 질문-답변과 같은 단위
- ⑥ 장 i) 도서나 장부 등의 낱장을 세는 단위로서, 낱장 한 장은 한 면으로 구성됨을 원칙으로 설정  
ii) 이미지 파일의 프레임을 세는 단위로서, 비디오 파일은 일반적으로 초당 24~60프레임(장)으로 구성  
※ 영화, 광고 : 24fps, 드라마 : 30fps, 게임, 스포츠, 3D : 60fps  
iii) 음성 녹음 시 스크립트 작성 단위로서, 스크립트 한 장은 A4용지 한 면으로 설정

### 3 비용 구성 요소

▶ AI 학습용 데이터 구축 전 과정에서 비용이 발생하는 구성 요소 항목은 3가지 예산 유형으로 구성

- 직접인건비 : 데이터 구축 작업 수행과 직접적인 관련이 있는 비용
- 제경비 : 직접비(직접인건비, 직접경비)에 포함되지 않고, 데이터 구축을 위한 기획, 경영, 총무 등에서 발생하는 간접 경비

※ 본 가이드의 분석 대상인 「AI 학습용 데이터 구축 사업」은 정보통신진흥기금 예산으로 운영('19~'22) 되고 있으며, 해당 사업 유형은 예산에 '제경비' 미반영

- 직접경비 : 데이터 구축 시 소요되는 직접적인 비용



| 그림 2-3 | 비용산정 예산 구성 요소

### | 참고 |

- **라벨링 데이터(Labeled Data)** 원천데이터에 부여한 '참값', 파일형식이나 해상도 등의 속성, 그리고 설명이나 주석 등이 포함된 '어노테이션(Annotation)'의 집합
- **원시데이터(Raw Data)** 기계학습을 목적으로 획득 단계에서 수집 또는 생성한 음성, 이미지, 영상, 텍스트 등의 데이터
- **원천데이터(Source Data, Unlabeled Data)** 원시데이터를 라벨링 공정에 투입하기 위해 필요한 전처리 등 정제 작업을 수행한 데이터로 라벨링데이터가 부여되지 않은 상태의 데이터
- **투입공수** 작업요소의 규모를 산정하기 위한 단위로 작업자가 한 달간 수행 가능한 작업량인 M/M(Man-Month) 방식으로 표기
- **반려율** 산출된 작업물이 단계별 품질 검사 기준을 충족하지 못하여 재작업이 요청되는 비율로, 데이터 구축 총작업량은 반려율을 가산하여 계산  
 ※ 반려율은 데이터 주제, 작업단계, 유형 등에 따라 편차가 있으나, 평균적으로 검사 대상 데이터 (총작업량)의 15~30% 수준이며, 본 가이드에서는 약 130%(산식 : 1.3)로 가산하여 계산(ρ24-)
- **보정요소** 작업요소별 비용 산식 도출 시, 작업의 특성 및 작업량을 정량적으로 표현한 것으로, 산식에 보정계수로 표기

AI 학습용 데이터 구축  
비용산정 가이드

# III

## AI 학습용 데이터 구축 비용산정

## 1 비용산정 절차

- ▶ AI 학습용 데이터 구축 비용은 기본적으로 투입공수에 의한 방식으로 산정
    - 투입공수에 의한 비용산정 방식은 M/M(Man-Month) 방식을 사용
    - 투입인력의 직접인건비는 한국소프트웨어산업협회에서 공표하는 “SW기술자 임금실태 조사”를 단가로 산정
- ※ 개별 단가 방식의 적용이 필요한 데이터는 세부 절차에서 설명함(상세한 설명은 III.2절 참조)  
 데이터 예시 : 클라우드워커 및 의사, 안무가 등 전문인력이 참여하는 데이터
- ▶ AI 학습용 데이터 구축 비용은 [표 3-1]와 같이 ① 직접인건비, ② 제경비 ③ 직접경비 3가지 요소로 구성되며, 각 요소별 설명은 II.3 절에서 기술

표 3-1 | AI 학습용 데이터 구축 비용의 구성 요소

구분	비용산정 핵심요소	비용 구성 요소
AI 학습용 데이터 구축 비용	투입공수	① 직접인건비 ② 제경비 ③ 직접경비

## 1) 비용산정 절차 및 절차별 주요 내용

- ▶ AI 학습용 데이터 구축 비용산정 절차는 데이터 구축 기획부터 수행할 작업 분석 및 관련 비용산정에 이르기까지 6단계로 구성
  - ① 작업 설계 ② 투입공수 산정 ③ 직접인건비 산정 ④ 제경비 산정 ⑤ 직접경비 산정 ⑥ 최종비용 산정

표 3-2 | 투입공수 방식에 의한 AI 학습용 데이터 구축 비용산정 절차

절차	주요내용	산출물
1. 작업 설계	<ul style="list-style-type: none"> <li>• AI 학습용 데이터 구축 대상 정의</li> <li>• 작업요소 정의</li> <li>• 작업요소별 물량 산정</li> </ul>	구축대상 작업요소 및 물량
↓		
2. 투입공수 산정	<ul style="list-style-type: none"> <li>• 작업요소별 투입공수 계산</li> </ul>	작업요소별 투입공수
↓		
3. 직접인건비 산정	<ul style="list-style-type: none"> <li>• 투입공수에 기초하여 작업요소별 직접인건비 산정</li> </ul>	직접인건비
↓		
4. 제경비 산정	<ul style="list-style-type: none"> <li>• 작업요소별 직접인건비에 기초하여 제경비 산정</li> </ul>	제경비
↓		
5. 직접경비 산정	<ul style="list-style-type: none"> <li>• 구축 업무와 관련된 직접경비 산정</li> </ul>	직접경비
↓		
6. 최종 비용 산정	<ul style="list-style-type: none"> <li>• AI 학습용 데이터 구축에 소요되는 모든 비용 합산</li> <li>- AI 학습용 데이터 구축 비용 = 직접인건비 + 제경비 + 직접경비</li> </ul>	AI 학습용 데이터 구축 비용

## 2) 절차별 설명

### 작업 설계

- ▶ AI 학습용 데이터 구축 대상 및 데이터 유형을 정의하고, 정의한 데이터 유형의 작업요소 그룹별 작업요소 정의
  - 어떤 데이터를 어떤 방식으로 구축할 것이며, 목표 지표 등을 포함하는 작업지시서를 작성하는 단계로 구축할 데이터에 대한 전반적인 이해를 바탕으로 설계 필요
    - ※ 구축할 데이터 특성의 이해도가 높은 고급인력의 참여가 필요하며, 본 비용산정 가이드는 실 구축 단계 (수집-정제-가공-검수)로 제한되어 있어, 필요시 작업 설계 단계에 대한 예산 검토 필요
  - [표 3-3]을 활용하여 구축할 데이터의 작업요소별 물량을 설정할 수 있으며, 구축 대상 및 물량에 따라 작업요소 변동 가능

표 3-3 | 작업요소별 작업량(물량) 정리 양식 (텍스트 유형의 예)

데이터 유형	작업요소그룹	작업요소		단위		
		구분	내용			
텍스트	수집 (원시데이터 수집)	구매	콘텐츠 구매		건	
			정제 (원천데이터 생성)	생성	질의응답	질의응답 생성
	유사질문 생성	건				
	가공 (라벨링데이터 생성)	가공		질의응답 검수		쌍
				대화 생성		건
				문서 생성		장
				대화 가공 (문장 생성)		건
	가공 (라벨링데이터 생성)	가공	문장 검수		문장	
			문서 가공		장	
			문서 검수		장	
			검수 (최종데이터 검수)		검수	장

### 작업요소별 투입공수 계산

AI 학습용 데이터 구축 비용은 데이터 유형 및 작업요소의 특성에 따라 달라지며, 각 작업그룹별로 정의되어 있는 작업요소와 각 작업요소에 대한 상세한 설명은 III.2절 참조

#### 작업요소별 비용 산정 요소 식별 및 투입공수 산정

- 작업요소별 투입공수(M/M)
  - = 작업요소별 물량 / 1일 작업량\* / 월평균 근무일수 x 반려율 x 보정요소\*\*\*
  - = 작업요소별 물량 / 1인 작업량\*\*\* x 반려율 x 보정요소
    - \* 1일 작업량 : 1일(8시간 기준)동안 1인이 수행할 수 있는 작업량
    - \*\* 1인 작업량 : 총작업량 중 1인이 수행할 수 있는 작업량
    - \*\*\* 보정요소 : 작업 특성에 해당하는 요소를 정량화하여 도출한 보정계수
    - ※ 현장에서 1일 총작업시간은 변동될 수 있으며, 산식에서는 1일 근무시간을 기준으로 제시
    - ※ 현재 적용되는 월평균 근무일수는 21이며, 정확한 환산기준은 한국소프트웨어산업협회에서 발표하는 "SW기술자 임금실태 조사" 최신 결과 참조
    - ※ 반려율은 총작업량의 15~30% 수준이며, 본 가이드에서는 30%로 적용하여 총작업량은 130% (산식:1.3) 가산
- 투입공수를 활용하지 않는 데이터는 '물량' 등의 비용산정 요소를 적용

#### 표 3-4]은 이미지 유형의 예시이며, 실제 활용시에는 작업요소를 수정하여 투입공수 계산

표 3-4 | 작업요소별 투입공수 산정 양식 (이미지 유형의 예)

작업요소		물량 (규모)	인원 (작업자)	기간 (소요시간)	장비 (활용기기)	장소 (작업장소)	투입공수
구매	① 콘텐츠 구매	건					
	촬영	② 이미지 촬영	장	인	일		
		③ 장소 임차		인	일		곳
④ 이미지 편집(합성)		장	인	일			
⑤ 정제		장	인	일			
⑥ 가공		장	인	일			
⑦ 검수(전수)		장	인	일			
합계							

❖ 직접인건비 산정

▶ 직접인건비는 작업요소별 투입공수에 기초하여 산정하며, 아래 산식 및 [표 3-5], [표 3-6] 참고

- 작업요소별 직접인건비 = 작업요소별 투입공수 × 작업요소별 단가
  - ※ 크라우드소싱 또는 전문인력(의사, 안무가 등)의 참여로 직접인건비에 개별 단가를 적용하여야 하는 경우, 투입공수 대신 물량 기준 방식으로 직접인건비 산정
- 인건비의 경우 한국소프트웨어산업협회에서 공표하는 “SW기술자 임금실태 조사”를 단가로 산정

**| 표 3-5 | 투입공수 기준 직접인건비 산정 양식**

순번	작업요소	투입공수	단가	직접인건비	비고
1					
2					
3					
직접인건비 합계					원

**| 표 3-6 | 물량 기준 직접인건비 산정 양식**

순번	작업요소	물량	단가	직접인건비	비고
4					
5					
6					
직접인건비 합계					원

❖ 제경비 산정

▶ 제경비는 작업요소별 직접인건비에 기초하여 산정하며 아래 산식 및 [표 3-7] 참고

- 작업요소별 제경비 = 작업요소별 직접인건비 × 제경비 비율
  - ※ 제경비는 직접인건비의 15~20% 수준이며, 본 가이드에서는 20%로 적용
  - ※ 본 가이드의 분석 대상인 「AI 학습용 데이터 구축 사업」은 정보통신진흥기금 예산으로 운영('19~'22) 되고 있으며, 해당 사업 유형은 예산에 '제경비' 미반영

