

인공지능 학습용 데이터  
품질관리 가이드라인 및 구축 안내서 v3.0

2023.02

# 품질관리 가이드라인 v3.0

## 요약본

# CONTENTS

제 1 장 요약본 구성 및 활용 .....	1
제 2 장 품질관리 프로세스 및 산출물 .....	5
1. 준비·계획단계 .....	7
2. 구축단계 .....	8
3. 운영·활용단계 .....	13
4. 단계별 산출물 .....	16
제 3 장 품질 자가점검 및 품질검증 .....	19
1. 품질 자가점검 .....	21
2. 제 3 자 품질검증 .....	28
〈부록〉 1. 주요 산출물 관계도 .....	41
1. 품질관리 가이드라인 및 구축 안내서 v3.0 과 주요 산출물 관계도 .....	41
〈부록〉 2. 품질 자가점검 체크리스트 양식 .....	44
1. 과제조정위원회 자가점검 체크리스트 .....	44
2. 샘플데이터 1-Cycle 자가점검 체크리스트 .....	49
3. 초기데이터 1-Cycle 자가점검 체크리스트 .....	51
4. 중간데이터 품질 자가점검 체크리스트 .....	53







## 제1장

# 요약본 구성 및 활용



## 제1장 | 요약본 구성 및 활용

- 인공지능 학습용 데이터 품질관리 가이드라인 요약본은 인공지능 학습용 데이터 품질관리 담당자를 주요 대상으로 업무 담당자, 구축 담당자 등 품질관리 관련 업무를 수행하는 실무자들이 상시 활용하여 인공지능 학습용 데이터 품질관리를 효율적으로 수행할 수 있도록 구성

### I. 총론 구성

- 요약본 활용 대상 및 구성 요약

### II. 품질관리 체계 구성

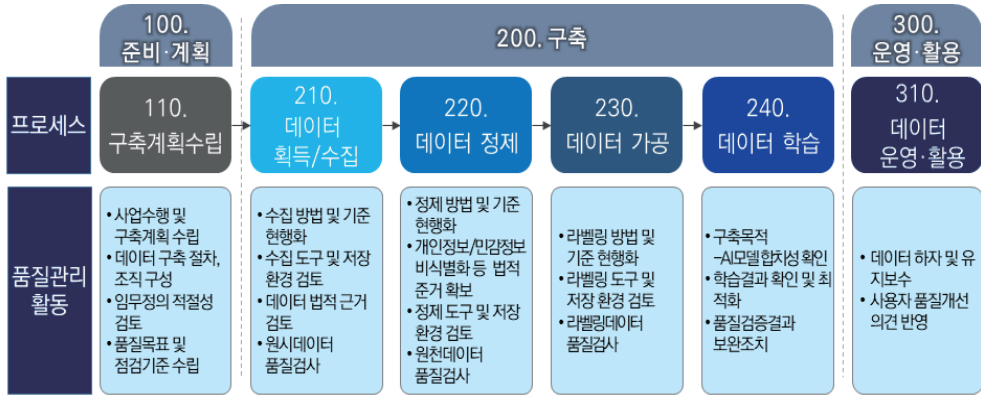
- ‘품질관리 프레임워크’ 단계별 프로세스와 품질관리 활동
- 품질관리 활동의 중요한 요소인 1-Cycle 자가점검에 대한 프로세스 및 품질관리 활동

- 인공지능 학습용 데이터 품질관리 프레임워크 구조

품질관리 프레임워크 구조	단계	• 인공지능 학습용 데이터의 생애주기 기반 단계 식별
	프로세스	• 인공지능 학습용 데이터의 공통된 구축 프로세스 정의
	산출물	<ul style="list-style-type: none"> <li>• 인공지능 학습용 데이터 구축 프로세스를 통해 생성되는 데이터 정의</li> <li>• 구축 프로세스 및 데이터 생성에 따른 구축 및 품질 관련 산출물 정의</li> </ul>
	구축 프로세스별 생성 데이터	
	구축 프로세스별 산출물	
	품질관리 활동	<ul style="list-style-type: none"> <li>• 구축 프로세스 품질관리 활동</li> <li>• 구축 데이터 품질관리 활동</li> <li>• 개방 데이터 품질관리 활동</li> <li>• 구축 시기별 1-Cycle 자가점검 활동</li> <li>• 제3자 품질검증 활동</li> </ul>
	구축 프로세스별 품질관리 활동	
	구축 시기별 품질 자가점검 활동	
	제3자 품질검증 활동	
	품질 관련 이해관계자	• 품질관리 주체 식별 및 이해관계자 정의

[그림 1-1] 인공지능 학습용 데이터 품질관리 프레임워크 구조

● 인공지능 학습용 데이터의 구축 및 운영단계별 품질관리 활동



[그림 1-2] 품질관리 프레임워크 프로세스별 품질관리 활동

단계	프로세스	품질관리 수행 활동
100. 준비·계획	110. 구축계획 수립	1101. 사업수행 및 구축계획 수립
		1102. 데이터 구축 절차, 조직 구성
		1103. 임무정의에 대한 적절성 검토
		1104. 품질목표 및 점검기준 수립
200. 구축	210. 데이터 획득/수집	2101. 획득/수집 방법 및 기준 현행화
		2102. 데이터 법적 근거 검토
		2103. 획득/수집 도구 및 저장환경 검토
		2104. 원시데이터 품질검사
	220. 데이터 정제	2201. 정제 방법 및 기준 현행화
		2202. 개인정보/민감정보 비식별화
		2203. 정제 도구 및 저장환경 검토
		2204. 원천데이터 품질검사
	230. 데이터 가공	2301. 라벨링 방법 및 기준 현행화
		2302. 라벨링 도구 및 저장환경 검토
		2303. 라벨링데이터 품질검사
	240. 데이터 학습	2401. 구축목적-AI모델 합치성 확인
		2402. 학습결과 확인 및 최적화
		2403. 품질검증결과 보완조치
300. 운영·활용	310. 데이터 운영·활용	3101. 데이터 하자 및 유지보수
		3102. 사용자 품질개선의견 반영

〈표 1-1〉 수행기관의 단계별 품질관리 수행 활동





## 제2장

# 품질관리 프로세스 및 산출물

1. 준비·계획단계
2. 구축단계
3. 운영·활용단계
4. 단계별 산출물

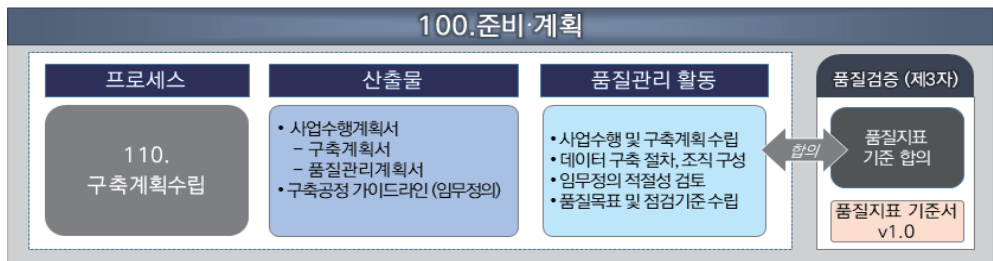




## 제2장 | 품질관리 프로세스 및 산출물

### 1 준비·계획단계(100)

#### 1.1 구축계획수립(110) 프로세스



[그림 2-1] 구축계획수립 프로세스

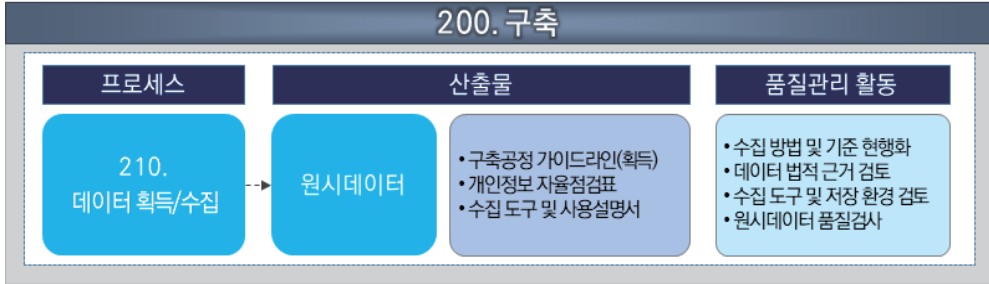
#### 1.2 구축계획수립(110) 프로세스 품질관리활동 및 산출물

프로세스	품질관리활동	주요 내용	구축 및 품질 산출물
110. 구축계획 수립	1101. 사업수행 및 구축계획 수립	<ul style="list-style-type: none"> <li>구축목적 및 범위, 수행계획 수립 검토</li> <li>품질관리방안 검토</li> <li>품질목표수준의 적정성 및 품질관리기준 정의 검토</li> </ul>	사업수행계획서 - 구축계획서 - 품질관리계획서
	1102. 데이터 구축 절차, 조직 구성	<ul style="list-style-type: none"> <li>데이터 명세 정보 및 데이터 구성 정의</li> <li>데이터 구축을 위한 절차, 조직, 도구 등 역할과 책임 수립 검토</li> <li>데이터셋 개요, 구성내역, 설계기준 및 분포 현황, 구조, 구축 과정, 검수 및 품질활동 내역 검토</li> </ul>	
	1103. 임무정의에 대한 적절성 검토	<ul style="list-style-type: none"> <li>구축목적의 일치성과 일관성 확보를 위한 계획 검토</li> </ul>	구축공정 가이드라인 (임무정의)
	1104. 품질목표 및 점검기준 수립	<ul style="list-style-type: none"> <li>사업수행기관이 제출한 사업수행계획서 등 검증 관련 자료를 토대로 품질지표기준을 수립하고, 품질검증 지표 및 목표에 대해 합의</li> <li>(사업수행기관 ↔ 제3자 품질검증기관, NIA)</li> </ul>	품질지표 기준서

〈표 2-1〉 구축계획수립 단계 품질관리활동 및 산출물

## 2 구축단계(200)

### 2.1 데이터 획득/수집(210) 프로세스



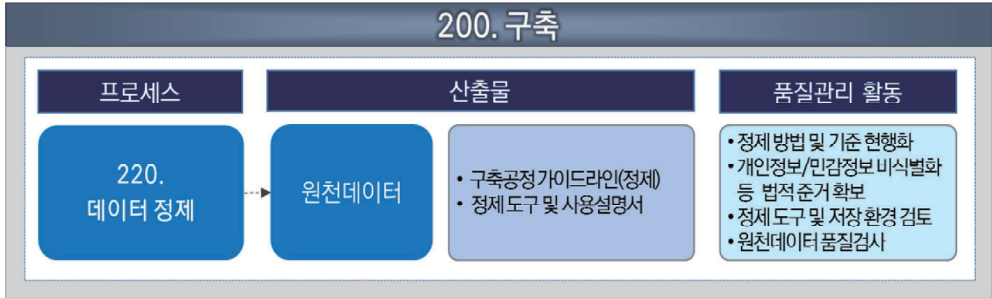
[그림 2-2] 데이터 획득/수집 프로세스

### 2.2 데이터 획득/수집(210) 프로세스 품질관리활동 및 산출물

프로세스	품질관리활동	주요 검토 내용	구축 및 품질 산출물
210. 데이터 획득/수집	2101. 획득/수집 방법 및 기준 현행화	<ul style="list-style-type: none"> <li>데이터 획득 및 확보 방법, 절차, 법 제도적 적절성</li> <li>획득항목 및 구축량 획득/수집 단계 품질 기준 및 조치, 절차, 도구</li> <li>데이터 획득 방법 및 기준, 교육, 검사에 대한 체계 수립</li> <li>데이터 획득 기준 변경 관리</li> </ul>	구축공정 가이드라인 (획득/수집)
	2102. 데이터 법적 근거 검토	<ul style="list-style-type: none"> <li>법 제도 준수(개인정보, 초상권, 저작권) 근거자료 확보</li> <li>데이터 획득/수집의 편향성 방지 방안</li> </ul>	개인정보 자율점검표
	2103. 수집 도구 및 저장환경 검토	<ul style="list-style-type: none"> <li>획득/수집에 사용될 도구 확보 및 사용 방법</li> <li>획득/수집 데이터 저장환경 구축</li> </ul>	수집 도구 수집 도구 사용설명서
	2104. 원시데이터 품질검사	<ul style="list-style-type: none"> <li>원시데이터가 인공지능을 활용하여 문제를 해결하는데 요구되는 다양성, 신뢰성, 충분성, 균일성, 사실성, 공정성 등 기준에 적합하게 구성되어 있는지 확인</li> <li>파일포맷, 해상도, 프레임 레이트 등 기술적합성 확인</li> <li>데이터에 대한 클래스 분포, 인스턴스 분포, 분석 가능 항목의 통계적 분포 확인</li> </ul>	원시데이터 품질관리결과서 품질검사 도구