**| 표 3-7 | 제경비 산정 양식**

순번	작업요소	직접인건비	제경비 비율	제경비	비고
1					
2					
3					
제경비 합계					원

❖ 직접경비 산정

▶ 직접경비는 II.3. 절 및 [표 3-8]을 활용하여 산정하며, 계상 시에는 정확한 내역 제시 필요

- 각 작업요소를 구성하는 비용구성 요소가 모두 포함된 산식 기재

**| 표 3-8 | 직접경비 산정 양식**

순번	작업요소	산식	직접경비	비고
1				
2				
3				
직접경비 합계				원

◆ 최종 비용 산정

- ▶ AI 학습용 데이터 구축 시 소요되는 모든 비용을 합산하여 최종 비용을 도출하며, 아래 산식 및 [표 3-9] 참고
  - AI 학습용 데이터 구축 비용 = 직접인건비 + 제경비 + 직접경비

| 표 3-9 | AI 학습용 데이터 구축 최종 비용 산정 양식

구분	금액	비고
직접인건비	원	
제경비	원	
직접경비	원	
소계 (부가세 별도)		원

## 2 데이터 4개 유형별 비용산정 절차 및 상세 내용

- ▶ 본 절은 AI 학습용 데이터 유형별 구축 절차와 관련 작업요소의 비용산정을 위한 내용을 상세히 설명
  - 데이터 유형별 비용산정 절차를 소개하기에 앞서, 각 데이터 구축 작업별 작업요소와 산식 산정시 기준이 된 작업량을 제시

| 참고사항 |

- 작업 단계(작업요소 그룹)별 소요 비용은 구축 데이터 주제에 따라 변동 가능
  - 공통적으로 '원천데이터'를 생성하는 정제 과정에 상대적으로 높은 비용이 투입되며, '원천데이터'의 완성도에 따라 가공 과정의 시간, 비용 변동 가능
- 각 데이터 유형별로 제시된 작업요소는 구축 작업 시 반복적으로 등장하는 작업들의 모음으로, 구축 대상 데이터의 특성을 고려하여 일부 작업요소만 활용하여 작업 진행
- 작업요소별 작업량은 분석 대상 과제의 데이터를 기반으로 도출된 통계에 의한 평균값임(보정요소로 연계)

작업요소 산식 표

데이터 유형	작업요소 그룹	작업요소	비용산정 산식
텍스트	수집 (원시데이터 수집)	① 콘텐츠 구매	물량 x 단가(콘텐츠 구매)
	정제 (원천데이터 생성)	② 질의응답 생성	물량(구축량) / 작업량(일 구축량) / 기간(월 근무일) x 반려율 x 보정요소 x 단가(월 인건비)
		③ 유사질문 생성	물량(구축량) x 물량(유사질문 개수) x 단가(질의생성)
		④ 질의응답 검수	물량(구축량) / 작업량(일 검수량) / 기간(월 근무일) x 반려율 x 보정요소 x 단가(월 인건비)
		⑤ 대화 생성	물량(구축량) x 인원 x 보정요소 x 단가(개인정보 처리)
		⑥ 문서 생성	물량(구축량) x 단가(문서 종류별 비용)
		가공 (라벨링데이터 생성)	⑦ 대화 가공 (문장 생성)
	⑧ 문장 검수		물량(구축량) / 작업량(일 검수량) / 기간(월 근무일) x 반려율 x 보정요소 x 단가(월 인건비)
	⑨ 문서 가공		물량(구축량) / 작업량(일 가공량) / 기간(월 근무일) x 반려율 x 단가(월 인건비)
	⑩ 문서 검수		물량(구축량) / 작업량(일 검수량) / 기간(월 근무일) x 반려율 x 단가(월 인건비)
	검수 (최종데이터 검수)	⑪ 검수(전수)	물량(총 텍스트 구축량) / 작업량(일 검수량) / 기간(월 근무일) x 반려율 x 단가(월 인건비)

데이터 유형	작업요소 그룹	작업요소	비용산정 산식
이미지	수집 (원시데이터 수집)	① 콘텐츠 구매	물량 x 단가(콘텐츠 구매)
		② 이미지 촬영	인원 x 기간 x 반려율 x 단가(이미지 촬영 비용)
		③ 장소 임차	① 장소 x 기간 x 단가(임차 비용) ② 인원(총 촬영인원) / 작업량(일 촬영인원) / 기간(월 근무일) x 단가(월 인건비)
		④ 이미지 편집(합성)	물량 x 단가(편집 비용)
	정제 (원천데이터 생성)	⑤ 정제	물량(총 이미지 구축량) / 작업량(일 정제량) / 기간(월 근무일) x 보정요소 x 반려율 x 단가(월 인건비)
	가공 (라벨링데이터 생성)	⑥ 가공	물량(총 이미지 구축량) / 작업량(일 가공량) / 기간(월 근무일) x 보정요소 x 반려율 x 단가(월 인건비)
	검수 (최종데이터 검수)	⑦ 검수(전수)	물량(총 이미지 구축량) / 작업량(일 검수량) / 기간(월 근무일) x 반려율 x 단가(월 인건비)
	음성	수집 (원시데이터 수집)	① 콘텐츠 구매
② 스크립트 작성			물량(총 스크립트분량) / 작업량(1인 스크립트작성분량) x 반려율 x 보정요소 x 단가(인당 인건비)
③ 스크립트 검수			물량(총 스크립트분량) / 작업량(1인 스크립트검수분량) / 월평균 근무일 x 반려율 x 보정요소 x 단가(월 인건비)
④ 일반 녹음			물량(총 녹음분량) / 작업량(인당 녹음분량) x 반려율 x 보정요소 x 단가(인당 인건비)
⑤ 스크립트 녹음			물량(총 녹음분량) x 단가(장당 녹음비용)
⑥ 특수 녹음			물량(총 녹음분량) x 작업량(시간당 녹음건수) x 단가(건당 비용)
⑦ 장소 임차			장소 x 단가(임차 비용)
⑧ 부대 업무			( ※ 변동 )
정제 (원천데이터 생성)		⑨ 녹음 검수	물량(총 녹음분량) / 월평균 근무일 / 1일 업무 시간 x 반려율 x 보정요소 x 단가(월 인건비)
가공 (라벨링데이터 생성)		⑩ 전사(라벨링)	(물량(총 녹음분량) x 보정계수) / 월평균 근무일 / 1일 업무 시간 x 반려율 x 단가(월 인건비)
검수 (최종데이터 검수)		⑪ 검수(전수)	물량(총 음성 구축량) / 작업량(일 검수량) / 기간(월 근무일) x 반려율 x 단가(월 인건비)

데이터 유형	작업요소 그룹	작업요소	비용산정 산식
비디오 (영상)	수집 (원시데이터 수집)	① 콘텐츠 구매	물량 x 단가(콘텐츠 구매)
		② 영상 촬영	인원 x 기간 x 반려율 x 단가(영상 촬영 비용)
		③ 장비 임차	물량 x 인원 x 기간 x 단가(영상 촬영 비용)
		④ 장소 임차	① 장소 x 기간 x 단가(장소 임차 비용) ② 인원(총 촬영인원) / 작업량(일 촬영인원) / 기간(월 근무일) x 단가(월 인건비)
		⑤ 부대 업무	물량 x 단가(촬영 부대 비용)
	정제 (원천데이터 생성)	⑥ 정제(영상)	물량(총 영상 구축량) / 작업량(일 정제량) / 기간(월 근무일) x 반려율 x 보정요소 x 단가(월 인건비)
	가공 (라벨링데이터 생성)	⑦ 가공(영상)	물량(총 영상 구축량) / 작업량(일 가공량) / 기간(월 근무일) x 반려율 x 보정요소 x 단가(월 인건비)
		⑧ 가공(이미지)	물량(총 영상 구축량) / 작업량(일 가공량) / 기간(월 근무일) x 반려율 x 보정요소 x 단가(월 인건비)
	검수 (최종데이터 검수)	⑨ 검수(전수)	물량(총 이미지 구축량) / 작업량(일 검수량) / 기간(월 근무일) x 반려율 x 단가(월 인건비)

작업량 전체 표

데이터 유형	작업요소	작업량	보정요소	보정계수	
텍스트	① 콘텐츠 구매	-	-	-	
	② 질의응답 생성	60쌍/일	원시데이터 유형	문서	0.63
			영상	1.00	
	③ 유사질문 생성	※ 질의응답 생성 작업량의 2~4배	-	-	-
	④ 질의응답 검수	100쌍/일	원시데이터 유형	문서	1.00
				영상	0.33
	⑤ 대화 생성	-	대화 유형	일상 대화	1.00
				목적지향 대화	1.67
	⑥ 문서 생성	-	-	-	-
	⑦ 대화 가공 (문장 생성)	2.5건/일	대화 유형	일상 대화	0.50
				목적지향 대화	1.00
⑧ 문장 검수	833문장/일	대화 유형	일상 대화	0.63	
			목적지향 대화	1.00	
⑨ 문서 가공	40장/일	-	-	-	
⑩ 문서 검수	133장/일	-	-	-	
⑪ 검수(전수)	-	-	-	-	

데이터 유형	작업요소	작업량	보정요소		보정계수
이미지	① 콘텐츠 구매	-	-	-	-
	② 이미지 촬영	모델 20~100장/인/일 촬영자 1~3인/일	-	-	-
	③ 장소 임차	-	-	-	-
	④ 이미지 편집(합성)	4장/일	-	-	-
	⑤ 정제	-	-	-	-
	⑥ 가공	50장/일	작업난이도	고	1.0
				저	0.5
⑦ 검수(전수)	-	-	-	-	
음성	① 콘텐츠 구매	-	-	-	-
	② 스크립트 작성	200문장/인	작업유형	한국어	1.00
				한영혼합	4.00
	③ 스크립트 검수	1,500문장/시간	작업 난이도	고	1.00
				저	0.67
	④ 일반 녹음	3.25시간/인	작업유형	스튜디오 녹음 및 영상촬영	0.25
				스튜디오 녹음	0.50
				클라우드소싱 (성인)	1.00
				클라우드소싱 (아동·노인)	1.625
	⑤ 스크립트 녹음	-	-	-	-
	⑥ 특수 녹음	2건/시간	-	-	-
⑦ 장소 임차	34개/월	-	-	-	
⑧ 부대 업무	-	-	-	-	
⑨ 녹음 검수	-	검수업무 가중치	저	1	
			중	2	
			고	3	
⑩ 가공 - 전사 (라벨링)	-	가공업무 가중치	※ 음성 유형(p.62) 참조		
⑪ 검수(전수)	-	-	-	-	

데이터 유형	작업요소	작업량	보정요소		보정계수
비디오 (영상)	① 콘텐츠 구매	-	-	-	-
	② 영상 촬영	촬영자 2~3시간/일 촬영자 1~3인/일	-	-	-
	③ 장비 임차	-	-	-	-
	④ 장소 임차	-	-	-	-
	⑤ 부대 업무	-	-	-	-
	⑥ 정제(영상)	-	-	-	-
	⑦ 가공(영상)	3개/일	작업난이도	고	1.0
				저	0.06
	⑧ 가공(이미지)	20장/일	작업난이도	고	1.0
저				0.2	
⑨ 검수(전수)	-	-	-	-	