〈표 2-2〉 데이터 획득/수집단계 품질관리활동 및 산출물

## 2.3 데이터 정제(220) 프로세스



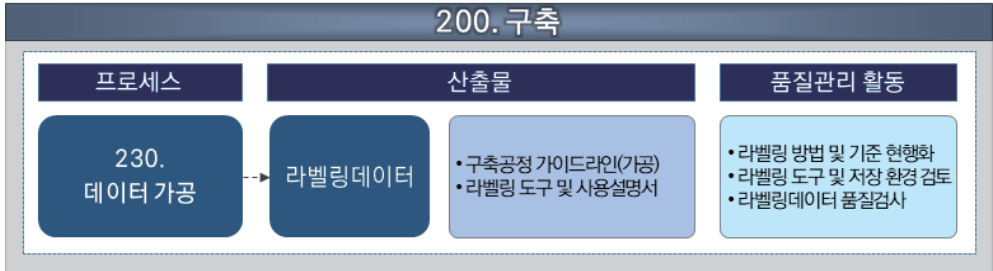
[그림 2-3] 데이터 정제 프로세스

## 2.4 데이터 정제(220) 프로세스 품질관리활동 및 산출물

프로세스	품질관리 활동	주요 검토 내용	구축 및 품질 산출물
220 데이터 정제	2201. 정제방법 및 기준 현행화	<ul style="list-style-type: none"> <li>구축목적에 적절한 데이터 선별을 위한 정제기준 수립</li> <li>정제단계 품질 기준 및 조직, 절차, 도구</li> <li>데이터 정제 방법 및 정제 결과에 대한 교육 및 훈련</li> <li>데이터 정제 결과에 대한 검사 절차와 기준 수립</li> <li>데이터 정제 기준 변경 관리</li> </ul>	구축과정 가이드라인 (정제)
	2202. 개인정보/민감정 보 비식별화	<ul style="list-style-type: none"> <li>개인정보보호 등 비식별화 기준 및 방안</li> </ul>	
	2203. 정제 도구 및 저장환경 검토	<ul style="list-style-type: none"> <li>정제단계에 사용될 저작도구 확보 및 사용방법</li> <li>정제 데이터 저장환경 구축</li> </ul>	저작도구(정제)  사용설명서
	2204. 원천데이터 품질검사	<ul style="list-style-type: none"> <li>사용하기 적합한 데이터 여부 품질검사</li> <li>원천데이터 품질검사를 위한 검사 도구 및 검사결과</li> <li>품질검사결과에 따른 보완조치</li> </ul>	원천데이터  품질관리결과서  품질검사 도구

〈표 2-3〉 데이터 정제단계 품질관리활동 및 산출물

## 2.5 데이터 가공(230) 프로세스



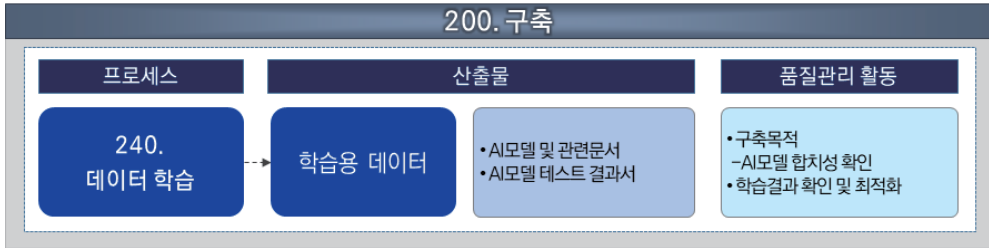
[그림 2-4] 데이터 가공 프로세스

## 2.6 데이터 가공(230) 프로세스 품질관리활동 및 산출물

프로세스	품질관리 활동	주요 검토 내용	구축 및 품질 산출물
230. 데이터 가공	2301. 라벨링 방법 및 기준 현행화	<ul style="list-style-type: none"> <li>• 라벨링 방법 및 기준 정의</li> <li>• 학습에 필요한 라벨링 작업 및 어노테이션 설정</li> <li>• 데이터 형식 및 입력값 범위 기준 마련</li> <li>• 데이터 유형별 객체 구분 작업</li> <li>• 데이터 가공 방법 및 기준, 교육, 검사체계 수립</li> <li>• 데이터 가공 기준변경 관리</li> </ul>	구축공정 가이드라인 (가공)
	2302. 라벨링 도구 및 저장환경 검토	<ul style="list-style-type: none"> <li>• 가공에 사용될 저작도구 확보 및 사용 방법</li> <li>• 데이터 가공 기준변경 관리</li> <li>• 라벨링데이터 저장환경 구축</li> </ul>	저작도구(가공)
	2303. 라벨링데이터 품질검사	<ul style="list-style-type: none"> <li>• 정확성(의미 정확성, 구문 정확성)을 중점으로 검사</li> <li>• 라벨링데이터 품질검사를 위한 검사 도구 및 검사 결과</li> <li>• 품질검사결과에 따른 보완 조치</li> </ul>	사용설명서
			라벨링데이터
			품질관리결과서
			품질검사 도구

〈표 2-4〉 데이터 가공단계 품질관리활동 및 산출물

## 2.7 데이터 학습(240) 프로세스



[그림 2-5] 데이터 학습 프로세스

## 2.8 데이터 학습(240) 프로세스 품질관리활동 및 산출물

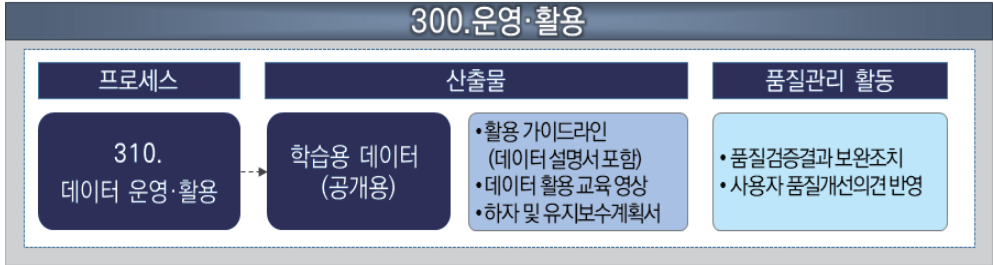
프로세스	품질관리활동	주요 검토 내용	구축 및 품질 산출물
240. 데이터 학습	2401. 구축목적-AI 모델 합치성 확인	<ul style="list-style-type: none"> <li>데이터를 인공지능 알고리즘으로 훈련시켰을 때 영역별 품질검사 기준 확인</li> <li>구축 목적에 부합하는 AI 모델 선정 확인</li> <li>AI 모델 구축/운영 및 테스트를 위한 환경 구축 정보</li> <li>AI 모델 정보/라이선스 가이드</li> </ul>	AI모델 설명서
	2402. 학습결과 확인 및 최적화	<ul style="list-style-type: none"> <li>학습용 데이터가 AI 알고리즘으로 훈련시켰을 때 목표로 했던 수준의 성능 달성이 가능한지 검사</li> <li>알고리즘 성능 검사</li> <li>모델 단계 유효성 검사</li> <li>학습모델, 활용 서비스 개발에 가능한 학습데이터셋 품질오류 및 이슈 내용을 분석하여 보완조치 반영</li> </ul>	AI모델 소스코드
			학습모델 파일
			모델 구축 도커 이미지
			AI모델 테스트 결과서
			품질검사 도구
	2403. 품질검증결과 보완조치	<ul style="list-style-type: none"> <li>최종 데이터 제3자 품질검증결과 미달성인 경우, 보완조치하여 결과 보고</li> </ul>	보완조치결과서

〈표 2-5〉 데이터 학습단계 품질관리활동 및 산출물

## 2.9 구축단계 품질검사

### 3 운영·활용단계(300)

#### 3.1 데이터 운영·활용(310) 프로세스



[그림 2-8] 데이터 운영·활용 프로세스

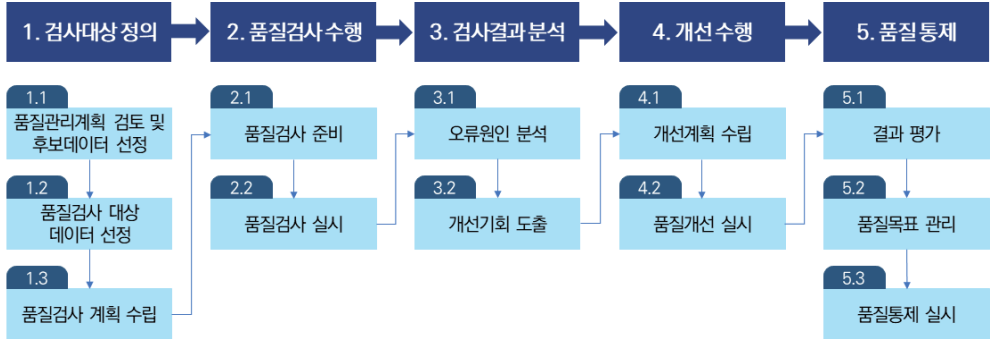
#### 3.2 데이터 운영·활용(310) 프로세스 품질관리 활동 및 산출물

프로세스	품질관리 활동	주요 검토 내용	구축 및 품질 산출물
310. 데이터 운영·활용	3101. 데이터 하자 및 유지보수	<ul style="list-style-type: none"> <li>• 시하브에 공개된 데이터에 대해 하자보수 및 유지보수</li> </ul>	하자 및 유지보수계획서
	3102. 사용자 품질개선의견 반영	<ul style="list-style-type: none"> <li>• 사용자들의 개발 및 활용 지원을 위한 산출물 공개</li> <li>• 데이터 품질에 대한 하자 및 유지보수 계획</li> <li>• 사용자들의 품질에 대한 개선 의견이 있는 경우 반영 등 유지보수</li> </ul>	활용 가이드라인 (데이터 설명서 포함)  데이터 활용 교육 영상

〈표 2-6〉 데이터 운영·활용단계 품질관리활동 및 산출물

### 3.3 운영·활용단계 품질검사 절차

- 인공지능 학습데이터 품질검사는 검사대상 정의(3), 품질검사 실시(2), 검사결과 분석(2), 개선수행(2), 품질 통제(3)의 5단계, 12개 상세 수행 절차로 구성



[그림 2-9] 운영·활용 단계 품질검사 상세 수행 절차

구 분	내 용
1. 검사대상 정의 (Define)	<ul style="list-style-type: none"> <li>원시데이터, 원천데이터, 데이터셋, 라벨링 데이터 등의 검사대상을 선정하고 해당 데이터가 요구되는 품질수준에 부합하는 상태인지를 판단하기 위한 품질검사 계획을 수립</li> </ul>
2. 품질검사 수행 (Measure)	<ul style="list-style-type: none"> <li>정의된 검사대상에 대해 적합성, 정확성, 유효성, 준비성, 완전성, 유용성 등의 품질관리 지표에 대해 체크리스트와 같은 검사기법을 적용하여 품질검사를 실시</li> </ul>
3. 검사결과 분석 (Analyze)	<ul style="list-style-type: none"> <li>품질검사 결과를 바탕으로 주요 품질문제를 식별하고 문제의 근본적 원인을 파악하여 품질문제를 해결하기 위한 개선 기회를 도출</li> </ul>
4. 개선수행 (Request & Implementation)	<ul style="list-style-type: none"> <li>품질개선을 위한 개선 방향을 정의하고 개선 우선순위를 결정하며, 개선 수행을 위한 구체적인 품질개선 계획을 수립한 후 수립된 개선계획에 따라 데이터 보정, 추가 작업 등 개선 영역별 품질활동을 요청</li> </ul>
5. 품질 통제 (Control)	<ul style="list-style-type: none"> <li>개선수행 결과의 확인 및 점검을 통해 품질 목표를 재설정하여 품질문제 재발 방지 및 고품질 데이터를 유지하기 위한 품질관리 활동을 수행</li> </ul>