### 1) 데이터 유형 : 텍스트

▶ 텍스트 유형 AI 학습용 데이터 구축 특징

- 원시데이터가 텍스트 유형인 AI 학습용 데이터는 '문서요약', '질의응답', '기계번역', '대화' 등의 다양한 문제 해결을 목적으로 구축
- 텍스트 유형의 데이터는 이미 존재하는 콘텐츠를 구매하여 정제하거나, 문장 등을 새롭게 생성하여 원천데이터를 생성
- 확보된 원천데이터는 가공(라벨링) 및 검수 과정을 통해 1쌍의 AI 학습용 데이터를 도출하며, 텍스트 유형은 대개 생성·가공·검수 작업이 동시 진행

[표 3-10]에 제시된 텍스트 유형 작업요소는 구축 작업 시 반복적으로 등장하는 작업들의 모음으로, 구축 대상 데이터의 특성에 따라, 일부 작업요소만 포함  
 (예) 민원 Q/A 텍스트 데이터의 작업요소 : ①, ②, ③, ④  
 문서 요약 텍스트 데이터의 작업요소 : ⑥, ⑨, ⑩

표 3-10 | 텍스트 유형 작업요소 및 작업요소별 비용산정 요소

작업그룹	작업요소	물량 (규모)	인원 (작업자)	기간 (소요시간)	장비 (활용기기)	장소 (작업장소)	비용 구분
수집	① 콘텐츠 구매	건					직접경비
정제	질의응답	② 질의응답 생성	쌍	인	일		직접인건비
		③ 유사질문 생성	건	인	일		직접인건비
		④ 질의응답 검수	쌍	인	일		직접인건비
	⑤ 대화 생성	건	인	일		직접인건비	
	⑥ 문서 생성	장	인	일		직접인건비	
가공	대화	⑦ 대화 가공 (문장 생성)	건	인	일		직접인건비
		⑧ 문장 검수	문장	인	일		직접인건비
	문서	⑨ 문서 가공	장	인	일		직접인건비
		⑩ 문서 검수	장	인	일		직접인건비
검수	⑪ 검수(전수)	장	인	일		직접인건비	

텍스트 - ① 콘텐츠 구매				
비용 구분	직접경비	작업량	-	
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)
	■	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기				

▶ **설명**  
 본 작업요소는 원시데이터 생성에 활용이 가능하나, 직접적인 생성이 어려운 문서나 영상 등의 콘텐츠\*를 구매하는 행위로 다음의 작업들을 포함하여 구성  
 \* 콘텐츠 : 행정문서, 뉴스 원문 데이터, 뉴스 원문 저작권, 방송 콘텐츠, 영화 등

- 콘텐츠의 기본 언어는 한국어로 설정
- 콘텐츠가 기획된 자료인지 확인 필요
- 구매하는 콘텐츠 간 중복 여부 확인 필요

▶ **산식 = 물량 x 단가**  
 위 산식을 활용하여 본 작업요소의 비용을 산정

- 물량(건) : 콘텐츠 총구매량
- 단가(원) : 콘텐츠 건당 구매 비용

▶ **기타 유의사항**

- 콘텐츠를 보유한 기관(기업) 내규에 따른 콘텐츠 구매단가의 편차 존재
- 콘텐츠 구매는 저작권료를 의미할 수 있으며, 각 콘텐츠의 특성에 따라 저작권의 유효기간 여부, 재가공 가능 여부 등 파악 필요
- 여러 유형의 콘텐츠를 구매하는 경우 각각의 콘텐츠 모두 직접경비 산정 양식에 계상 필요

텍스트 - ② 질의응답 생성																
비용 구분	직접인건비	작업량	60쌍/일													
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)											
	■	□	□	□	□											
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																
<p>▶ <b>설명</b></p> <p>본 작업요소는 질의응답을 생성하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 특정 종류의 정보에 관한 질문과 그에 대한 해답을 질의응답 1쌍으로 설정</li> <li>- 서술형 답변, 예/아니오 형식의 단문형 답변 등 답변의 형태를 고려한 질의응답 설계 필요</li> <li>- 답변이 특정 값에 편향되지 않는 질의응답 설계 필요</li> <li>- 질의응답의 기본 언어는 한국어로 설정</li> <li>- 생성하는 질의응답 간 중복 여부 확인 필요.</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 보정요소 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(쌍) : 질의응답 총생성량</li> <li>• 작업량(쌍) : 1일 질의응답 구축량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(질의응답 생성)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율 (p.14 참조, 이하 생략)</li> <li>- 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용                     <table border="1" data-bbox="257 1050 952 1168"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">원시데이터 유형</td> <td>문서</td> <td>0.63</td> <td>pdf, hwp, docx, html 등 원문에서 텍스트 및 JSON을 추출 가능한 유형</td> </tr> <tr> <td>영상</td> <td>1.00</td> <td>mp4 등 동영상에서 한국어 문장을 추출 가능한 유형</td> </tr> </tbody> </table> </li> </ul> <p>- 단가(원) : 참여 인력 월 인건비</p> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>						보정요소		보정계수	보정요소 설명	원시데이터 유형	문서	0.63	pdf, hwp, docx, html 등 원문에서 텍스트 및 JSON을 추출 가능한 유형	영상	1.00	mp4 등 동영상에서 한국어 문장을 추출 가능한 유형
보정요소		보정계수	보정요소 설명													
원시데이터 유형	문서	0.63	pdf, hwp, docx, html 등 원문에서 텍스트 및 JSON을 추출 가능한 유형													
	영상	1.00	mp4 등 동영상에서 한국어 문장을 추출 가능한 유형													

텍스트 - ③ 유사질문 생성					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 유사질문*을 생성하기 위한 다음의 작업을 포함하여 구성</p> <p>* 유사질문 : 질의응답의 질문과 내용이 같으면서 문장 구조가 다른 질문</p> <ul style="list-style-type: none"> <li>- 유사질문의 기본 언어는 한국어로 설정</li> <li>- 생성하는 유사질문 간 중복 여부 확인 필요</li> </ul> <p>▶ <b>산식 = 물량 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 물량(건) : 유사질문 총생성량</li> <li>- 단가(원) : 유사질문 건당 생성 비용                     <ul style="list-style-type: none"> <li>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</li> </ul> </li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>					

텍스트 - ④ 질의응답 검수																
비용 구분	직접인건비	작업량	100쌍/일													
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)											
	■	□	□	□	□											
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																
<p>▶ <b>설명</b></p> <p>본 작업요소는 생성한 질의응답을 검수하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 생성한 질의응답에서 개인정보 비식별화, 민감성*, 오타자 등을 검수하고 오류를 수정</li> <li>* 민감성 : 정치적 견해, 특정인물 비하, 성적인 표현 등</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 보정요소 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(쌍) : 질의응답 총생성량</li> <li>• 작업량(쌍) : 1일 질의응답 검수량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(질의응답 검수본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용                     <ul style="list-style-type: none"> <li>※ '④ 질의응답 검수' 작업 시에는 '영상' 내 정제된 문장을 바탕으로 진행되어 '② 질의응답 생성' 작업에 비해 상대적으로 낮은 보정계수 적용</li> </ul> </li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">원시데이터 유형</td> <td>문서</td> <td>1.00</td> <td>pdf, hwp, docx, html 등 원문에서 텍스트 및 JSON을 추출 가능한 유형</td> </tr> <tr> <td>영상</td> <td>0.33</td> <td>mp4 등 동영상에서 한국어 문장을 추출 가능한 유형</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>- 단가(원) : 참여 인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>						보정요소		보정계수	보정요소 설명	원시데이터 유형	문서	1.00	pdf, hwp, docx, html 등 원문에서 텍스트 및 JSON을 추출 가능한 유형	영상	0.33	mp4 등 동영상에서 한국어 문장을 추출 가능한 유형
보정요소		보정계수	보정요소 설명													
원시데이터 유형	문서	1.00	pdf, hwp, docx, html 등 원문에서 텍스트 및 JSON을 추출 가능한 유형													
	영상	0.33	mp4 등 동영상에서 한국어 문장을 추출 가능한 유형													

텍스트 - ⑤ 대화 생성																
비용 구분	직접인건비	작업량	-													
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)											
	■	■	□	□	□											
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																
<p>▶ <b>설명</b></p> <p>본 작업요소는 대화를 생성하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 대화는 여러 건의 대화 턴으로 구성되며, 대화의 기본 언어는 한국어로 설정                     <ul style="list-style-type: none"> <li>※ 대화 턴(turn) : 하나의 발화와 이에 대한 하나의 응답을 묶어 하나의 대화 턴으로 정의</li> </ul> </li> <li>- 대화는 2인 이상의 화자, 2개 이상의 대화 턴, 문장은 3개 이상의 어절로 구성</li> <li>- 생성하는 대화 간 중복 여부 확인 필요</li> </ul> <p>▶ <b>산식 = 투입공수 x 보정요소 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 × 인원                     <ul style="list-style-type: none"> <li>• 물량(건) : 대화 총생성량</li> <li>• 인원(인) : 대화 1건당 참여하는 평균 화자 인원</li> </ul> </li> <li>- 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">대화 유형</td> <td>일상 대화</td> <td>1.00</td> <td>특정 주제에 대해 자유롭게 나눈 대화</td> </tr> <tr> <td>목적지향 대화</td> <td>1.67</td> <td>식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>- 단가(원) : 화자 개인정보 활용 등의 비용                     <ul style="list-style-type: none"> <li>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</li> </ul> </li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>						보정요소		보정계수	보정요소 설명	대화 유형	일상 대화	1.00	특정 주제에 대해 자유롭게 나눈 대화	목적지향 대화	1.67	식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화
보정요소		보정계수	보정요소 설명													
대화 유형	일상 대화	1.00	특정 주제에 대해 자유롭게 나눈 대화													
	목적지향 대화	1.67	식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화													

텍스트 - ⑥ 문서 생성					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 문서*를 생성하기 위한 다음의 작업을 포함하여 구성</p> <p>* 문서 : 뉴스, 보도자료, 역사기록물/문화재소개, 보고서, 자기소개서, 메일, 회의록, 사설, 간행물, 연설문, 일기/소설, 나레이션 등</p> <ul style="list-style-type: none"> <li>- 문서의 기본 언어는 한국어로 설정</li> <li>- 생성하는 문서 간 중복 여부 확인 필요</li> </ul> <p>▶ <b>산식 = 물량 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 물량(장) : 문서 총생성량</li> <li>- 단가(원) : 문서 건당 생성 비용</li> </ul> <p>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</p> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 여러 유형의 문서를 생성하는 경우 각각의 문서 모두 투입공수 산정 양식 및 물량 기준 직접인건비 산정 양식에 계상 필요</li> </ul>					

텍스트 - ⑦ 대화 가공(문장 생성)																
비용 구분	직접인건비	작업량	2.5건/일													
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)											
	■	□	□	□	□											
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																
<p>▶ <b>설명</b></p> <p>본 작업요소는 생성한 대화를 가공하여 문장을 생성하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 생성한 대화에 오류가 없음(대화 생성 작업의 반려율 저조)을 전제하여 대화를 가공하여 문장 생성</li> <li>- 대화에서 의미가 모호하거나 대화 진행에 불필요한 음절 및 어절의 제거</li> <li>- 대화에 포함된 은어, 비속어 등의 제거</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 보정요소 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(건) : 대화 총생성량</li> <li>• 작업량(건) : 1일 대화 가공량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(대화 가공본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> <table border="1" data-bbox="1339 1093 2033 1209"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">대화 유형</td> <td>일상 대화</td> <td>0.50</td> <td>특정 주제에 대해 자유롭게 나눈 대화</td> </tr> <tr> <td>목적지향 대화</td> <td>1.00</td> <td>식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>- 단가(원) : 참여 인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>						보정요소		보정계수	보정요소 설명	대화 유형	일상 대화	0.50	특정 주제에 대해 자유롭게 나눈 대화	목적지향 대화	1.00	식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화
보정요소		보정계수	보정요소 설명													
대화 유형	일상 대화	0.50	특정 주제에 대해 자유롭게 나눈 대화													
	목적지향 대화	1.00	식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화													

텍스트 - ㉔ 문장 검수																
비용 구분	직접인건비	작업량	833문장/일													
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)											
	■	□	□	□	□											
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																
<p>▶ <b>설명</b></p> <p>본 작업요소는 대화 가공을 통해 생성된 문장을 검수하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 생성한 문장의 개인정보 비식별화, 민감성*, 오타자 등을 검수하고 오류를 수정</li> <li>* 민감성 : 정치적 견해, 특정인물 비하, 성적인 표현 등</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 보정요소 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(문장) : 문장 총생성량</li> <li>• 작업량(문장) : 1일 문장 검수량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(문장 검수본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">대화 유형</td> <td>일상 대화</td> <td>0.63</td> <td>특정 주제에 대해 자유롭게 나눈 대화</td> </tr> <tr> <td>목적지향 대화</td> <td>1.00</td> <td>식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>- 단가(원) : 참여 인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>						보정요소		보정계수	보정요소 설명	대화 유형	일상 대화	0.63	특정 주제에 대해 자유롭게 나눈 대화	목적지향 대화	1.00	식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화
보정요소		보정계수	보정요소 설명													
대화 유형	일상 대화	0.63	특정 주제에 대해 자유롭게 나눈 대화													
	목적지향 대화	1.00	식당 예약, 결제 취소 및 환불, 고객센터 상담 등 특정한 목적을 달성하기 위한 대화													

텍스트 - ㉕ 문서 가공					
비용 구분	직접인건비	작업량	40장/일		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 문서를 가공하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 생성한 문서에 오류가 없음(문서 생성 작업의 반려율 저조)을 전제하여 문서 가공</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(장) : 문서 총생성량</li> <li>• 작업량(장) : 1일 문서 가공량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(문서 가공본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 참여 인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>					

텍스트 - ⑩ 문서 검수					
비용 구분	직접인건비	작업량	133장/일		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 가공된 문서를 검수하기 위한 다음의 작업을 포함</p> <ul style="list-style-type: none"> <li>- 가공된 문서에서 개인정보 비식별화, 민감성*, 오타자 등을 검수하고 오류를 수정 * 민감성 : 정치적 견해, 특정인물 비하, 성적인 표현 등</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(장) : 문서 총생성량</li> <li>• 작업량(장) : 1일 문서 검수량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 문서 검수 중 작업에 오류가 발생하여 재작업을 수행해야 하는 작업의 비율</li> <li>- 단가(원) : 참여 인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>					

텍스트 - ⑪ 검수(전수)					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 최종 산출물을 생성하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 최종 데이터 1쌍(원천, 라벨링 데이터)을 생성하기 위해 앞 단계에서 산출된 데이터의 품질을 전수 검사하는 작업</li> <li>- 본 단계 외에도, 데이터 구축 작업단계(작업그룹)마다 교차 품질 검사는 병행되며, 텍스트 데이터의 반려율 비율이 높은 작업은 작업그룹(④, ⑧, ⑩)으로 명시</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(장) : 텍스트 원시, 원천데이터 총생성량</li> <li>• 작업량(장) : 1일 원시, 원천데이터 검사 수량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(최종파일)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가 : 참여인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당사항 없음</li> </ul>					

## 2) 데이터 유형 : 이미지

### ▶ 이미지 유형 AI 학습용 데이터 구축 특징

- 원시데이터가 이미지 유형인 AI 학습용 데이터는 '이미지 분류', '객체 인식', '영역 구분' 등의 다양한 문제 해결을 위한 목적으로 구축
- 이미지 유형의 데이터는 이미 존재하는 콘텐츠를 구매하여 정제하거나, 직접 촬영, 이미지 편집본 등을 정제하여 이미지 원천데이터 생성
- 확보된 이미지 원천데이터는 가공(라벨링) 과정을 통해 라벨링 데이터를 생성하고, 데이터 검수를 통해 1쌍의 AI 학습용 데이터를 도출

[표 3-11]에 제시된 이미지 유형 작업요소는 구축 작업 시 반복적으로 등장하는 작업들의 모음으로, 구축 대상 데이터의 특성에 따라, 일부 작업요소만 포함  
 (예) 건물 사진 데이터의 작업요소 : ②, ⑤, ⑥, ⑦  
 안면 이미지 데이터의 작업요소 : ①, ④, ⑤, ⑥, ⑦

표 3-11 | 이미지 유형 작업요소 및 작업요소별 비용산정 요소

작업그룹	작업요소		물량 (규모)	인원 (작업자)	기간 (소요시간)	장비 (활용기기)	장소 (작업장소)	비용 구분
수집	구매	① 콘텐츠 구매	건					직접경비
	촬영	② 이미지 촬영	장	인	일			직접인건비
		③ 장소 임차		인	일		곳	직접경비
		④ 이미지 편집(합성)	장	인	일			직접인건비
정제	⑤ 정제	장	인	일			직접인건비	
가공	⑥ 가공	장	인	일			직접인건비	
검수	⑦ 검수(전수)	장	인	일			직접인건비	

이미지 - ① 콘텐츠 구매					
비용 구분	직접경비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□

※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기

**▶ 설명**  
 본 작업요소는 원시데이터 생성에 활용이 가능하나, 직접적인 구축이 어려운 문서나 이미지 등의 콘텐츠\*를 구매하는 행위로 다음의 작업들을 포함하여 구성  
 \* 콘텐츠 : 행정문서, 뉴스 원문 데이터, 뉴스 원문 저작권, 방송 콘텐츠, 영화 등

- 콘텐츠의 기본 언어는 한국어로 설정
- 콘텐츠가 확보된 자료인지 확인 필요
- 구매하는 콘텐츠 간 중복 여부 확인 필요
- 특정 상황에 대한 대규모 이미지 수집이 필요할 시, 클라우드소싱 방식으로 크라우드 워커들이 촬영한 이미지에 건별 비용을 책정하여 데이터를 구매하는 작업을 의미

**▶ 산식 = 물량 x 단가**  
 위 산식을 활용하여 본 작업요소의 비용을 산정

- 물량(개) : 이미지 콘텐츠 총구매량
- 단가(원) : 이미지 콘텐츠 개당 구매 비용

**▶ 기타 유의사항**

- 이미지 콘텐츠를 보유한 (공공)기관 기업의 내규에 따른 콘텐츠 구매단가의 편차 존재  
 ※ 기관을 통한 구매 방식 외, 클라우드소싱 방식으로 일반인의 이미지 파일을 수집(구매)하는 경우, 대상의 개인정보 민감도, 촬영 기법 등에 따라 개당 구매 비용의 편차가 매우 큼(예 : 약 1,000원 ~ 120,000원)
- 콘텐츠 구매는 저작권료를 의미할 수 있으며, 각 콘텐츠의 특성에 따라 저작권의 유효기간 여부, 재가공 가능 여부 등 파악 필요
- 여러 유형의 콘텐츠를 구매하는 경우 각각의 콘텐츠 모두 직접경비 산정 양식에 계상 필요
- 크라우드워커 참여시 구매 대상 콘텐츠에 대한 표준 규격 제시 필요

이미지 - ② 이미지 촬영					
비용 구분	직접인건비	작업량	모델1인 20~100장/일 촬영자 1~3인/일		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		■	■	■	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 이미지 데이터를 획득(촬영)하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 기존에 존재하지 않는 이미지 원시데이터를 획득하기 위하여 카메라 등의 촬영 장비를 활용하여 데이터를 생성하는 작업</li> <li>- 이미지 촬영 작업의 참여인력은 ① 촬영을 직접하는 촬영자(및 촬영 보조인력), ② 촬영 대상이 되는 모델 인력으로 구분</li> <li>- 구축 대상 이미지 데이터의 작업(촬영) 난이도에 따라, 참여인력 ①, ②는 특정 분야 전문인력의 참여 필요</li> </ul> <p>▶ <b>산식 = 인원 x 기간 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 인원(인) : 1일 이미지 촬영 참여인력</li> <li>- 기간(일) : 이미지 촬영 총소요기간</li> <li>- 반려율 : 작업물(촬영본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 1일 참여인력의 이미지 촬영 비용</li> </ul> <p>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</p> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 총구축량 중 클라우드워커의 참여 비중에 대한 사전 검토가 필요하며, 워커 참여시 통일된 이미지 촬영 규격에 대한 충분한 교육을 통해 반려율을 낮추는 노력 필요</li> <li>※ 반려된 이미지를 재촬영하는데 소요되는 시간(비용)을 낮추기 위한 교육 필수</li> <li>- 촬영 장비는 2D 이미지의 경우, 일반적으로 100만 화소 이상 이미지 촬영이 가능한 카메라 수준을 요구하며, 3D 이상의 차원이 요구되는 경우 특수 장비 필요</li> </ul>					

이미지 - ③ 장소 임차					
비용 구분	직접경비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		□	■	■	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 이미지 콘텐츠를 생성하기 위하여 물리적 공간을 임차하는 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 비용이 수반되는 대부분의 장소 임차는 실내 공간인 스튜디오 임차를 의미</li> </ul> <p>▶ <b>산식① = 장소 x 기간 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 장소(곳) : 촬영이 진행되는 장소의 총수</li> <li>- 기간(일) : 촬영장소 임차 총소요기간</li> <li>- 단가(원) : 1일 촬영장소 임차 비용</li> </ul> <p>※ 장기 대여의 경우, 투입공수를 적용하여 총 임차비용을 산출함</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p><b>산식② = 스튜디오 대여 투입공수 x 단가</b></p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 인원 / 작업량 / 월평균 근무일</li> <li>• 인원(인) : 총촬영 인원</li> <li>• 작업량(인) : 1일 촬영 참여 인원</li> <li>• 월평균 근무일(일) : 21일</li> <li>- 단가(원) : 참여인력 월 인건비</li> </ul> </div> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 생성하는 콘텐츠의 특성에 따라, 실외 공간에 대한 임차료가 필요한 경우도 존재</li> <li>예 : 여러 실외 장소를 배경으로 하는 이미지 콘텐츠 생성 등</li> </ul>					

이미지 - ④ 이미지 편집(합성)					
비용 구분	직접경비	작업량	4장/일		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		■	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 이미지 콘텐츠를 생성하기 위하여 이미지 편집 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 이미지 콘텐츠를 컴퓨터 이미지 편집 프로그램을 이용하여 편집(합성) 하는 작업</li> </ul> <p>▶ <b>산식 = 물량 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 물량(장) : 이미지 파일 편집 총수량</li> <li>- 단가(원) : 이미지 파일 1장당 편집 비용                             <ul style="list-style-type: none"> <li>※ 이미지 편집 작업은 전문성이 요구되며, 전문성에 따른 차등 편집 단가 책정 필요</li> <li>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</li> </ul> </li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 편집된 파일(원천데이터)의 검수 과정에서 수정사항 반영을 위해 편집 프로그램의 통일 필요</li> </ul>					