〈표 2-7〉 운영·활용 단계 품질검사 수행 절차 정의



번호	상세 수행 절차	주요 활동
1.1	품질관리계획 검토 및 후보데이터 선정	<ul style="list-style-type: none"> <li>품질관리 담당자는 구축사업의 품질관리계획과 관련된 자료를 수집하고 데이터 구축 및 인공지능 학습모델 담당자와 협의를 통해 변경된 요구사항, 업무절차 및 작업방법 등의 부가정보를 획득</li> </ul>
1.2	품질검사 대상 데이터 선정	<ul style="list-style-type: none"> <li>품질관리 계획 검토 및 후보 데이터 선정을 식별한 후보 데이터 중에서 품질검사의 중요성, 시급성 관점으로 검사대상 데이터를 선정</li> </ul>
1.3	품질검사 계획 수립	<ul style="list-style-type: none"> <li>품질검사 대상 데이터 선정이 완료되면 선정한 데이터를 대상으로 품질검사계획서를 작성하는 단계</li> </ul>
2.1/ 2.2	품질검사 준비 / 품질검사 실시	<ul style="list-style-type: none"> <li>품질검사 준비는 본격적인 품질검사 수행에 앞서 품질검사를 효율적으로 수행하기 위해 수행기관 및 참여기관에서 사전에 준비해야 할 검사 환경을 준비하는 단계</li> </ul>
3.1	오류원인 분석	<ul style="list-style-type: none"> <li>오류 원인을 추적하기 위해서는 선행적으로 품질검사 수행단계에서 조사된 추정 오류를 대상으로 실제로 오류 여부를 분류하고, 설사 오류라고 할지라도 임무와 데이터 특성을 고려하여 분석 대상 포함 여부를 확인</li> </ul>
3.2	개선기회 도출	<ul style="list-style-type: none"> <li>도출된 구체적인 개선기회는 향후 “개선 수행(Implement)” 단계의 실제적인 활동 내용으로 연결되어야 함</li> </ul>
4.1	개선계획 수립	<ul style="list-style-type: none"> <li>검사결과 분석 단계에서 도출된 개선기회를 종합하여 개선과제를 도출하고, 분류된 개선 과제별로 상세 추진계획을 수립</li> </ul>
4.2	품질개선 실시	<ul style="list-style-type: none"> <li>수립된 개선계획 중 단기 과제를 우선으로, 개선수행을 위한 환경을 준비하고, 실제 개선을 수행</li> </ul>
5.1	결과 평가	<ul style="list-style-type: none"> <li>품질 관리체계 수립, 데이터 운영계획 수립, 데이터 보정으로 구분</li> <li>구축 가이드라인만 개정 한 영역과 실제 개선을 한 영역을 종합적으로 표현, 실제 개선한 영역에 개선 효과를 정량적으로 평가</li> </ul>
5.2	품질목표 관리	<ul style="list-style-type: none"> <li>품질통제 대상에 대한 품질 목표를 설정</li> <li>품질을 유지하거나 더 높은 품질을 확보할 수 있는 방안을 수립</li> </ul>
5.3	품질통제 실시	<ul style="list-style-type: none"> <li>품질목표 관리 대상에 대해 수시로 모니터링을 수행</li> <li>수행기관 및 참여기관 구성원들에 대한 교육, 홍보 활동으로 품질목표 관리활동에 대한 공감대 형성</li> <li>적극적인 참여를 유도과 소통 절대적으로 필요</li> </ul>

〈표 2-8〉 운영·활용 단계 품질검사 상세 수행 절차 주요 활동

## 4 단계별 산출물

### 4.1 구축 및 품질 산출물

※ 제출여부 - Y : 필수 제출 산출물, C : 선택적 제출 산출물

구분	산출물명	유형	필수 여부	AI 허브 공개산출물
구축계획 수립	사업수행계획서 (※ 구축계획서, 품질관리계획서 포함)	문서	Y	-
데이터 획득/수집, 정제, 가공	개인정보 자율점검표	문서	Y	-
	구축공정 가이드라인	문서	Y	-
	저작도구(Viewer) 사용설명서	문서	Y	Y
	저작도구/Viewer	어플리케이션	Y	Y
	원시데이터(획득/수집데이터)	데이터	C	-
	원천데이터(이미지, 영상, 음성 등)	데이터	Y	-
	라벨링데이터(JSON, XML, CSV 등)	데이터	Y	-
데이터 학습	AI모델 설명서	문서	Y	Y
	AI모델 소스코드	Code	Y	Y
	학습모델 파일	binary	Y	Y
	모델 구축 도커이미지	tar	Y	Y
	AI모델 테스트 결과서	문서	C	-
데이터 품질	품질관리결과서	문서	Y	-
	품질검증결과서	문서	Y	-
	품질지표 기준서(사업수행계획서 내 포함)	문서	Y	-
데이터 운영·활용	활용 가이드라인	문서	Y	Y
	데이터 활용 교육 영상	동영상	Y	Y
	하자 및 유지보수 계획서	문서	Y	-

〈표 2-9〉 구축 및 품질 관련 산출물 목록

## 4.2 구축단계별 산출물

※ 범례 - \* 필수 제출 산출물, \*\* 필수(공개) 제출 산출물

구분	구축계획 수립	데이터 획득/수집	데이터 정제	데이터 가공	데이터 학습	데이터 운영·활용
문서	사업수행계획서 (별첨포함)*					사업수행계획서 (별첨포함)*
	구축공정 가이드라인 (임무정의)*	구축공정 가이드라인 (획득)*	구축공정 가이드라인 (정제)*	구축공정 가이드라인 (가공)*		
						활용 가이드라인**
		획득 저작도구 사용설명서	정제 저작도구 사용설명서**	라벨링 저작도구 사용설명서**		
		품질관리결과서*	품질관리결과서*	품질관리결과서*	품질관리결과서*	
		개인정보 자율점검표*				
					시모델 설명서**	
					시모델 테스트 결과서	
						하자 및 유지보수 계획서*
동영상						데이터 활용 교육 영상**
데이터		원시데이터 (획득/수집 데이터)	원천데이터 (이미지, 영상, 음성 등)**	라벨링데이터 (JSON, XML, CSV 등)**		
어플리케이션		저작도구	저작도구**	저작도구**		
기타					시모델 소스코드**	
					학습모델 파일**	
					모델 구축 도커이미지**	

〈표 2-10〉 구축단계별 산출물



## 제3장

# 품질 자가점검 및 품질검증

1. 품질 자가점검
2. 제3자 품질검증



## 제3장 | 품질 자가점검 및 품질검증

### 1 품질 자가점검

#### 1.1 품질자가점검

- 인공지능 학습용 데이터 구축사업 진행 시 구축공정별 품질 이슈를 최소화하기 위해 사업수행기관의 초기데이터, 중간데이터 품질 자가점검에 대한 품질 자가점검을 필수적으로 수행하고, 그 외 품질 자가점검은 사업규모, 사업범위, 일정 등에 따라 선택적으로 수행

##### 1.1.1 품질 자가점검 프로세스

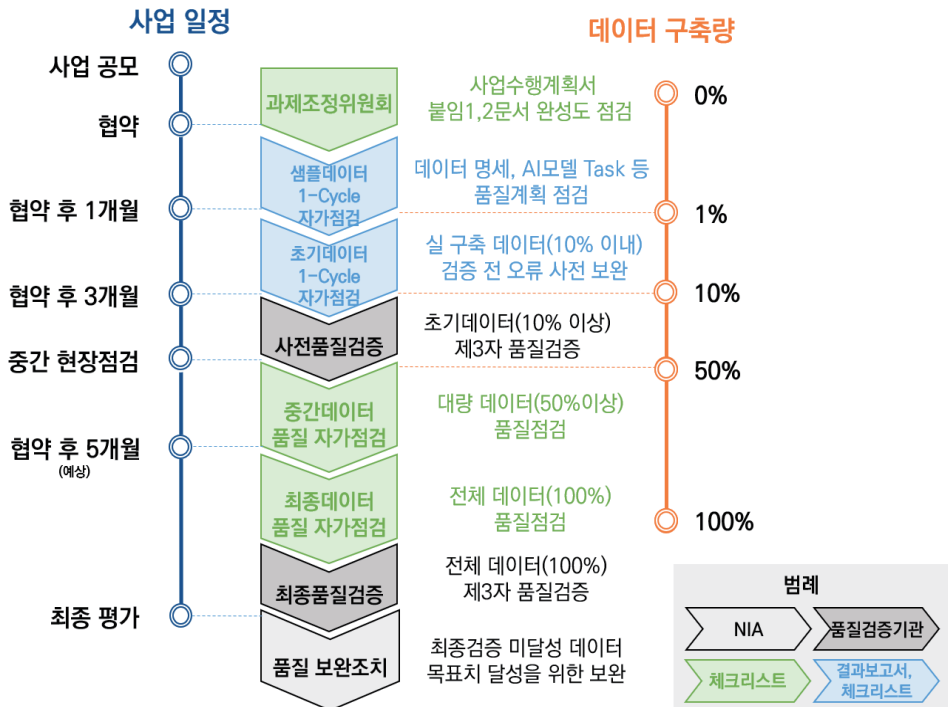


[그림 3-1] 품질 자가점검 프로세스

### 1.1.2 품질자가점검 프로세스 주요 내용

사업일정	품질 자가점검 활동	주요 검토내용
과제조정위원회	과제조정위원회 자가점검	<ul style="list-style-type: none"> <li>사업수행계획서 내 구축계획서, 품질관리계획서의 적정성 검토</li> </ul>
사업 1개월 시점	샘플데이터 1-Cycle 자가점검	<ul style="list-style-type: none"> <li>데이터 설계의 적정성 점검 및 데이터 구축 목적과 AI 학습모델 후보군의 합치성 및 준비도 점검</li> </ul>
사업 3개월 시점	초기데이터 1-Cycle 자가점검	<ul style="list-style-type: none"> <li>측정지표의 적정성 및 정량목표의 달성도 확인</li> <li>데이터 형태별 오류유형에 따른 보완조치 방안 마련</li> </ul>
중간시점	중간데이터 품질 자가점검	<ul style="list-style-type: none"> <li>초기데이터 1-Cycle 결과 반영 여부 검토</li> <li>대량 데이터(50% 이상) 기반의 유효성 점검</li> </ul>
최종시점	최종데이터 품질 자가점검	<ul style="list-style-type: none"> <li>최종 데이터 품질 전반 점검</li> </ul>

〈표 3-1〉 품질 자가점검 활동의 주요 검토내용

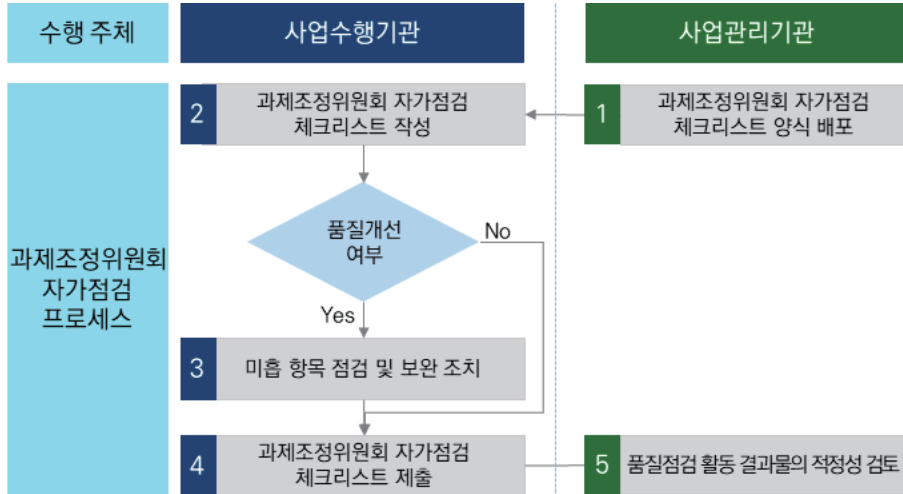


[그림 3-2] 사업 일정에 따른 품질 자가점검 활동



## 1.2 품질자가점검 세부 활동 정의

### 1.2.1 과제조정위원회(과조위) 자가점검 프로세스

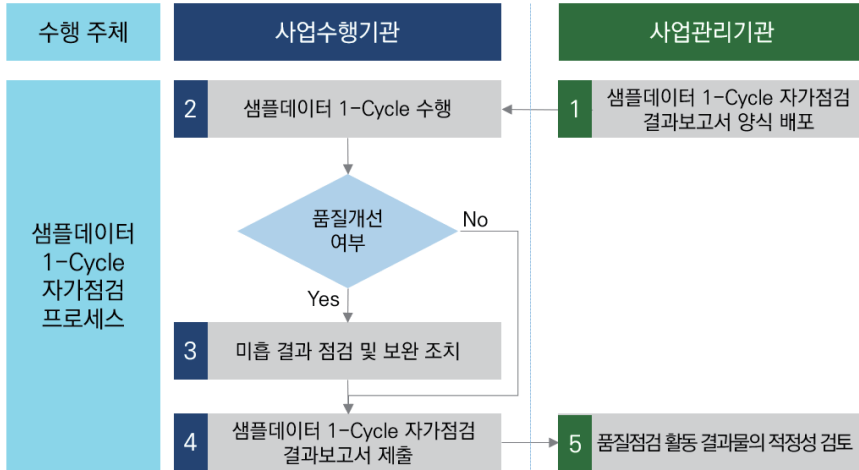


[그림 3-3] 과제조정위원회(과조위) 자가점검 프로세스

번호	활동	주체	주요 내용	산출물
1	과제조정위원회 자가점검 체크리스트 양식 배포	사업관리기관	<ul style="list-style-type: none"> <li>사업수행계획서 내 구축계획서 및 품질계획서 자가점검 체크리스트 양식 배포</li> <li>우선협상대상자에게 전달하여 작성요청</li> </ul>	과제조정위원회 자가점검 체크리스트 양식
2	과제조정위원회 체크리스트 작성	사업수행기관	<ul style="list-style-type: none"> <li>사업수행기관이 자가점검 체크리스트 작성</li> </ul>	과제조정위원회 체크리스트
3	미흡항목 점검 및 보완	사업수행기관	<ul style="list-style-type: none"> <li>자가점검 체크리스트 점검항목 중 미흡 항목에 대해 점검 및 보완</li> </ul>	과제조정위원회 체크리스트, 사업수행계획서
4	과제조정위원회 체크리스트 제출	사업수행기관	<ul style="list-style-type: none"> <li>최종 작성된 자가점검 체크리스트 제출</li> </ul>	과제조정위원회 체크리스트
5	품질점검 활동 결과물의 적정성 검토	사업관리기관	<ul style="list-style-type: none"> <li>제출된 자가점검 체크리스트는 과조위 위원들과 공유, 과조위 개최 시 점검 및 활용</li> <li>최종 과제조정위원회 종료 후, 유닛별 품질담당자에게 최종본 회신</li> </ul>	과제조정위원회 검토의견서, 건설팅보고서

〈표 3-2〉 과제조정위원회 자가점검 품질관리 활동 및 산출물

## 1.2.2 샘플데이터 1-Cycle 자가점검 프로세스

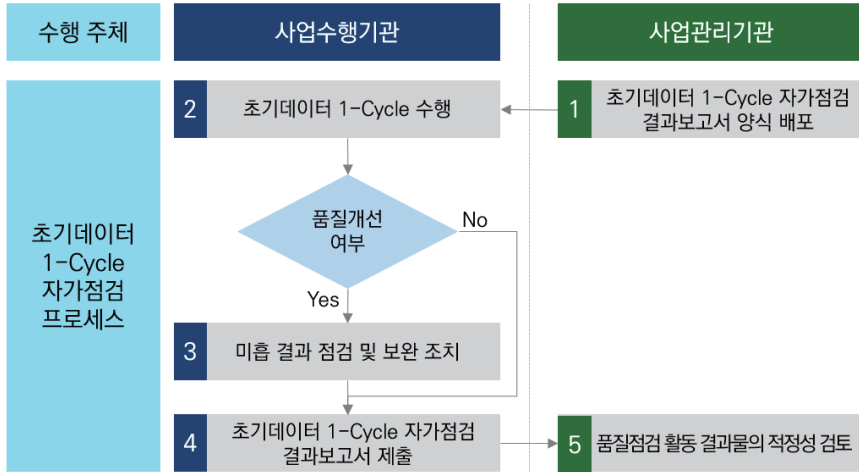


[그림 3-4] 샘플데이터 1-Cycle 자가점검 프로세스

번호	활동	주체	주요 내용	산출물
1	샘플데이터 1-Cycle 자가점검 결과보고서 양식 배포	사업관리기관	<ul style="list-style-type: none"> <li>최소 목표 수량(1% 이내)으로 샘플데이터의 획득/수집, 정제, 가공, AI 모델 적용방안 검토 및 데이터 명세 확정을 위한 자가점검 결과보고서 양식 배포</li> </ul>	샘플데이터 1-Cycle 자가점검 결과보고서 양식
2	샘플데이터 1-Cycle 수행	사업수행기관	<ul style="list-style-type: none"> <li>샘플데이터 1-Cycle 수행</li> <li>자가점검 결과보고서 작성</li> </ul>	샘플데이터 1-Cycle 결과보고서, 체크리스트
3	미흡 결과 점검 및 보완	사업수행기관	<ul style="list-style-type: none"> <li>샘플데이터 1-Cycle 결과에 따라 보완</li> <li>단계별 구축공정 가이드라인에 반영</li> </ul>	보완된 결과보고서 등 산출물
4	샘플데이터 1-Cycle 자가점검 결과보고서 제출	사업수행기관	<ul style="list-style-type: none"> <li>최종 작성된 자가점검 결과보고서 제출</li> </ul>	샘플데이터 1-Cycle 결과보고서, 체크리스트
5	품질점검 활동 결과물의 적정성 검토	사업관리기관	<ul style="list-style-type: none"> <li>사업담당자, 유닛별 품질담당자 등 공유</li> <li>품질점검 활동 결과 분석 및 적정성 검토</li> </ul>	컨설팅보고서

<표 3-3> 샘플데이터 1-Cycle 자가점검 품질관리 활동 및 산출물

### 1.2.3 초기데이터 1-Cycle 자가점검 프로세스

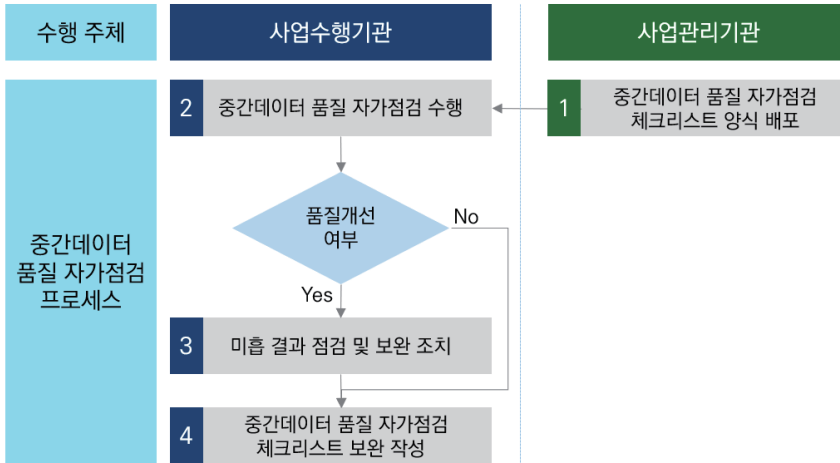


[그림 3-5] 초기데이터 1-Cycle 자가점검 프로세스

번호	활동	주체	주요 내용	산출물
1	초기데이터 1-Cycle 자가점검 결과보고서 양식 배포	사업관리기관	<ul style="list-style-type: none"> <li>초기데이터 수량(10% 이상)으로 품질 정량 목표에 부합하는지에 대한 자가점검 결과보고서 양식 배포</li> </ul>	초기데이터 1-Cycle 자가점검 결과보고서 양식
2	초기데이터 1-Cycle 수행	사업수행기관	<ul style="list-style-type: none"> <li>초기데이터 1-Cycle 수행</li> <li>측정지표의 적정성 및 정량 목표의 달성도 점검</li> <li>자가점검 결과보고서 작성</li> </ul>	초기데이터 1-Cycle 결과보고서, 체크리스트
3	미흡 결과 점검 및 보완	사업수행기관	<ul style="list-style-type: none"> <li>샘플데이터 1-Cycle 결과 반영여부 점검</li> <li>데이터 오류 유형에 따른 보완 방안 마련</li> <li>단계별 구축공정 가이드라인에 반영</li> </ul>	초기데이터, 보완된 결과보고서 등 산출물
4	초기데이터 1-Cycle 자가점검 결과보고서 제출	사업수행기관	<ul style="list-style-type: none"> <li>최종 작성된 자가점검 결과보고서 제출</li> </ul>	초기데이터 1-Cycle 결과보고서, 체크리스트
5	품질점검 활동 결과물의 적정성 검토	사업관리기관	<ul style="list-style-type: none"> <li>사업담당자, 유닛별 품질담당자 등 공유</li> <li>품질점검 활동 결과 분석 및 적정성 검토</li> </ul>	컨설팅보고서

〈표 3-4〉 초기데이터 1-Cycle 자가점검 품질관리 활동 및 산출물

## 1.2.4 중간데이터 품질 자가점검 프로세스

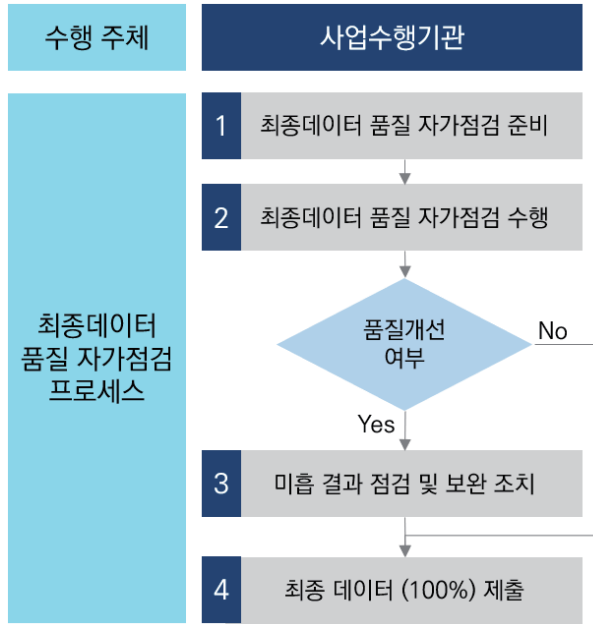


[그림 3-6] 중간데이터 품질 자가점검 프로세스

번호	활동	주체	주요 내용	산출물
1	중간데이터 품질 자가점검 체크리스트 양식 배포	사업관리기관	<ul style="list-style-type: none"> <li>대량 데이터(50% 이상)에 대한 품질 자가점검을 위한 체크리스트 양식 배포</li> </ul>	중간데이터 품질 자가점검 체크리스트 양식
2	중간데이터 품질 자가점검 수행	사업수행기관	<ul style="list-style-type: none"> <li>초기데이터 1-Cycle 결과 반영 여부 점검</li> <li>대량 데이터(50% 이상) 기반의 정확성 및 유효성 점검(대량 데이터에서 다양성 기반 균등 샘플링 후 점검)</li> </ul>	중간데이터 품질 자가점검 체크리스트
3	미흡 결과 점검 및 보완	사업수행기관	<ul style="list-style-type: none"> <li>중간데이터 품질점검결과 보완 조치</li> <li>단계별 구축공정 가이드라인에 반영</li> </ul>	산출물 보완
4	중간데이터 품질 자가점검 체크리스트 보완 작성	사업수행기관	<ul style="list-style-type: none"> <li>자가점검 체크리스트 재검토 및 보완</li> </ul>	중간데이터 품질 자가점검 체크리스트