이미지 - ⑤ 정제					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		■	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 이미지 원시데이터 파일을 정제하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 이미지 원시데이터의 파일 규격 통일, 비식별 처리 등 전처리(정제)하여 원천 데이터를 생성하는 작업</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                             <ul style="list-style-type: none"> <li>• 물량(장) : 이미지 데이터 총생성량</li> <li>• 작업량(장) : 1일 이미지 데이터 정제 수량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(정제본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가 : 참여인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>					

이미지 - ⑥ 가공															
비용 구분	직접인건비	작업량	50장/일												
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)										
	■	□	□	□	□										
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기															
<p>▶ <b>설명</b></p> <p>본 작업요소는 이미지 원천데이터를 가공하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 이미지 원천데이터에 라벨링 정보를 부착하여 라벨링 데이터를 생성하는 작업을 포함</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일 x 보정요소                     <ul style="list-style-type: none"> <li>• 물량(장) : 이미지 데이터 총생성량</li> <li>• 작업량(장) : 1일 이미지 데이터 가공 수량</li> <li>• 월평균 근무일(일) : 21일</li> <li>• 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> </li> </ul> <table border="1"> <thead> <tr> <th>보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">작업 난이도</td> <td>고</td> <td>1.00</td> <td>다차원(3D 이상) 이미지 데이터를 가공 작업자가 처리할 수 있는 물량</td> </tr> <tr> <td>저</td> <td>0.50</td> <td>일차원 이미지 데이터를 가공 작업자가 처리할 수 있는 물량</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>- 반려율 : 작업물(가공파일)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가 : 참여인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 사항 없음</li> </ul>						보정요소	보정계수	보정요소 설명	작업 난이도	고	1.00	다차원(3D 이상) 이미지 데이터를 가공 작업자가 처리할 수 있는 물량	저	0.50	일차원 이미지 데이터를 가공 작업자가 처리할 수 있는 물량
보정요소	보정계수	보정요소 설명													
작업 난이도	고	1.00	다차원(3D 이상) 이미지 데이터를 가공 작업자가 처리할 수 있는 물량												
	저	0.50	일차원 이미지 데이터를 가공 작업자가 처리할 수 있는 물량												

이미지 - ⑦ 검수(전수)					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 최종 산출물을 생성하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 최종 데이터 1쌍(원천, 라벨링 데이터)을 생성하기 위해 앞 단계에서 산출된 데이터의 품질을 전수 검사하는 작업</li> <li>- 본 단계 외에도, 데이터 구축 작업단계(작업그룹)마다 교차 품질 검사는 병행</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일                     <ul style="list-style-type: none"> <li>• 물량(장) : 이미지 원시, 원천데이터 총생성량</li> <li>• 작업량(장) : 1일 원시, 원천데이터 검사 수량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(최종파일)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가 : 참여인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당사항 없음</li> </ul>					

### 3) 데이터 유형 : 음성

#### ▶ 음성 유형 AI 학습용 데이터 구축 특징

- 음성 AI 학습용 데이터는 사람의 목소리나 말소리, 음악, 효과음, 자연음 등이 기록(녹음)된 형태를 의미
- 원시데이터가 음성 유형인 AI 학습용 데이터는 '음성 인식', '음성합성' 등 음성 언어 처리와 관련된 문제 해결을 목적으로 구축
- 음성 유형의 데이터는 이미 존재하는 콘텐츠를 구매하여 정제하거나, 스크립트, 녹음본 등의 검수 공정을 통해 원천데이터를 생성
- 확보된 원천데이터는 메타데이터 입력, 전사 입력·검수 등의 가공 공정을 통해 1쌍의 AI 학습용 데이터를 도출

[표 3-12]에 제시된 음성 유형 작업요소는 구축 작업 시 반복적으로 등장하는 작업들의 모음으로, 구축 대상 데이터의 특성에 따라, 일부 작업요소만 포함

(예) 노인 음성 데이터의 작업요소 : ④, ⑨, ⑩

고객응대 음성 데이터의 작업요소 : ②, ③, ⑤, ⑦, ⑨, ⑩

표 3-12 | 음성 유형 작업요소 및 작업요소별 비용산정 요소

작업그룹	작업요소	물량 (규모)	인원 (작업자)	기간 (소요시간)	장비 (활용기기)	장소 (작업장소)	비용 구분	
수집	① 콘텐츠 구매	건					직접경비	
	스크립트	② 스크립트 작성	문장	인				직접인건비
		③ 스크립트 검수	문장	인				직접인건비
	녹음	④ 일반 녹음	시간	인				직접인건비
		⑤ 스크립트 녹음	장	인				직접인건비
		⑥ 특수 녹음	건	인				직접인건비
		⑦ 장소 임차		인			곳	직접경비
		⑧ 부대 업무	건					직접경비
정제	⑨ 녹음 검수	시간	인				직접인건비	
가공	⑩ 전사(라벨링)	시간	인				직접인건비	
검수	⑪ 검수(전수)		인				직접인건비	

음성 - ① 콘텐츠 구매

비용 구분	직접경비	작업량	(변동)		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□

※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기

**▶ 설명**

본 작업요소는 원시데이터 생성에 활용이 가능하나, 직접적인 구축이 어려운 문서나 음성 등의 콘텐츠\*를 구매하는 행위로 다음의 작업들을 포함하여 구성

\* 콘텐츠 : 행정문서, 뉴스 원문 데이터, 뉴스 원문 저작권, 방송 콘텐츠, 영화 등

- 콘텐츠가 기확보된 자료인지 확인
- 구매하는 콘텐츠 간 중복 여부 확인

**▶ 산식 = 물량 x 단가**

위 산식을 활용하여 본 작업요소의 비용을 산정

- 물량(건) : 콘텐츠 총구매량
- 단가(원) : 콘텐츠 구매단가

**▶ 기타 유의사항**

- 콘텐츠를 보유한 (공공)기관·기업의 내규에 따른 콘텐츠 구매단가의 편차 존재
- 콘텐츠 구매는 저작권료를 의미할 수 있으며, 각 콘텐츠의 특성에 따라 저작권의 유효기간 여부, 재가공 가능 여부 등 파악 필요
- 여러 유형의 콘텐츠를 구매하는 경우 각각의 콘텐츠 모두 직접경비 산정 양식에 계상 필요

음성 - ② 스크립트 작성																							
비용 구분	직접인건비	작업량	-																				
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)																		
	■	■	□	□	□																		
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																							
<p>▶ <b>설명</b></p> <p>본 작업요소는 음성 유형의 원시데이터 생성(녹음)을 위한 스크립트*를 작성하는 다음의 작업을 포함하여 구성</p> <p>* 스크립트 : 녹음작업의 대본과 각본 따위의 원고(시나리오)</p> <p>- 음성 데이터 구축 물량(시간)은 다음과 같은 음성-텍스트 환산기준에 따라 텍스트 기준 작업량으로 환산</p> <table border="1"> <thead> <tr> <th>음성 데이터</th> <th>실제 녹음 시간</th> <th>텍스트(A4 기준)</th> <th>텍스트(문장 기준)</th> </tr> </thead> <tbody> <tr> <td>녹음파일 5분</td> <td>10분 ※ 녹음 준비시간 5분</td> <td>A4 1장</td> <td>67문장</td> </tr> </tbody> </table> <p>- 녹음을 수행하는 주체(성별, 나이 등), 녹음 환경 및 언어 등을 고려하여 스크립트 작성</p> <p>- 스크립트는 한 번에 수집(녹음)할 수 있는 분량으로 작성하되, 발화의 길이는 일정 길이 이상의 문장 혹은 단어로 이루어져 녹음할 수 있도록 구성</p> <p>▶ <b>산식 = 투입공수 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <p>- 투입공수(M/M) = 물량 / 작업량 x 반려율 x 보정요소</p> <ul style="list-style-type: none"> <li>• 물량(문장) : 스크립트 총작성 분량</li> <li>• 작업량(문장) : 작업자 1인의 스크립트 작성 분량</li> <li>• 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">작업 유형</td> <td>한국어</td> <td>1.00</td> <td>공사장, 전철, 매장, 거리 등 소음 환경 발화, 아동의 실제 발화 및 대화</td> </tr> <tr> <td>한영 혼합</td> <td>4.00</td> <td>일상 대화에서 한국어-외래어 혼합 문장표현 및 발화</td> </tr> </tbody> </table>					음성 데이터	실제 녹음 시간	텍스트(A4 기준)	텍스트(문장 기준)	녹음파일 5분	10분 ※ 녹음 준비시간 5분	A4 1장	67문장	보정요소		보정계수	보정요소 설명	작업 유형	한국어	1.00	공사장, 전철, 매장, 거리 등 소음 환경 발화, 아동의 실제 발화 및 대화	한영 혼합	4.00	일상 대화에서 한국어-외래어 혼합 문장표현 및 발화
음성 데이터	실제 녹음 시간	텍스트(A4 기준)	텍스트(문장 기준)																				
녹음파일 5분	10분 ※ 녹음 준비시간 5분	A4 1장	67문장																				
보정요소		보정계수	보정요소 설명																				
작업 유형	한국어	1.00	공사장, 전철, 매장, 거리 등 소음 환경 발화, 아동의 실제 발화 및 대화																				
	한영 혼합	4.00	일상 대화에서 한국어-외래어 혼합 문장표현 및 발화																				

<ul style="list-style-type: none"> <li>• 반려율 : 작업물(스크립트)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> </ul> <p>- 단가(원) : 작업자 1인의 인건비</p>
<p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 투입공수는 보정요소인 작업 유형에 따라 변동</li> <li>- 전체 데이터의 스크립트 구성의 다양성을 위해 작업자 1인의 스크립트 작성 분량을 제한</li> <li>- 단가는 작업자 1인의 스크립트 작성 분량에 대한 인건비로 작업 난이도에 따라 변동 가능</li> </ul>

음성 - ③ 스크립트 검수															
비용 구분	직접인건비	작업량	1,500문장/시간												
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)										
	■	□	■	□	□										
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기															
<p>▶ <b>설명</b></p> <p>본 작업요소는 ‘녹음’ 작업의 기초가 되는 스크립트를 검수하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 작성된 스크립트에 의도한 상황 및 환경 요인이 반영되어 있고, 이를 바탕으로 녹음하여 데이터를 획득할 수 있는지 검수하는 작업</li> <li>- 음성 데이터의 정확성 및 다양성에 영향을 끼칠 수 있는 스크립트 작성 내용 검토</li> <li>- 스크립트의 전체 분량, 문장의 구성 등을 통해 발화자가 한 번에 녹음이 가능한 분량인지, 발화의 길이가 밸런싱 있게 녹음될 수 있는지 검토</li> <li>- 중복 문장 및 내용, 비속어, 은어, 개인정보 등 오류 사항을 검토</li> </ul> <p>▶ <b>산식 = 투입공수 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일 x 반려율 x 보정요소                     <ul style="list-style-type: none"> <li>• 물량(문장) : 스크립트 총작성 분량</li> <li>• 작업량(문장) : 1인 스크립트 검수 분량</li> <li>• 월평균 근무일(일) : 21일</li> <li>• 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> </li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">작업 난이도</td> <td>고</td> <td>1.00</td> <td>공사장, 전철, 매장, 거리 등 소음 환경 발화, 아동의 실제 발화 및 대화</td> </tr> <tr> <td>저</td> <td>0.67</td> <td>일상 대화에서 한국어-외래어 혼합 문장표현 및 발화</td> </tr> </tbody> </table>					보정요소		보정계수	보정요소 설명	작업 난이도	고	1.00	공사장, 전철, 매장, 거리 등 소음 환경 발화, 아동의 실제 발화 및 대화	저	0.67	일상 대화에서 한국어-외래어 혼합 문장표현 및 발화
보정요소		보정계수	보정요소 설명												
작업 난이도	고	1.00	공사장, 전철, 매장, 거리 등 소음 환경 발화, 아동의 실제 발화 및 대화												
	저	0.67	일상 대화에서 한국어-외래어 혼합 문장표현 및 발화												

<ul style="list-style-type: none"> <li>• 반려율 : 작업물(스크립트 검수본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 참여 인력 월 인건비</li> </ul>
<p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 다른 유형의 콘텐츠(소설, 영화 및 드라마 대본) 내용을 차용한 경우, 저작권 협의(위반) 여부 및 저작권료 검토 필요</li> </ul>

음성 - ④ 일반 녹음																				
비용 구분	직접인건비	작업량	3.25시간/인																	
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)																
	■	■	□	□																
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																				
<p>▶ <b>설명</b></p> <p>본 작업요소는 스크립트가 없고 대화 주제만 주어진 일반 상황에서 자유 발화를 녹음하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 자유롭고 편안한 상황에서 대화할 수 있도록 그룹 단위로(예. 1조 2인) 화자를 모집하여 진행</li> <li>- 녹음 화자가 사회적 중립성(민감성*)을 벗어나거나, 편중된 내용을 발언하지 않도록 사전에 체계적인 주제와 자료를 제시</li> <li>* 민감성 : 정치적 견해, 특정인물 비하, 성적인 표현 등</li> <li>- 녹음 화자가 녹음 상황임을 인지하지 않고 자연스럽게 참여할 수 있도록 환경 조성</li> </ul> <p>▶ <b>산식 = 물량 / 작업량 x 반려율 x 보정요소 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 물량(시간) : 총녹음 분량</li> <li>- 작업량(시간) : 작업자 1인의 녹음 분량(발화시간)</li> <li>- 반려율 : 작업물(일반 녹음본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="4">작업 유형</td> <td>스튜디오 녹음 및 영상 촬영</td> <td>0.25</td> <td>스튜디오 환경에서 녹음과 영상 촬영을 병행</td> </tr> <tr> <td>스튜디오 녹음</td> <td>0.50</td> <td>스튜디오 환경에서 자유 발화 녹음</td> </tr> <tr> <td rowspan="2">클라우드소싱</td> <td>1.00</td> <td>클라우드소싱을 통한 일반년배 성인의 자유 발화 녹음</td> </tr> <tr> <td>1.625</td> <td>클라우드소싱을 통한 아동, 노인의 자유 발화 녹음</td> </tr> </tbody> </table>					보정요소		보정계수	보정요소 설명	작업 유형	스튜디오 녹음 및 영상 촬영	0.25	스튜디오 환경에서 녹음과 영상 촬영을 병행	스튜디오 녹음	0.50	스튜디오 환경에서 자유 발화 녹음	클라우드소싱	1.00	클라우드소싱을 통한 일반년배 성인의 자유 발화 녹음	1.625	클라우드소싱을 통한 아동, 노인의 자유 발화 녹음
보정요소		보정계수	보정요소 설명																	
작업 유형	스튜디오 녹음 및 영상 촬영	0.25	스튜디오 환경에서 녹음과 영상 촬영을 병행																	
	스튜디오 녹음	0.50	스튜디오 환경에서 자유 발화 녹음																	
	클라우드소싱	1.00	클라우드소싱을 통한 일반년배 성인의 자유 발화 녹음																	
		1.625	클라우드소싱을 통한 아동, 노인의 자유 발화 녹음																	

<ul style="list-style-type: none"> <li>- 단가 : 작업자 1인의 인건비</li> <li>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</li> </ul>
<p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 투입공수는 보정요소인 작업 유형에 따라 변동</li> <li>- 전체 데이터의 구성을 위해 작업자 1인의 녹음 분량(발화시간)을 제한</li> <li>- 단가는 작업자 1인의 녹음 분량에 대한 인건비로 작업 난이도에 따라 변동 가능</li> </ul>

음성 - ⑤ 스크립트 녹음					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		■	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 작성된 스크립트를 바탕으로 발화된 음성을 녹음을 하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 준비된 스크립트(시나리오)를 발화자가 발화</li> <li>- 녹음된 음성을 들어보고 수정이 필요한 경우 재녹음</li> </ul> <p>▶ <b>산식 = 물량 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 물량(장) : 총녹음 분량</li> <li>- 단가(원) : A4용지 1장의 녹음 비용</li> </ul> <p>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</p> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 물량은 문장단위로 작성된 스크립트를 A4 용지 1장으로 환산한 수치</li> <li>- 단가는 스크립트를 A4 용지 기준으로 환산한 A4 용지 1장의 녹음 비용으로 책정하며 작업 난이도에 따라서 변동</li> </ul>					

음성 - ⑥ 특수 녹음					
비용 구분	직접인건비	작업량	2건/시간		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		■	□	■	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 콜센터 및 고객센터 전화 통화 등 특수한 상황에서 발생하는 대화를 녹음하는 것으로 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 콜센터 및 고객센터의 경우, 상담 음성 녹음 및 AI 데이터로 활용 동의의 진행 후 녹음</li> </ul> <p>▶ <b>산식 = 물량 x 작업량 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 물량(시간) : 총녹음 분량</li> <li>- 작업량(건) : 시간당 녹음 건수</li> <li>- 단가(원) : 녹음 1건의 개인정보 제공 및 활용 동의를 위한 비용</li> </ul> <p>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</p> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 실제 상황의 음성 데이터를 수집해야 하므로 참여 당사자의 개인정보 등 법적 침해요인을 점검하여 침해요인 사전 차단 필요</li> <li>- 개별 녹음 1건당 개인정보 제공 및 활용 동의를 위한 비용 책정 필요</li> </ul>					

음성 - ⑦ 장소 임차				
비용 구분	직접경비	작업량	34개/월	
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)
	□	□	■	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기				
<p>▶ <b>설명</b></p> <p>본 작업요소는 음성 녹음을 위하여 물리적 공간을 임차하기 위한 다음의 작업을 포함</p> <ul style="list-style-type: none"> <li>- 비용이 수반되는 대부분의 장소 임차는 실내 공간인 스튜디오 임차를 의미</li> <li>- 스튜디오 임차 비용에는 공간 뿐 아니라, 스튜디오 내 마이크 등 녹음 장비 포함 가능</li> </ul> <p>▶ <b>산식 = 장소 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 장소(곳) : 녹음이 진행되는 공간의 임차 수량</li> <li>- 단가(원) : 1개월 임차 비용</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 장소는 1개월 단위로 스튜디오를 임차할 경우 전체 음성 데이터 녹음에 필요한 스튜디오의 수를 의미</li> <li>- 스튜디오 녹음 작업량 및 스튜디오의 규모에 따라 필요한 장소의 수와 단가가 변동</li> <li>- 구축하고자 하는 데이터의 특성에 따라 온라인 작업장이 필요한 경우도 존재</li> </ul>				

음성 - ⑧ 부대 업무				
비용 구분	직접경비	작업량	(변동)	
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)
	□	□	■	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기				
<p>▶ <b>설명</b></p> <p>본 작업요소는 원시데이터 생성에 필요한 기타 부대 업무(운영 등)에 대한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 전화망 데이터 수집을 위한 콜센터 및 고객센터 운영비 등</li> </ul> <p>▶ <b>산식</b></p> <p>※ 변동</p> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당 없음</li> </ul>				

음성 - ⑨ 녹음 검수																		
비용 구분	직접인건비		작업량		-													
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)													
	■	■	□	□	□													
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																		
<p>▶ <b>설명</b></p> <p>본 작업요소는 녹음된 음성을 검수하기 위한 다음의 작업들을 포함</p> <ul style="list-style-type: none"> <li>- 음원의 왜곡, 불필요한 무음 등을 검토하여 녹음 수집 데이터의 적/부 판별</li> </ul> <p>▶ <b>산식 = 투입공수 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 월평균 근무일 / 1일 업무 시간 x 반려율 x 보정요소                     <ul style="list-style-type: none"> <li>• 물량(시간) : 총녹음 분량</li> <li>• 월평균 근무일(일) : 21일</li> <li>• 1일 업무 시간(시간) : 8시간</li> <li>• 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> </li> </ul> <table border="1"> <thead> <tr> <th>보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="3">검수업무 가중치</td> <td>저</td> <td>1</td> <td>화자인식 등 검수 난이도가 낮은 경우</td> </tr> <tr> <td>중</td> <td>2</td> <td>일반 검수 작업인 경우</td> </tr> <tr> <td>고</td> <td>3</td> <td>아동 음성 등 검수 난이도가 높은 경우</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>• 반려율 : 작업물(녹음 검수본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> </ul> <ul style="list-style-type: none"> <li>- 단가(원) : 참여 인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 음성 유형의 '⑨ 녹음 검수' 작업요소는 '②, ③스크립트 작성·검수', '④~⑧녹음' 등의 작업요소 대비 보다 높은 숙련도를 요구하므로 검수 업무의 가중치를 적용하여 투입공수 보정</li> </ul>						보정요소	보정계수	보정요소 설명	검수업무 가중치	저	1	화자인식 등 검수 난이도가 낮은 경우	중	2	일반 검수 작업인 경우	고	3	아동 음성 등 검수 난이도가 높은 경우
보정요소	보정계수	보정요소 설명																
검수업무 가중치	저	1	화자인식 등 검수 난이도가 낮은 경우															
	중	2	일반 검수 작업인 경우															
	고	3	아동 음성 등 검수 난이도가 높은 경우															