〈표 3-5〉 중간데이터 품질 자가점검 품질관리활동 및 산출물

### 1.2.5 최종데이터 품질 자가점검 프로세스



[그림 3-7] 최종데이터 품질 자가점검 프로세스

번호	활동	주체	주요 내용	산출물
1	최종데이터 품질 자가점검 준비	사업수행기관	<ul style="list-style-type: none"> <li>• 전체 데이터(100%) 구축</li> <li>• 사업수행계획서, 품질지표 기준서 v2.0, 품질검사 도구 등 준비</li> </ul>	전체 데이터, 최종 산출물 등
2	최종데이터 품질 자가점검 실행	사업수행기관	<ul style="list-style-type: none"> <li>• 구축완료 데이터(100%)로 사업수행계획서 전 공정 결과값이 품질기준에 부합하는지 여부 점검</li> </ul>	체크리스트
3	미흡 결과 점검 및 보완	사업수행기관	<ul style="list-style-type: none"> <li>• 최종 품질점검결과 보완 조치</li> <li>• 오류 원인 및 유형 분석 후 개선 조치</li> </ul>	산출물 보완
4	최종 데이터 제출	사업수행기관	<ul style="list-style-type: none"> <li>• 최종 데이터를 사업관리기관에 제출</li> <li>• 제3자 품질검증을 위한 데이터 제출</li> </ul>	학습용 데이터 100%

〈표 3-6〉 최종데이터 품질 자가점검 품질관리 활동 및 산출물

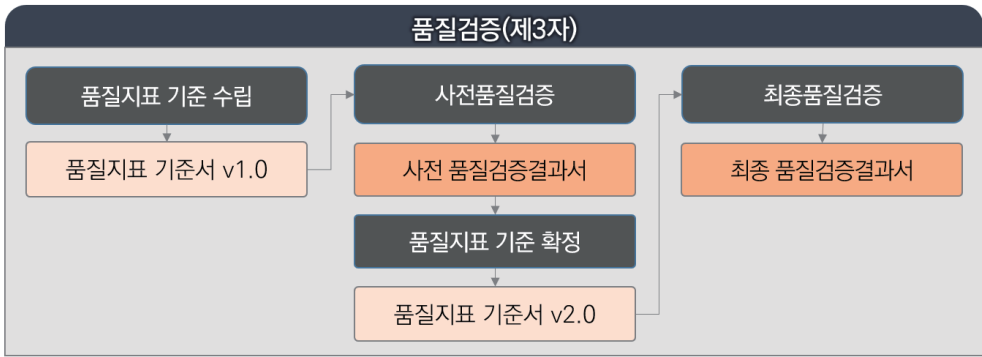
## 2 제3자 품질검증

### 2.1 제3자 품질검증

- 인공지능 학습용 데이터 개별 특성에 따른 객관적 품질지표 기준을 마련하고, 지표 기준에 따라 제3자 품질검증을 수행하여 구축 데이터의 품질목표 달성여부 측정

※ 제3자 품질검증 : 인공지능 학습용 데이터의 품질관리 기준, 절차 및 방법을 이해하고 있는 기관 또는 업체를 통해 사업수행기관이 구축한 인공지능 학습용 데이터 품질을 검증하는 활동

#### 2.1.1 제3자 품질검증 프로세스



[그림 3-8] 제3자 품질검증 프로세스

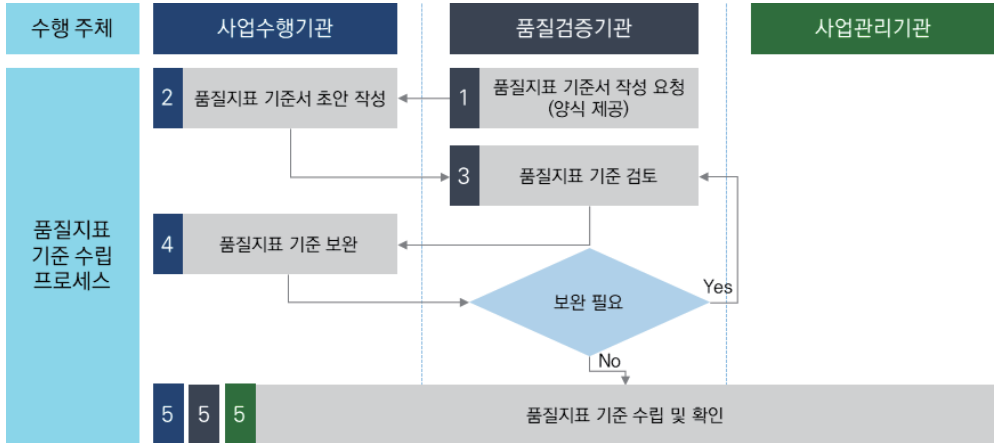
#### 2.1.2 제3자 품질검증 프로세스 주요 내용

활동	주요 내용	산출물
품질지표 기준 수립	<ul style="list-style-type: none"> <li>사업수행기관이 제출한 사업수행계획서를 토대로 사전 품질검증을 위한 품질지표 기준 수립</li> </ul>	품질지표 기준서 v1.0
사전품질검증	<ul style="list-style-type: none"> <li>구축 중인 학습용 데이터셋의 오류를 사전에 검출하여 수정하고 예방</li> <li>검증 대상 : 다양성, 구문 정확성, 의미 정확성</li> </ul>	사전 품질검증 결과서
품질지표 기준 확정	<ul style="list-style-type: none"> <li>사전 품질검증 결과를 토대로 이슈 및 변경 사항 발생 시 협의하여 품질지표 기준서에 반영</li> <li>최종품질검증을 위한 품질지표 기준 최종 확정</li> </ul>	품질지표 기준서 v2.0
최종품질검증	<ul style="list-style-type: none"> <li>최종 구축된 학습용 데이터셋을 대상으로 품질목표 달성 여부 측정</li> <li>검증 대상 : 다양성, 구문 정확성, 의미 정확성, 유효성</li> </ul>	품질검증 결과서

<표 3-7> 제3자 품질검증 프로세스 주요 내용

## 2.2 제3자 품질검증 세부 활동 정의

### 2.2.1 품질지표기준 합의 프로세스



[그림 3-9] 품질지표 기준 수립 프로세스

번호	활동	주체	주요 내용	산출물
1	품질지표 기준서 작성 요청	품질검증기관	• 품질지표 기준서 양식 및 안내자료 배포	• 품질지표 기준서 양식 • 품질지표 기준서 작성 안내자료
2	품질지표 기준서 초안 작성	사업수행기관	• 사업수행계획서를 토대로 품질지표 기준서 초안 작성 • 항목별 측정조건 작성 • 구문 정확성 검사규칙 작성 • 의미 정확성 검사기준서 작성	• 품질지표 기준서(초안) • 항목별 측정조건 • 구문 정확성 검사 규칙 • 의미 정확성 검사 기준서
3	품질지표 기준 검토	품질검증기관	• 품질지표 기준 제반문서 검토	-
4	품질지표 기준 보완	사업수행기관	• 검토내용을 토대로 품질지표 기준 수정 및 보완	• 품질지표 기준서(수정) • 항목별 측정조건 • 구문 정확성 검사규칙 • 의미 정확성 검사기준서
5	품질지표 기준 수립 및 확인	사업수행기관 품질검증기관 사업관리기관	• 최종 보완된 품질지표 및 목표에 대한 합의 • 합의된 품질지표 기준 확인	• 품질지표 기준서 v1.0

〈표 3-8〉 품질지표 기준 수립 활동 및 산출물

## 2.2.2 사전품질검증 프로세스



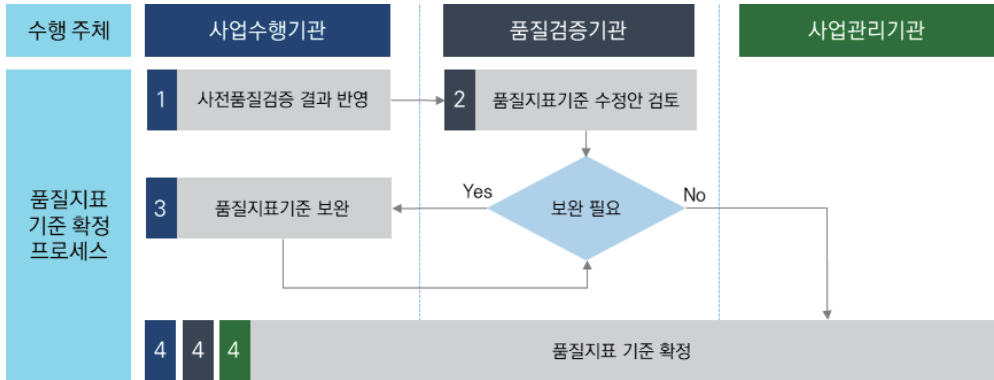
[그림 3-10] 사전품질검증 프로세스

번호	활동	주체	주요 내용	산출물
1	사전품질검증 안내	품질검증기관	<ul style="list-style-type: none"> <li>품질검증 안내자료 배포</li> <li>품질검증 교육 제공</li> </ul>	<ul style="list-style-type: none"> <li>품질검증안내자료</li> <li>품질검증 양식</li> </ul>
		사업관리기관	<ul style="list-style-type: none"> <li>초기데이터(10% 이상) 제출 안내</li> </ul>	<ul style="list-style-type: none"> <li>요청 공문</li> </ul>
2	사전품질검증 준비	품질검증기관	<ul style="list-style-type: none"> <li>품질지표 기준서 검토</li> </ul>	<ul style="list-style-type: none"> <li>초기데이터</li> </ul>
		사업수행기관	<ul style="list-style-type: none"> <li>데이터셋 제출 (초기데이터 10% 이상)</li> <li>품질검증 참고자료 산출물 제출</li> <li>검증 준비 현황 점검</li> </ul>	
3	사전품질검증 의뢰	사업수행기관	<ul style="list-style-type: none"> <li>온라인 검증의뢰</li> </ul>	-
4	사전품질검증 수행	품질검증기관	<ul style="list-style-type: none"> <li>사전품질검증 수행 및 검증 결과 작성</li> </ul>	사전품질검증 결과서
5	사전품질검증 결과 검토	사업수행기관 품질검증기관	<ul style="list-style-type: none"> <li>검증 결과 상호검토 및 합의</li> </ul>	-
6	사전품질검증 결과 공유	품질검증기관	<ul style="list-style-type: none"> <li>사전품질검증 결과를 사업관리기관 및 사업수행기관에 공유</li> </ul>	-
7	사전품질검증 결과 확인	사업수행기관 사업관리기관	<ul style="list-style-type: none"> <li>사전품질검증 결과 확인</li> <li>※ 검증 결과 오류는 반드시 수정 · 반영</li> </ul>	사업관리대장

〈표 3-9〉 사전품질검증 활동 및 산출물



## 2.2.3 품질지표 기준 확정 프로세스



[그림 3-11] 품질지표 기준 확정 프로세스

번호	활동	주체	주요 내용	산출물
1	사전품질검증 결과 반영	사업수행기관	<ul style="list-style-type: none"> <li>사전품질검증 결과를 품질지표 기준서 v1.0에 반영</li> <li>항목별 측정조건 변경 반영</li> <li>구문 정확성 검사규칙 변경 반영</li> <li>의미 정확성 검사기준서 변경 반영</li> </ul>	<ul style="list-style-type: none"> <li>품질지표 기준서 v1.0 (수정)</li> <li>항목별 측정조건</li> <li>구문 정확성 검사 규칙</li> <li>의미 정확성 검사 기준서</li> </ul>
2	품질지표 기준 수정안 검토	품질검증기관	<ul style="list-style-type: none"> <li>사업수행기관이 수정한 품질지표 기준 자료 검토</li> </ul>	-
3	품질지표 기준 보완	사업수행기관	<ul style="list-style-type: none"> <li>검토내용을 토대로 품질지표 기준 수정 및 보완</li> </ul>	<ul style="list-style-type: none"> <li>품질지표 기준서 v1.0 (수정)</li> <li>항목별 측정조건</li> <li>구문 정확성 검사 규칙</li> <li>의미 정확성 검사 기준서</li> </ul>
4	품질지표 기준 확정	사업수행기관 품질검증기관 사업관리기관	<ul style="list-style-type: none"> <li>최종 보완된 내용으로 품질지표 및 목표에 대한 합의</li> <li>합의된 품질지표기준 확인</li> </ul>	<ul style="list-style-type: none"> <li>품질지표 기준서 v2.0</li> </ul>