음성 - ⑩ 가공																																
비용 구분	직접인건비		작업량		-																											
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)																											
	■	■	□	□	□																											
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																																
<p>▶ <b>설명</b></p> <p>본 작업요소는 녹음된 음성 데이터를 가공(라벨링)하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 음성 파일과 발화 텍스트 파일 지정, 메타데이터 입력 등</li> <li>- 전사 및 표준어 이종 전사 등 음성을 텍스트로 옮기는 라벨링 작업</li> <li>- 전사된 텍스트 및 라벨링 데이터를 검수하여 적/부 판별</li> </ul> <p>▶ <b>산식 = 투입공수 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 x 보정요소 / 월평균 근무일 / 1일 업무 시간 x 반려율                     <ul style="list-style-type: none"> <li>• 물량(시간) : 총 녹음 분량</li> <li>• 보정요소 및 보정계수 : 보정 요소별 가중치의 합(①+②+③+④)</li> </ul> </li> </ul> <table border="1"> <thead> <tr> <th>보정 요소</th> <th>보정 요소 설명</th> <th>가중치</th> </tr> </thead> <tbody> <tr> <td rowspan="3">① 메타데이터 입력</td> <td>메타데이터 입력을 수행하지 않는 경우</td> <td>0</td> </tr> <tr> <td>메타데이터 입력을 수행하는 경우</td> <td>1</td> </tr> <tr> <td>화자인식 등 메타데이터 입력 작업량이 많은 경우</td> <td>2</td> </tr> <tr> <td rowspan="3">② 전사 입력</td> <td>전사 입력을 수행하지 않는 경우</td> <td>0</td> </tr> <tr> <td>전사 입력을 수행하는 경우</td> <td>1</td> </tr> <tr> <td>자유 발화 등 전사 작업량이 많은 경우</td> <td>3</td> </tr> <tr> <td rowspan="2">③ 표준어 이종 전사</td> <td>표준어 이종 전사 작업이 없는 경우</td> <td>0</td> </tr> <tr> <td>자유 발화 등 표준어 이종 전사 작업이 필요한 경우</td> <td>0.5</td> </tr> <tr> <td rowspan="2">④ 전사 검수</td> <td>전사 검수를 수행하는 경우</td> <td>1</td> </tr> <tr> <td>저음질 등 검수 난이도가 높은 경우</td> <td>2</td> </tr> </tbody> </table> <p>보정계수 = 보정 요소별 가중치의 합 (① + ② + ③ + ④)</p> <ul style="list-style-type: none"> <li>• 월평균 근무일(일) : 21일</li> <li>• 1일 업무 시간(시간) : 6.8시간</li> </ul>						보정 요소	보정 요소 설명	가중치	① 메타데이터 입력	메타데이터 입력을 수행하지 않는 경우	0	메타데이터 입력을 수행하는 경우	1	화자인식 등 메타데이터 입력 작업량이 많은 경우	2	② 전사 입력	전사 입력을 수행하지 않는 경우	0	전사 입력을 수행하는 경우	1	자유 발화 등 전사 작업량이 많은 경우	3	③ 표준어 이종 전사	표준어 이종 전사 작업이 없는 경우	0	자유 발화 등 표준어 이종 전사 작업이 필요한 경우	0.5	④ 전사 검수	전사 검수를 수행하는 경우	1	저음질 등 검수 난이도가 높은 경우	2
보정 요소	보정 요소 설명	가중치																														
① 메타데이터 입력	메타데이터 입력을 수행하지 않는 경우	0																														
	메타데이터 입력을 수행하는 경우	1																														
	화자인식 등 메타데이터 입력 작업량이 많은 경우	2																														
② 전사 입력	전사 입력을 수행하지 않는 경우	0																														
	전사 입력을 수행하는 경우	1																														
	자유 발화 등 전사 작업량이 많은 경우	3																														
③ 표준어 이종 전사	표준어 이종 전사 작업이 없는 경우	0																														
	자유 발화 등 표준어 이종 전사 작업이 필요한 경우	0.5																														
④ 전사 검수	전사 검수를 수행하는 경우	1																														
	저음질 등 검수 난이도가 높은 경우	2																														

- 반려율 : 작업물(음성 가공본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율
- 단가(원) : 참여 인력 월 인건비

▶ 기타 유의사항

- 물량은 보정요소인 가공업무의 수행여부에 따라 가중치가 부여되며 가중치의 합으로 보정

음성 - ⑩ 검수(전수)					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□

※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기

▶ 설명

본 작업요소는 최종 산출물을 생성하기 위한 다음의 작업을 포함하여 구성

- 최종 데이터 1쌍(원천, 라벨링 데이터)을 생성하기 위해 앞 단계에서 산출된 데이터의 품질을 전수 검사하는 작업
- 본 단계 외에도, 데이터 구축 작업단계(작업그룹)마다 교차 품질 검사는 병행되며, 음성 데이터의 반려율 비율이 높은 작업은 작업그룹(③, ⑤)으로 명시

▶ 산식 = 투입공수 x 반려율 x 단가

위 산식을 활용하여 본 작업요소의 비용을 산정

- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일
  - 물량(장/건) : 음성 원시, 원천데이터 총생성량
  - 작업량(건) : 1일 원시, 원천데이터 검사 수량
  - 월평균 근무일(일) : 21일
- 반려율 : 작업물(최종파일)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율
- 단가 : 참여인력 월 인건비

▶ 기타 유의사항

- 해당사항 없음

#### 4) 데이터 유형 : 비디오(영상)

##### ▶ 비디오 유형 AI 학습용 데이터 구축 특징

- 원시데이터가 비디오 유형인 AI 학습용 데이터는 '동영상 분류', '객체 인식', '객체 추적' 등의 문제 해결을 위한 목적으로 구축
- 비디오 유형의 데이터는 이미 존재하는 비디오 유형의 콘텐츠를 구매하여 정제하거나, 직접 촬영한 파일을 정제하여 원천데이터를 생성
- 확보된 원천데이터는 가공(라벨링) 과정을 통해 라벨링 데이터를 생성하고, 데이터 검수를 통해 1쌍의 인공지능 데이터를 도출
  - ※ 영상 데이터는 프레임 단위의 이미지가 연속적으로 구성된 것으로, 데이터의 목적에 따라, 영상 파일 자체를 가공하거나, 프레임 단위에서 가공 후 영상 파일로 통합

[표 3-13]에 제시된 텍스트 유형 작업요소는 구축 작업 시 반복적으로 등장하는 작업들의 모음으로, 구축 대상 데이터의 특성에 따라, 일부 작업요소만 포함

(예) 회사생활 비디오 데이터의 작업요소 : ②, ③, ④, ⑧, ⑨  
 스포츠영상 비디오 데이터의 작업요소 : ①, ⑤, ⑥, ⑦, ⑨

표 3-13 | 비디오 유형 작업요소 및 작업요소별 비용산정 요소

작업그룹	작업요소	물량 (규모)	인원 (작업자)	기간 (소요시간)	장비 (활용기기)	장소 (작업장소)	비용 구분	
수집	구매 ① 콘텐츠 구매	건					직접경비	
	촬영	② 영상 촬영	건	인	일			직접인건비
		③ 장비 임차	개	인	일			직접경비
		④ 장소 임차		인	일		곳	직접경비
		⑤ 부대 업무		인	일			직접경비
정제	⑥ 정제	건	인	일			직접인건비	
가공	⑦ 가공(영상)	건	인	일			직접인건비	
	⑧ 가공(이미지)	장	인	일			직접인건비	
검수	⑨ 검수(전수)	건	인	일			직접인건비	

비디오 - ① 콘텐츠 구매				
비용 구분	직접경비	작업량	-	
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)
	■	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기				

▶ **설명**  
 본 작업요소는 원시데이터 생성에 활용이 가능하나, 직접적인 구축이 어려운 문서나 영상 등의 콘텐츠\*를 구매하는 행위로 다음의 작업들을 포함하여 구성  
 \* 콘텐츠 : 행정문서, 뉴스 원문 데이터, 뉴스 원문 저작권, 방송 콘텐츠, 영화 등

- 콘텐츠의 기본 언어는 한국어로 설정
- 콘텐츠가 확보된 자료인지 확인 필요
- 구매하는 콘텐츠 간 중복 여부 확인 필요

▶ **산식 = 물량 x 단가**  
 위 산식을 활용하여 본 작업요소의 비용을 산정

- 물량(개) : 콘텐츠 총구매량
- 단가(원) : 콘텐츠 건당 구매 비용

▶ **기타 유의사항**

- 콘텐츠를 보유한 기관(기업) 내규에 따른 콘텐츠 구매단가의 편차 존재
- 콘텐츠 구매는 저작권료를 의미할 수 있으며, 각 콘텐츠의 특성에 따라 저작권의 유효기간 여부, 재가공 가능 여부 등 파악 필요
- 여러 유형의 콘텐츠를 구매하는 경우 각각의 콘텐츠 모두 직접경비 산정 양식에 계상 필요

비디오 - ② 영상 촬영					
비용 구분	직접인건비	작업량	촬영자 2~3시간/일 촬영자 1~3인/일		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		■	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 영상 데이터를 획득(촬영)하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 기존에 존재하지 않는 영상 원시데이터를 획득하기 위하여 카메라 등의 촬영 장비를 활용하여 영상 데이터를 생성하는 작업</li> <li>- 영상 촬영 작업의 참여인력은 ① 촬영을 직접하는 촬영자(및 촬영 보조인력), ② 촬영 대상(피사체)이 되는 모델 인력으로 구분되며, 대체로 ①의 촬영자를 의미</li> <li>- 구축할 영상 데이터의 작업(촬영) 난이도에 따라, 참여인력 ①, ②는 특정 도메인 전문인력의 참여 필요</li> </ul> <p>▶ <b>산식 = 인원 x 기간 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 인원(인) : 영상 촬영 총참여인력</li> <li>- 기간(일) : 영상 촬영 총소요기간</li> <li>- 반려율 : 작업물(촬영본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 참여인력 1인당 영상 촬영 비용</li> </ul> <p>※ 본 산식은 물량 기준 직접인건비 산정 양식에 적용</p> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 촬영 대상(피사체)이 사람(모델)인 경우, 데이터의 노출범위와 사용범위에 대해 고지하고 초상권이용 동의서 등 데이터의 저작권에 대한 사전 협의(관련 서류, 비용) 필수</li> <li>- 촬영대상이 사람이 아니어도 개인의 식별이 가능하게 하는 객체(자동차, 소지품 등)가 포함된 경우 저작권획득 필요</li> </ul>					

비디오 - ③ 장비 임차					
비용 구분	직접경비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
		□	■	■	■
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 영상 원시데이터를 생성하기 위하여 활용되는 모든 기기(사물)를 임차하는 작업을 포함</p> <ul style="list-style-type: none"> <li>- 영상 촬영시 활용되는 기기는 대체로 영상 촬영 필수 장비인 카메라와 촬영 환경 조성을 위한 부대장비, 즉 조명, 지미집, 크레인 등이며, 해당 기기 임차 비용을 의미</li> <li>- 촬영 대상(피사체)에 사물이 포함되고, 임차가 필요한 경우 장비 임차 진행</li> <li>- 구축할 영상 데이터가 특정 품질 수준이 요구되는 경우, 촬영본(원시데이터)의 품질 확보를 위한 특수 촬영 장비 및 부속 장비에 대한 고려(비용) 필요</li> </ul> <p>※ 예 : 3D, 4D 등 다차원 영상은 다시점 카메라 등 특수장비 필요 모션 영상은 카메라와 데이터를 연계하는 센서 장비들을 함께 고려 필요</p> <p>▶ <b>산식 = 장비 x 인원 x 기간 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 장비(개) : 영상 촬영장비 및 피사체(사물) 총임차 수량</li> <li>- 인원(인) : 장비 활용 총인원</li> <li>- 기간(일) : 장비 임차 총소요기간</li> <li>- 단가(원) : 장비 개당 임차 비용</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 데이터 구축 작업에 장기적으로 사용되며, 민감 장비의 경우 자산 구매 고려 필요</li> <li>※ 일반 촬영 카메라와 특수 촬영장비 카메라의 임차 비용은 최소 약 5배 이상 차이</li> <li>- 고해상도 영상 촬영 기기 중 해외 수급이 필요한 경우 장비 확보 일정을 포함하여 촬영 계획 수립 필요</li> <li>- 영상 촬영본은 구축 목적에 따라 편차는 있으나, FHD(1920*1080px) 수준 이상 준수 필요</li> </ul>					

비디오 - ④ 장소 임차					
비용 구분	직접경비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	□	□	■	□	■
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					

**▶ 설명**  
 본 작업요소는 영상 원시데이터를 생성하기 위하여 물리적 공간을 임차하는 작업 포함

- 비용이 수반되는 대부분의 장소 임차는 실내 공간인 스튜디오 임차를 의미하며, 구축할 데이터의 특성에 따라 필요로 하는 공간의 규모 및 비용 변동  
 ※ 예 : 사람 1인 촬영 vs 여러사람 활동 촬영, 일반 카메라 촬영 vs 다차원 카메라 촬영
- 스튜디오 임차 비용에는 공간 뿐 아니라 스튜디오 내에 설치된 장비(카메라, 조명 등)가 기본 사항으로 포함 가능

**▶ 산식 ① = 장소 x 기간 x 단가**  
 위 산식을 활용하여 본 작업요소의 비용을 산정

- 장소(곳) : 촬영이 진행되는 공간의 임차 수량
- 기간(일) : 촬영장소 임차 총소요기간
- 단가(원) : 촬영장소 1곳당 임차 비용  
 ※ 장기 대여의 경우, 투입공수를 적용하여 총 임차비용 산출

**산식 ② = 스튜디오 대여 투입공수 x 단가**

- 투입공수(M/M) = 인원 / 작업량 / 월평균 근무일
  - 인원(인) : 총 촬영 인원
  - 작업량(인) : 1일 촬영 인원
  - 월평균 근무일(일) : 21일
- 단가(원) : 촬영장소 임차 비용

**▶ 기타 유의사항**

- 생성하는 콘텐츠의 특성에 따라, 실외 공간에 대한 임차료가 필요한 경우도 존재  
 ※ 예 : 여러 실외 장소를 배경으로 하는 영상 콘텐츠 생성 등

비디오 - ⑤ 부대 업무					
비용 구분	직접경비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					

**▶ 설명**  
 본 작업요소는 영상 원시데이터를 생성하기 위해 주 작업요소에 수반되는 작업을 포함

- 부대 업무는 데이터 구축 진행 과정에서 사전 계획 없이 추가로 발생할 수 있으며, 총 예산 가용 범위를 넘지 않도록 구축 설계 단계에서 면밀한 작업요소 설계 및 예비비 준비 필요  
 ※ 예 : 스포츠 경기 촬영을 위한 관람 비용 지원 등

**▶ 산식 = 물량 x 단가**  
 위 산식을 활용하여 본 작업요소의 비용을 산정

- 물량(개) : 부대 업무 수량
- 단가(원) : 업무 건당 지원 비용

**▶ 기타 유의사항**

- 해당 사항 없음

비디오 - ⑥ 정제					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 영상 원시데이터의 파일 규격 통일, 비식별화 등을 전처리(정제)하여 원천데이터를 생성하는 작업을 포함</p> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일</li> <li>• 물량(건) : 영상 데이터 총생성량</li> <li>• 작업량(건) : 1일 영상 데이터 정제 수량</li> <li>• 월평균 근무일(일) : 21일</li> <li>- 반려율 : 작업물(정제본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 참여인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 특수영상(모션캡처) 데이터는 센서 데이터와 동기화 여부 등이 전제되어야 활용 가능</li> <li>- 모션캡처 영상은 인체 구조를 촬영한 데이터로 비식별화 이슈가 없음</li> </ul>					

비디오 - ⑦ 가공(영상)																
비용 구분	직접인건비	작업량	3개/일													
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)											
	■	□	□	□	□											
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기																
<p>▶ <b>설명</b></p> <p>본 작업요소는 영상 원천데이터를 가공하여 라벨링데이터를 생성하는 작업을 포함</p> <ul style="list-style-type: none"> <li>- 영상 원천데이터에 의미있는 정보를 라벨링하는 작업으로, 여러 정보(영상/음성/언어 등)가 포함된 영상 데이터에는 다양한 의미정보를 부착하여 멀티모달 데이터 셋으로 활용 가능</li> <li>※ 영상 멀티모달 데이터는 영상 데이터에 등장인물의 정보 및 감정, 발화 정보 및 의도, 개체 정보 및 관계 등의 메타 정보를 포함</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일 x 보정요소</li> <li>• 물량(건) : 영상 데이터 총생성량</li> <li>• 작업량(건) : 1일 영상 데이터 가공 수량</li> <li>• 월평균 근무일(일) : 21일</li> <li>• 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">작업 난이도</td> <td>고</td> <td>1.00</td> <td>특정분야 전문가가 필수적으로 참여하여 처리할 수 있는 물량</td> </tr> <tr> <td>저</td> <td>0.06</td> <td>일반 데이터 가공 작업자가 처리할 수 있는 물량</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>- 반려율 : 작업물(가공파일)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 참여인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 편집 영상의 경우 가공시간은 원본 영상의 3배로 상정 필요</li> </ul>						보정요소		보정계수	보정요소 설명	작업 난이도	고	1.00	특정분야 전문가가 필수적으로 참여하여 처리할 수 있는 물량	저	0.06	일반 데이터 가공 작업자가 처리할 수 있는 물량
보정요소		보정계수	보정요소 설명													
작업 난이도	고	1.00	특정분야 전문가가 필수적으로 참여하여 처리할 수 있는 물량													
	저	0.06	일반 데이터 가공 작업자가 처리할 수 있는 물량													

비디오 - ⑧ 가공(이미지)															
비용 구분	직접인건비	작업량	20장/일												
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)										
	■	□	□	□	□										
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기															
<p>▶ <b>설명</b></p> <p>본 작업요소는 영상 원천데이터를 프레임 단위로 라벨링데이터를 생성하는 작업을 포함</p> <ul style="list-style-type: none"> <li>- 영상 파일은 이미지(프레임) 파일로 구성되며, 본 작업은 영상 데이터 파일을 프레임 단위의 각 파일로 가공하는 작업을 의미하며, 통상 영상 라벨링 작업은 본 작업을 의미</li> <li>※ 영상 1초당 프레임 각각에 라벨링 작업을 수행하며, 초당 프레임 수(fps)는 파일 특성에 따라 변동 가능</li> <li>- 일반적으로 가공 대상인 영상 파일의 전체 이미지 파일에 라벨링을 수행하지만, 데이터 특성 또는 목표 구축량에 따라 초당 3프레임, 5프레임 등의 단위로 가공 작업을 수행</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일 x 보정요소                     <ul style="list-style-type: none"> <li>• 물량(장) : 영상 데이터 총생성량의 이미지(프레임) 단위 변환 수량</li> <li>• 작업량(장) : 1일 이미지 가공 수량</li> <li>• 월평균 근무일(일) : 21일</li> <li>• 보정요소 및 보정계수 : 본 작업요소는 다음과 같은 보정요소 및 보정계수 적용</li> </ul> </li> </ul> <table border="1"> <thead> <tr> <th colspan="2">보정요소</th> <th>보정계수</th> <th>보정요소 설명</th> </tr> </thead> <tbody> <tr> <td rowspan="2">작업 난이도</td> <td>고</td> <td>1</td> <td>다차원(3D 이상) 이미지 데이터를 가공 작업자가 처리할 수 있는 물량</td> </tr> <tr> <td>저</td> <td>0.2</td> <td>일차원 이미지 데이터를 가공 작업자가 처리할 수 있는 물량</td> </tr> </tbody> </table>					보정요소		보정계수	보정요소 설명	작업 난이도	고	1	다차원(3D 이상) 이미지 데이터를 가공 작업자가 처리할 수 있는 물량	저	0.2	일차원 이미지 데이터를 가공 작업자가 처리할 수 있는 물량
보정요소		보정계수	보정요소 설명												
작업 난이도	고	1	다차원(3D 이상) 이미지 데이터를 가공 작업자가 처리할 수 있는 물량												
	저	0.2	일차원 이미지 데이터를 가공 작업자가 처리할 수 있는 물량												

<ul style="list-style-type: none"> <li>- 반려율 : 작업물(가공본)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 참여인력 월 인건비</li> </ul>
<p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 정적인 영상파일의 경우, 각 이미지(프레임) 전후의 변동이 크지 않을 가능성이 높으며, 중복 라벨링 작업으로 인한 시간-비용을 줄일 수 있는 자동화 기술 고려 필요</li> <li>※ 예 : 이미지 파일 전 프레임의 라벨링 정보를 후 프레임에 적용하는 등의 라벨링 자동화 기술</li> </ul>

비디오 - ㉔ 검수					
비용 구분	직접인건비	작업량	-		
비용산정 요소	물량(규모)	인원(작업자)	기간(소요기간)	장비(활용기기)	장소(작업장소)
	■	□	□	□	□
※ 작업요소의 투입 공수를 산정하는데 포함된 비용산정 요소 표기					
<p>▶ <b>설명</b></p> <p>본 작업요소는 최종 산출물을 생성하기 위한 다음의 작업을 포함하여 구성</p> <ul style="list-style-type: none"> <li>- 최종 데이터 1쌍(원천, 라벨링 데이터)을 생성하기 위해 앞 단계에서 산출된 데이터의 품질을 전수 검사하는 작업</li> <li>- 본 단계 외에도, 데이터 구축 작업단계(작업그룹)마다 교차 품질 검사는 병행</li> </ul> <p>▶ <b>산식 = 투입공수 x 반려율 x 단가</b></p> <p>위 산식을 활용하여 본 작업요소의 비용을 산정</p> <ul style="list-style-type: none"> <li>- 투입공수(M/M) = 물량 / 작업량 / 월평균 근무일 <ul style="list-style-type: none"> <li>• 물량(개) : 영상 원시, 원천데이터 총 생성량</li> <li>• 작업량(건) : 1일 원시, 원천데이터 검수 수량</li> <li>• 월평균 근무일(일) : 21일</li> </ul> </li> <li>- 반려율 : 작업물(최종파일)이 품질 검사 기준을 충족하지 못하여 재작업이 요청된 비율</li> <li>- 단가(원) : 참여인력 월 인건비</li> </ul> <p>▶ <b>기타 유의사항</b></p> <ul style="list-style-type: none"> <li>- 해당사항 없음</li> </ul>					

## AI 학습용 데이터 구축 비용산정 가이드

**발행** 2022년 9월  
**행인** 황 중 성  
**기획** 한국지능정보사회진흥원  
지능데이터본부 고윤석 본부장, AI데이터기획팀 윤미영 팀장  
**작성** 한국지능정보사회진흥원 지능데이터본부 AI데이터기획팀  
박지영 책임, 임민성 선임, 추하원 선임  
**문의** jiyoung.park@nia.or.kr(한국지능정보사회진흥원 AI데이터기획팀 박지영 책임)  
**주소** (41068) 대구광역시 동구 첨단로 53  
**홈페이지** www.nia.or.kr / www.aihub.or.kr  
**ISBN** 978-89-8483-724-9

- 본 가이드의 내용은 한국지능정보사회진흥원의 공식견해와 다를 수 있습니다.
- 본 가이드 내용에 대한 무단전제를 금하며, 가공 및 인용할 때에는 반드시 출처를 명기하여 주시기 바랍니다.
- 본 책자는 정보통신기금으로 제작되었습니다.(비매품)

## AI 학습용 데이터 구축 비용산정 가이드