〈표 3-10〉 품질지표 기준 확정 활동 및 산출물

## 2.2.4 최종품질검증 프로세스



[그림 3-12] 최종품질검증 프로세스

번호	활동	주체	주요 내용	산출물
1	품질검증안내	품질검증기관	• 품질검증 안내자료 배포	• 품질검증안내자료
		사업관리기관	• 최종데이터(100%) 제출 안내	• 요청 공문
2	품질검증 준비	품질검증기관	• 품질지표 기준서 검토	• 최종데이터 • 유효성 증빙자료
		사업수행기관	• 데이터셋 제출(최종데이터 100%) • 품질검증 참고자료 산출물 제출 • 검증 준비 현황 점검 • 유효성 증빙자료 제출	
3	품질검증 의뢰	사업수행기관	• 온라인 검증의뢰	-
4	품질검증 수행	품질검증기관	• 품질검증 수행 및 검증 결과 작성	• 품질특성별 검증결과
5	검증결과 검토	품질검증기관 사업수행기관	• 검증결과 상호검토 및 합의	-
6	품질검증결과서 발급	품질검증기관	• 사업관리기관 및 사업수행기관에 품질검증 결과서 발급	• 품질검증결과서
7	품질검증결과 확인	사업수행기관 사업관리기관	• 최종 품질검증결과 확인	• 사업관리대장

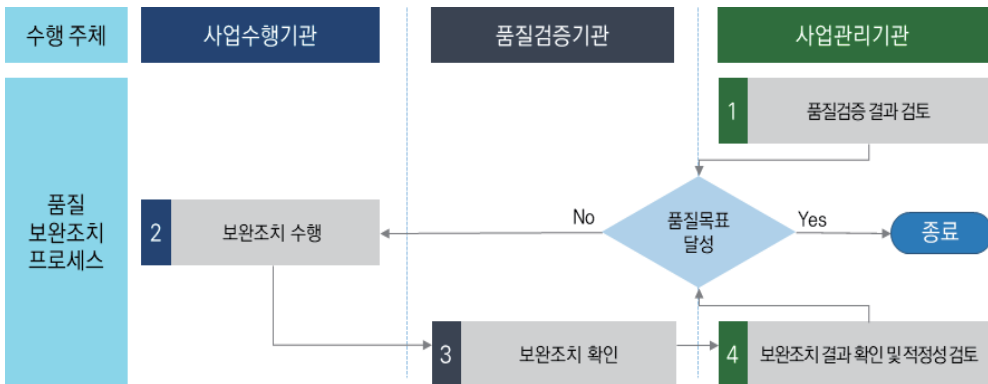
〈표 3-11〉 최종품질검증 활동 및 산출물

● 품질검증항목

번호	품질지표	주요 내용	기준
1	다양성(통계)	• 관심객체, 카테고리, 수집환경 등 인공지능 학습용 데이터의 주요특성을 통계적 방법으로 확인	충분성, 균등성, 편향성 여부 확인
2	다양성(요건)	• 사업수행기관의 구축목표 대비 구축결과물 비율(수량)이 충족하는지 통계적 방법으로 확인	구축목표 대비 구축결과물 비율(수량) 기준 설정
3	구문 정확성	• 라벨 데이터 포맷과 값이 정확하게 입력되어 있고 필수항목 누락 여부 검사	정확도 99.5% 이상 권고
4	의미 정확성	• 어노테이션 값이 의미적으로 정답(GT)인지 확인하는 항목으로 실제적인 정답 비율 확인	정확도 95% 이상 권고
5	유효성	• 구축한 데이터세트를 잘 알려진 인공지능 학습 모델로 훈련시킨 후 목표로 했던 수준의 성능 달성이 가능한지 확인	인공지능 학습 모델 별 적정 기준 설정

〈표 3-12〉 품질검증항목 내용

## 2.2.5 품질 보완조치 프로세스



[그림 3-13] 품질 보완조치 프로세스

번호	활동	주체	주요 내용	산출물
1	품질검증 결과 검토	사업관리기관	<ul style="list-style-type: none"> <li>최종품질검증 결과 달성여부 검토</li> <li>검증 결과에 따라 사업수행기관에 보완조치 시행 요구</li> </ul>	<ul style="list-style-type: none"> <li>사업관리대장</li> </ul>
2	보완조치 수행	사업수행기관	<ul style="list-style-type: none"> <li>미달성 항목에 대한 보완 수행 (데이터, AI모델 등)</li> <li>보완 완료 결과를 업로드</li> </ul>	<ul style="list-style-type: none"> <li>보완 완료한 데이터 혹은 AI모델</li> </ul>
3	보완조치 확인	품질검증기관	<ul style="list-style-type: none"> <li>사업관리기관에 보완조치 확인 결과 통보</li> </ul>	<ul style="list-style-type: none"> <li>보완조치 확인 결과 (공문)</li> </ul>
4	보완조치 결과 확인 및 적정성 검토	사업관리기관	<ul style="list-style-type: none"> <li>보완조치 결과 확인</li> <li>필요시 사업수행기관에 재보완 요구</li> </ul>	<ul style="list-style-type: none"> <li>사업관리대장</li> </ul>

〈표 3-13〉 품질 보완조치 활동 및 산출물

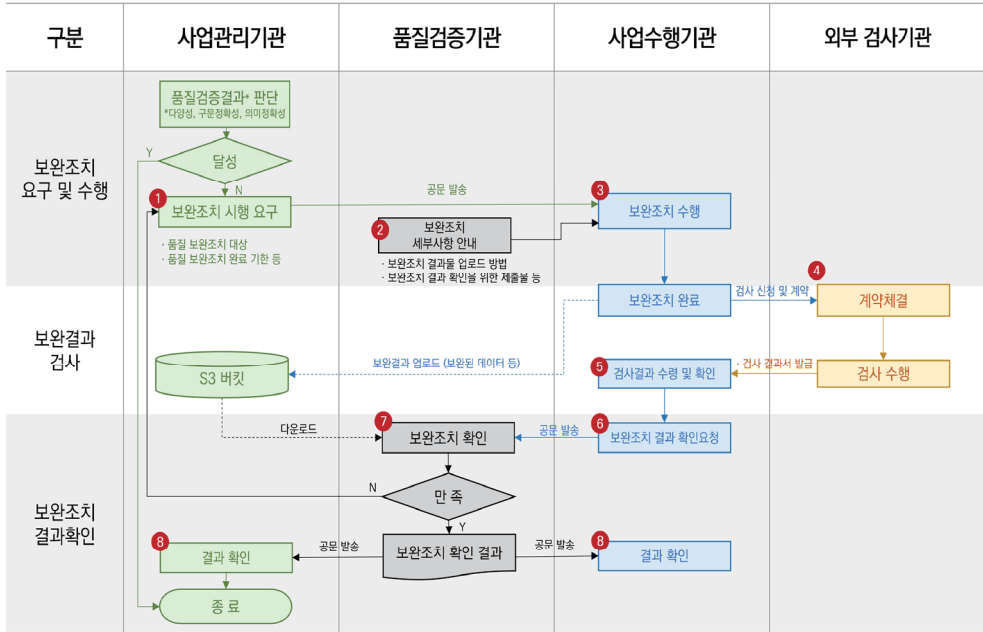
● 품질지표에 따른 보완조치 항목

※ 품질지표별 보완 대상, 세부 절차 등 일부 변동 가능

품질지표	보완 대상	검사 주체	세부 절차
다양성	전체 데이터	외부 검사기관	[그림 3-14]
구문 정확성			
의미 정확성			
유효성	AI모델, 데이터, 관련 서류	품질검증기관	[그림 3-15]

〈표 3-14〉 품질 보완조치 대상 구분

- 다양성, 구문 정확성, 의미 정확성 보완조치 절차



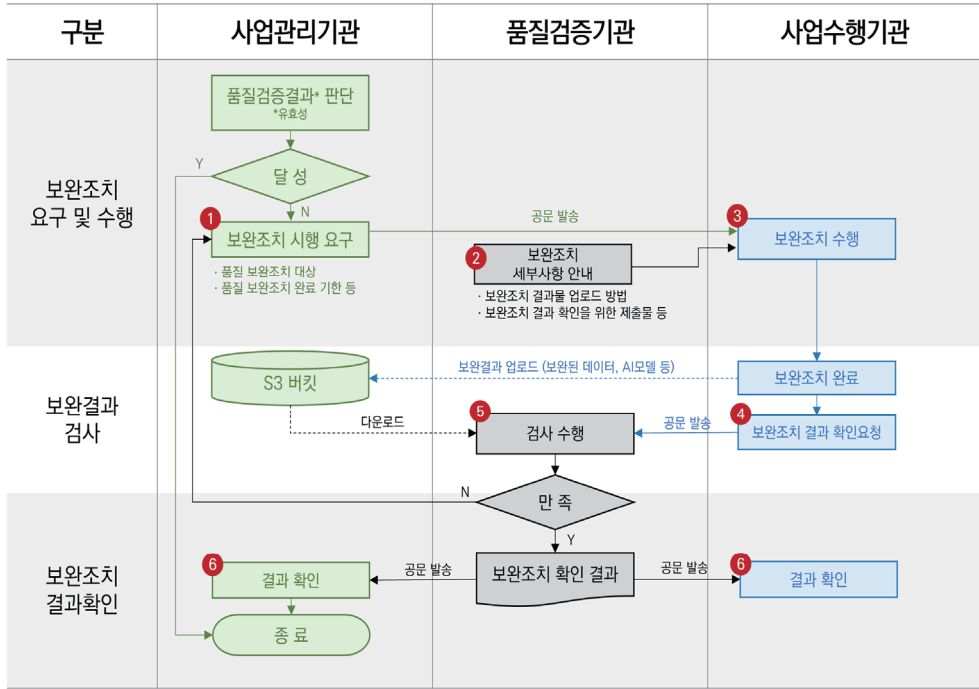
[그림 3-14] 품질 보완조치 세부절차(다양성, 구문 정확성, 의미 정확성)

번호	활동	주체	주요 내용
1	보완조치 시행 요구	사업관리기관	<ul style="list-style-type: none"> <li>사업관리기관은 품질검증 결과에 따라 사업수행기관에 품질 보완조치 시행 요구 공문 발송</li> <li>품질 보완조치 시행 요구 시, 아래 항목 포함               <ul style="list-style-type: none"> <li>품질 보완조치 대상</li> <li>데이터 보완 완료 기한 등</li> </ul> </li> </ul>
2	보완조치 세부사항 안내	품질검증기관	<ul style="list-style-type: none"> <li>사업수행기관은 품질검증기관에 연락하여 품질 보완조치에 대한 세부 안내 확인               <ul style="list-style-type: none"> <li>품질검증기관은 데이터 업로드 방법, 제출서류, 외부 검사기관 리스트 등에 대해 안내</li> </ul> </li> <li>※ 계약 추진 및 보완조치 검사 일정 협의 등은 사업수행기관이 직접 외부 검사기관과 진행함</li> </ul>

번호	활동	주체	주요 내용
3	보완조치 수행	사업수행기관	<ul style="list-style-type: none"> <li>사업수행기관은 전체 데이터에 대해 보완 수행</li> <li>사업수행기관은 보완 완료된 데이터를 S3 버킷에 업로드</li> </ul>
4	계약 체결	사업수행기관 외부 검사기관	<ul style="list-style-type: none"> <li>사업수행기관은 외부 검사기관과 개별적으로 검증 의뢰</li> <li>사업수행기관과 외부 검사기관은 검사 대상, 검사 기준, 검사 기간 등 협의 후 계약 체결</li> <li>※ 사업수행기관은 검증 일정을 사업관리기관, 품질검증기관에 공유하여야 함</li> </ul>
5	검사결과 수령 및 확인	사업수행기관	<ul style="list-style-type: none"> <li>사업수행기관은 외부 검사기관이 발급한 검사결과를 검토 후 이상없음을 확인</li> </ul>
6	보완조치 결과 확인요청	사업수행기관	<ul style="list-style-type: none"> <li>사업수행기관은 외부 검사기관의 검사결과를 바탕으로 품질검증기관에 보완조치 결과 확인요청 공문 발송                             <ul style="list-style-type: none"> <li>- 보완조치 결과 확인요청 공문에 외부 검사기관으로부터 발급받은 검사 결과를 첨부</li> </ul> </li> </ul>
7	보완조치 확인	품질검증기관	<ul style="list-style-type: none"> <li>품질검증기관은 사업수행기관의 제출서류(공문 및 외부 검사기관의 검사결과)를 검토                             <ul style="list-style-type: none"> <li>- 보완결과 만족 시, 품질검증기관은 보완조치 확인 결과를 사업관리기관 및 사업수행기관에 공문으로 통보</li> <li>- 보완결과 불만족 시, 품질검증기관은 사업관리기관에 이를 알리고, 사업관리기관은 사업수행기관에 재보완 요구</li> </ul> </li> </ul>
8	결과 확인	사업관리기관 사업수행기관	<ul style="list-style-type: none"> <li>사업관리기관과 사업수행기관은 품질검증기관의 보완조치 확인 결과 공문 수령 및 세부내역(엑셀) 최종 확인</li> </ul>

〈표 3-15〉 품질 보완조치 세부절차(다양성, 구문 정확성, 의미 정확성)

－ 유효성 보완조치 세부절차



[그림 3-15] 품질 보완조치 세부절차(유효성)

번호	활동	주체	주요 내용
1	보완조치 시행 요구	사업관리기관	<ul style="list-style-type: none"> <li>사업관리기관은 품질검증 결과에 따라 사업수행기관에 품질 보완조치 시행 요구 공문 발송</li> <li>품질 보완조치 시행 요구 시, 아래 항목 포함                             <ul style="list-style-type: none"> <li>품질 보완조치 대상</li> <li>데이터/AI모델 보완 완료 기한 등</li> </ul> </li> </ul>
2	보완조치 세부사항 안내	품질검증기관	<ul style="list-style-type: none"> <li>사업수행기관은 품질검증기관에 연락하여 품질 보완조치에 대한 세부 안내 확인                             <ul style="list-style-type: none"> <li>품질검증기관은 데이터 업로드 방법, 제출서류 등에 대해 안내</li> </ul> </li> </ul>
3	보완조치 수행	사업수행기관	<ul style="list-style-type: none"> <li>사업수행기관은 데이터, AI 모델, 관련 서류 등 보완 수행</li> </ul>

번호	활동	주체	주요 내용
4	보완조치 결과 확인요청	사업수행기관	<ul style="list-style-type: none"> <li>사업수행기관은 보완 완료된 결과물을 S3 버킷에 업로드 후 품질검증기관에 보완조치 결과 확인요청 공문 발송</li> </ul>
5	검사 수행 및 보완조치 확인	품질검증기관	<ul style="list-style-type: none"> <li>품질검증기관은 사업수행기관이 보완한 결과 및 제출서류를 기반으로 검사 수행 및 보완조치 확인                             <ul style="list-style-type: none"> <li>보완결과 만족 시, 품질검증기관은 보완조치 확인 결과를 사업관리기관 및 사업수행기관에 공문으로 통보</li> <li>보완결과 불만족 시, 품질검증기관은 사업관리기관에 이를 알리고, 사업관리기관은 사업수행기관에 재보완 요구</li> </ul> </li> </ul>
6	결과 확인	사업관리기관 사업수행기관	<ul style="list-style-type: none"> <li>사업관리기관과 사업수행기관은 품질검증기관의 보완조치 확인 결과 공문 수령 및 세부내역(엑셀) 최종 확인</li> </ul>

〈표 3-16〉 품질 보완조치 세부절차(유효성)





## 부 록

1. 주요 산출물 관계도
2. 품질 자가점검 체크리스트 양식



## 부록 | 1. 주요 산출물 관계도

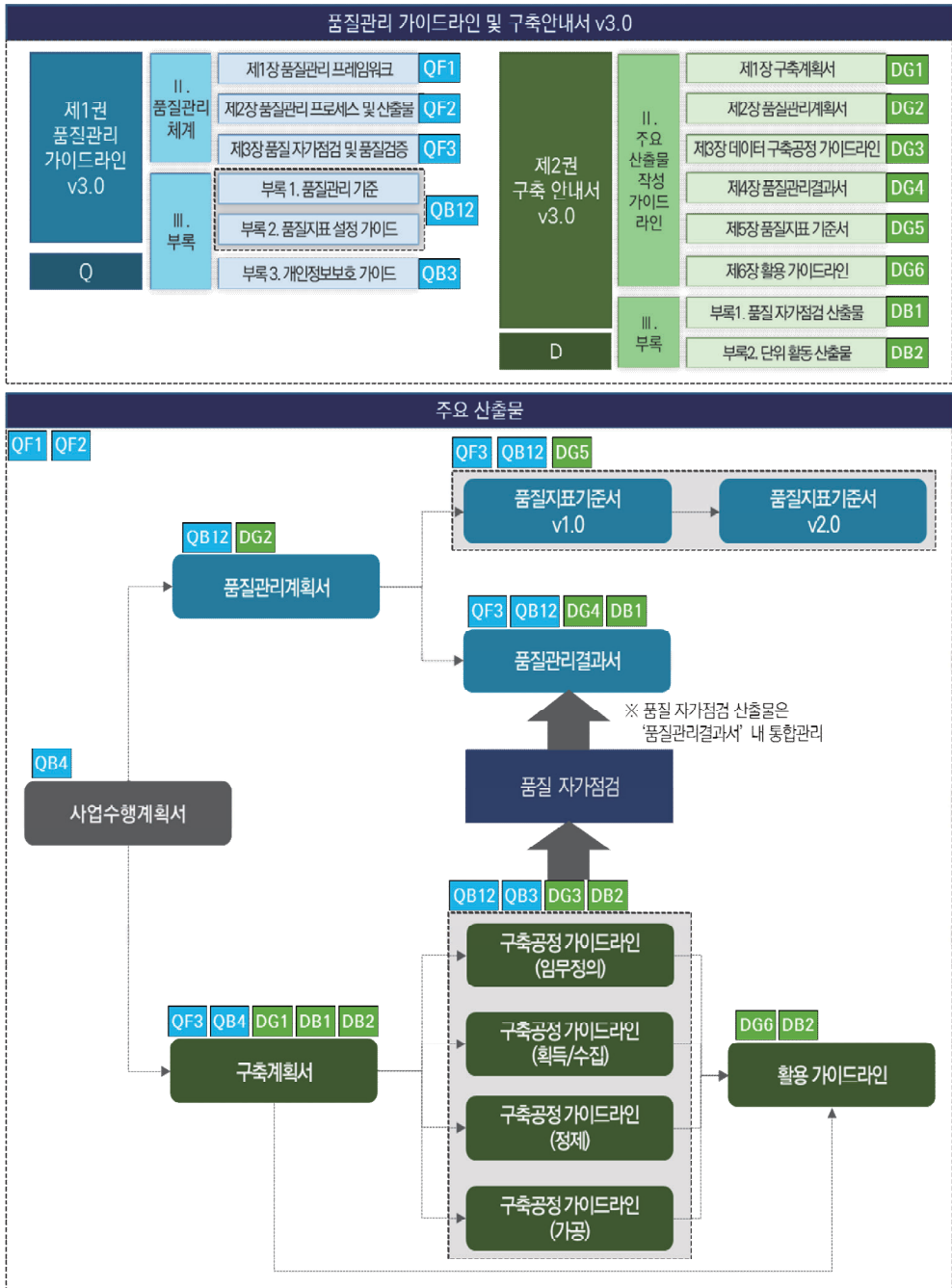
### 1 품질관리 가이드라인 및 구축 안내서 v3.0과 주요 산출물 관계도

- 구축안내서 본문에 작성된 구축계획서, 품질관리계획서, 구축공정 가이드라인, 품질관리결과서, 품질지표기준서, 활용 가이드라인 산출물을 중심으로 구축단계에 따른 산출물 흐름 정의
- 품질관리 가이드라인 및 구축안내서 v3.0에 작성된 내용과 구축단계별 주요 산출물과의 관계 정의는 다음과 같음

※ 본 관계도의 산출물 작성 시기, 변경 및 최종 제출 관련 사항은 제1권 품질관리 가이드라인 품질관리 체계 참조

※ '품질관리 가이드라인 및 구축안내서 v3.0과 주요 산출물 관계도' 그림 약어 참고

Q : 제1권 품질관리 가이드라인 v3.0	D : 제2권 구축안내서 v3.0
QF : 품질관리체계	DG : 주요 산출물 작성 안내
QB : 부록	DB : 부록



[그림 부록 1-1] 품질관리 가이드라인 및 구축 안내서 v3.0과 주요 산출물 관계도

- 품질관리 가이드라인 및 구축안내서 v3.0(제1권 품질관리 가이드라인 v3.0, 제2권 구축 안내서 v3.0)과 산출물 작성에 참조 가능한 부분과의 연관관계는 다음과 같음

품질관리 가이드라인 및 구축안내서 v3.0			산출물						제3자 품질검증
구분	목차		사업수행 계획서	구축 계획서	품질관리 계획서	구축공정 가이드 라인	품질관리 결과서	활용 가이드 라인	품질지표 기준서
제1권 품질관리 가이드 라인 v3.0	Ⅱ. 품질관 리체계	1. 품질관리 프레임워크	●	●	●	●	●	●	●
		2. 품질관리 프로세스 및 산출물	●	●	●	●	●	●	
		3. 품질 자가점검 및 품질검증		●			●		●
	Ⅲ. 부록	1. 품질관리 기준			● (체크리 스트 활용)	● (다양성, 정확성)			●
		2. 품질지표 설정 가이드			●	● (다양성, 정확성)			●
		3. 개인정보보호 가이드	●	●		●			
제2권 구축 안내서 v3.0	Ⅰ. 총론	3. 산출물 정의	●	●	●	●	●	●	●
	Ⅱ. 주요 산출물 작성 안내	1. 구축 계획서		●					
		2. 품질관리 계획서			●				
		3. 데이터 구축공정 가이드라인				●			
		4. 품질관리 결과서					●		
		5. 품질지표 기준서							●
		6. 활용 가이드라인						●	
	Ⅲ. 부록	1. 품질 자가점검 체크리스트		●			●		
		2. 단위 활동 산출물		●		●		●	

〈표 부록 1-1〉 품질관리 가이드라인 및 구축안내서 v3.0 산출물 관계 매트릭스

## 부록 | 2. 품질 자가점검 체크리스트 양식

### 1 과제조정위원회 자가점검 체크리스트

〈인공지능 학습용 데이터 구축계획서 및 품질관리계획서 자가점검 체크리스트〉

과제명	[차시]-[과제번호]. [과제명] (예. 1-001. 포괄적 비디오 이해를 위한 멀티모달 데이터)
데이터명	[차시]-[과제번호]-[데이터번호]. [데이터명] (예. 1-001-001. 픽셀 단위 장면 인지 멀티모달 데이터)
사업담당자	○○○ (예. 홍길동)
수행기관 작성자	[소속명] [이름] (예. (주)철수회사 김철수)

**붙임**

**1. 구축계획서 자가점검 체크리스트**

구분	점검항목		확인 (O/X)
제1장 구축 개요	1-01	데이터 구축목적이 명확하게 기술되어 있는가?	
	1-02	데이터 활용분야와 활용방안이 명확하게 기술되어 있는가?	
제2장 데이터 구축	2-01	데이터 활용분야가 분류되어 있는가?	
	2-02	원시데이터 유형과 포맷이 명시되어 있는가?	
	2-03	원시데이터 획득 규모가 명시되어 있는가?	
	2-04	원시데이터 획득방법이 명시되어 있는가?	
	2-05	원시데이터 획득 기준 수립 시 전문가 (도메인, 인공지능 등)의 검토 및 자문을 받았는가?	
	2-06	원시데이터의 저작권, 지적재산권 등 법적권리 확보 방안이 명시되어 있는가?	
	2-07	원천데이터 유형과 포맷이 명시되어 있는가?	
	2-08	원천데이터 규모가 명시되어 있는가?	
	2-09	원천데이터 정제 기준이 명시되어 있는가?	
	2-10	원천데이터의 환경이 원시데이터의 구축환경을 반영하고 있는가?	
	2-11	데이터 라벨링 유형이 명시되어 있는가?	
	2-12	라벨링 데이터 구성 방식이 명시되어 있는가?	
	2-13	라벨링 데이터 규모가 명시되어 있는가?	
	2-14	라벨링 데이터 중 메타데이터에 포함되어야 하는 항목이 별도로 명시되어 있는가?	

구분	점검항목		확인 (O/X)
	2-15	라벨링된 데이터에 대한 교차검증 방안이 있는가?	
	2-16	학습모델 임무유형이 명시되어 있는가?	
	2-17	학습모델의 구조가 제시되었는가?	
	2-18	학습모델 평가지표가 제시되었는가?	
	2-19	학습모델 임무 유형과 데이터 활용방안이 부합하는가?	
제3장 개인정보보호 및 저작권	3-01	개인정보보호 조치 및 대응 방안이 수립되어 있는가?	
	3-02	데이터 및 학습모델에 대한 저작권 대응방안 또는 확보계획이 수립되어 있는가?	
제4장 작업자 운영방안	4-01	작업자 가이드 마련을 위한 세부 계획이 수립되어 있는가?	
	4-02	각 구축단계별 작업에 소요되는 인원의 적정 요구수준과 규모가 제시되어 있는가?	
	4-03	작업자 이탈에 따른 위험관리 절차가 마련되어 있는가?	
	4-04	작업자별 작업진도 관리를 위한 모니터링 환경(SW)을 보유하고 있는가?	
	4-05	데이터 작업 내역에 대한 산출 근거 및 기준 등을 명확하게 가지고 있는가?	



붙임

2. 품질관리계획서 자가점검 체크리스트

구분	점검항목		확인 (O/X)
제1장 개요	1-01	구축공정에 대한 품질지표와 목표가 설정되어 있는가?	
	1-02	데이터 적합성(기준적합성, 기술적합성, 통계적 다양성)에 대한 품질지표와 목표가 설정되어 있는가?	
	1-03	데이터 정확성(구문정확성, 의미정확성)에 대한 품질지표와 목표가 설정되어 있는가?	
	1-04	학습모델에 대한 품질지표와 목표가 설정되어 있는가?	
	1-05	품질관리 조직체계와 각 조직의 역할과 책임이 명확한가?	
	1-06	품질관리 수행을 위한 주관/참여 컨소시엄의 역할과 책임이 명확한가?	
	1-07	품질관리조직 구성에 따른 역할에 중복/누락된 것이 없는가?	
	1-08	품질관리 역할에 따른 책임자 및 실무자를 빠짐없이 기재하였는가?	
제2장 품질 관리 기준	2-01	데이터 구축과정 품질관리를 위한 세부지표와 점검방법이 명시되어 있는가?	
	2-02	기준 적합성 세부지표와 검사방법이 명시되어 있는가?	
	2-03	기술 적합성 세부지표와 검사방법이 명시되어 있는가?	
	2-04	통계적 다양성 세부지표와 검사방법이 명시되어 있는가?	
	2-05	구문 정확성 세부지표와 검사방법이 명시되어 있는가?	
	2-06	의미 정확성 세부지표와 검사방법이 명시되어 있는가?	
	2-07	데이터 품질관리를 위한 데이터 정의서가 마련되어 있는가?	

구분	점검항목		확인 (O/X)
제3장 학습모델 기준	3-01	임무정의가 데이터 구축목적에 부합하는가?	
	3-02	제시된 활용모델이 임무정의에 부합하는 모델인가?	
	3-03	활용모델의 평가지표(Evaluation Metric)가 제시된 활용모델과 합치하는가?	
	3-04	활용모델의 평가지표(Evaluation Metric) 별 목표기준 설정이 적정한가?	
제4장 품질검사 절차	4-01	구축공정 단계별 품질관리 과업 및 일정계획이 마련되어 있는가?	
	4-02	데이터 품질검사 과업 및 일정계획이 마련되어 있는가?	
	4-03	학습모델 개발, 훈련 및 평가계획이 마련되어 있는가?	
제5장 품질검사 교육 방법	5-01	학습모델 개발 담당 조직이 사업초기에 적극적으로 가담하는가?	
	5-02	외부 참여인원(클라우드워커)에 대한 교육계획이 마련되어 있는가?	
제6장 품질관리 도구	6-01	데이터 구문정확성 검사를 위한 도구(SW)를 보유하고 있는가?	
	6-02	데이터 의미정확성 검사를 위한 도구(SW)를 보유하고 있는가?	
	6-03	데이터 유효성 검사를 위한 도구(SW)를 보유하고 있는가?	

## 2 샘플데이터 1-Cycle 자가점검 체크리스트

과제명	[차시]-[과제번호]. [과제명] (예. 1-001. 포괄적 비디오 이해를 위한 멀티모달 데이터)
데이터명	[차시]-[과제번호]-[데이터번호]. [데이터명] (예. 1-001-001. 픽셀 단위 장면 인지 멀티모달 데이터)
사업담당자	○○○ (예. 홍길동)
수행기관 작성자	[소속명] [이름] (예. (주)철수회사 김철수)
1-Cycle 점검기간	202X. XX. XX ~ 202X. XX. XX
실행환경	(개발환경, 개발언어, 활용 알고리즘 패키지 등)

구분	점검항목		확인 (O/X)
제1장 샘플데이터 획득 및 정제	1-1	기술 적합성, 통계적 다양성을 고려한 원시데이터 획득/수집 가이드라인이 존재하는가?	
	1-2	원천데이터 품질확보를 위한 정제 가이드라인(지침)이 존재하는가?	
	1-3	원천데이터 저장소(파일디렉토리) 구조가 정의되어 있는가?	
	1-4	파일 유형 및 포맷이 수행계획서 상에 제시한 기준과 일치하는가?	
제2장 샘플데이터 가공	2-1	라벨링 작업의 오차범위를 줄이기 위한 가공(라벨링) 가이드라인이 존재하는가?	
	2-2	샘플데이터 내 수행계획서에서 제시한 모든 카테고리의 데이터가 포함되어 있는가?	
제3장 원천데이터 품질	3-1	원천데이터 특성에 따른 필수 점검항목이 모두 기술되어 있는가?	
	3-2	각 점검항목 별 오류사항에 대한 보완조치 방안이 마련되어 있는가?	

구분	점검항목		확인 (O/X)
제4장 라벨링데이터 품질	4-1	데이터 라벨링 환경(클라우드플랫폼) 혹은 저장도구(SW)를 확보하고 있는가?	
	4-2	샘플데이터는 자체 품질책임자가 검사 완료한 데이터인가?	
제5장 샘플데이터 학습	5-1	1-Cycle 점검 시 구현한 학습모델이 수행계획서상 학습모델 후보군에 해당하는 모델인가?	
	5-2	학습모델은 기존의 공개된 오픈소스를 활용하여 구현하였는가?	

※ 샘플데이터 1-Cycle 자가점검 체크리스트의 결과보고서 양식은 ‘구축안내서 v3.0’ 참조

### 3 초기데이터 1-Cycle 자가점검 체크리스트

과제명	[차시]-[과제번호]. [과제명] (예. 1-001. 포괄적 비디오 이해를 위한 멀티모달 데이터)
데이터명	[차시]-[과제번호]-[데이터번호]. [데이터명] (예. 1-001-001. 픽셀 단위 장면 인지 멀티모달 데이터)
사업담당자	000 (예. 홍길동)
수행기관 작성자	[소속명] [이름] (예. (주)철수회사 김철수)
1-Cycle 점검기간	202X. XX. XX ~ 202X. XX. XX
실행환경	(개발환경, 개발언어, 활용 알고리즘 패키지 등)

구분	점검항목		확인 (O/X)
1. 원천데이터	1-1	원천데이터 필수 점검항목 중 오류율이 10% 이상인 항목이 존재하는가?	
	1-2	필수 점검항목에 대한 보완조치 방법을 정제 가이드라인에 반영하였는가?	
2. 다양성	2-1	클래스 분포도, 인스턴스 분포도, 문장길이 등 다양성 지표 측정기준을 마련하였는가?	
	2-2	초기데이터 다양성 검사를 수행하였는가?	
	2-3	카테고리별 데이터 수량이 수행계획서에서 명시한 다양성 분포기준을 만족하는가?	
	2-4	분포기준을 만족하지 못하는 카테고리에 대한 추가 구축 계획을 수립하였는가?	
3. 구문 정확성	3-1	라벨링데이터 구조, 형태 등 구문 정확성 측정을 위한 구문규칙을 마련하였는가?	
	3-2	자동화 검사 툴을 이용하여 구문 정확성 검사(전수) 수행 후 육안검사(샘플링)를 수행하였는가?	

구분	점검항목		확인 (O/X)
	3-3	구축한 초기데이터가 수행계획서에 명시한 구문 정확성 정량목표를 만족하는가?	
	3-4	구문규칙에 반하는 오류(구조, 형식) 사례를 발굴하여 가공 가이드라인에 반영하였는가?	
4. 의미 정확성	4-1	의미 정확성 측정을 위한 라벨링 유형별 검사기준을 마련하였는가?	
	4-2	의미 정확성 검사 시 2인 이상의 교차검사 과정을 거쳤는가?	
	4-3	구축한 초기데이터가 수행계획서에 명시한 의미 정확성 정량목표를 만족하는가?	
	4-4	검사결과에 따른 의미적 오류 사례를 발굴하여 가공 가이드라인에 반영하였는가?	
5. 유효성	5-1	AI 모델 학습 전, 학습용 데이터셋을 전수검사하여 모든 오류를 제거하였는가?	
	5-2	학습용 데이터셋은 편향성을 방지하여 구성하였는가?	
	5-3	AI 모델의 과적합 방지를 위한 사전 조치방안을 마련하였는가?	
	5-4	유효성 측정결과 출력과 관련하여 오픈소스를 활용하였는가?	

※ 초기데이터 1-Cycle 자가점검 체크리스트의 결과보고서 양식은 ‘구축안내서 v3.0’ 참조

## 4 중간데이터 품질 자가점검 체크리스트

과제명	[차시]-[과제번호]. [과제명] (예. 1-001. 포괄적 비디오 이해를 위한 멀티모달 데이터)
데이터명	[차시]-[과제번호]-[데이터번호]. [데이터명] (예. 1-001-001. 픽셀 단위 장면 인지 멀티모달 데이터)
사업담당자	000 (예. 홍길동)
수행기관 작성자	[소속명] [이름] (예. (주)철수회사 김철수)
중간품질 점검기간	202X. XX. XX ~ 202X. XX. XX

구분	점검항목		확인 (O/X)
제1장 계획이행 및 진도점검	1-1	초기데이터 1-Cycle 자가점검 시 발견한 수정사항을 대량 데이터 구축 중에도 유지 반영하였는가?	
	1-2	사전 품질검증에서 도출된 초기데이터 검토결과를 구축과정에 수정 반영하였는가?	
	1-3	컨소시엄 내 품질관리 조직의 역할이 원활하게 이행되고 있는가?	
	1-4	사업종료 1개월 전 품질지표 기준서 v2.0을 확정 및 변경관리를 수행하였는가?	
제2장 유효성	2-1	구축한 대량(50%이상) 데이터가 품질지표 기준서 및 사업수행계획서에 명시한 유효성 정량목표를 만족하는가?	

## 품질관리 가이드라인 v3.0

# 요약본

2023년 2월 발행

발행처: 한국지능정보사회진흥원

### 〈작성 참여진〉

한국지능정보사회진흥원 이용진 본부장

한국지능정보사회진흥원 손기문 팀장

한국지능정보사회진흥원 오현목 팀장

한국지능정보사회진흥원 유호진 팀장

한국지능정보사회진흥원 신다울 팀장

한국지능정보사회진흥원 박현우 수석

한국지능정보사회진흥원 전진우 선임

한국지능정보사회진흥원 윤주미 연구원

(주)티지 오윤환 이사

(주)비트레스 류동주 대표

(주)아이티지엔 양원 대표

### 〈자문 위원〉

한국정보통신기술협회 신준호 단장

가톨릭대학교 이상국 교수

세종대학교 김진철 교수

한국전자기술연구원 김동철 책임

한국표준과학연구원 이재두 초빙연구위원

(주)아이삭 김상복 기술사

- 본 『품질관리 가이드라인 v3.0 요약본』 내용의 무단전제 및 재배포를 금하며, 가공·인용 시에는 반드시 과학기술정보통신부, 한국지능정보사회진흥원의 「인공지능 학습용 데이터 품질관리 가이드라인 및 구축 안내서 v3.0 - 품질관리 가이드라인 v3.0 요약본」임을 밝혀주시기 바랍니다.
- 본 『품질관리 가이드라인 v3.0 요약본』은 지능정보산업 인프라 조성을 위한 인공지능 학습용 데이터 구축사업 중 ‘인공지능 학습용 데이터 품질관리 컨설팅 및 체계 고도화’ 영역 사업의 결과 산출물입니다.
- 한국지능정보사회진흥원 지능데이터본부 시데이터활용팀 품질유닛 : aidq@nia.or.kr, 053-230-4220