

KODIA

디자인데이터 품질진단 프레임워크

현장 검수 담당자를 위한 실무 가이드

10가지 평가 기준 · 5단계 프로세스 · 파일별 체크포인트

(사)한국디자인산업연합회 디자인데이터 품질연구소

Framework v1.0 | 2024 AI 학습용 데이터 품질관리 가이드라인 v3.1 기반

목차

1장	이 가이드는 무엇인가?	03
2장	KODIA 프레임워크 개요	04
3장	10가지 평가 기준 상세	05
4장	리커트 5점 척도 판정 방법	09
5장	5단계 평가 프로세스	10
6장	파일별 체크포인트	13
7장	약식 검수 3가지 기준 (Q1·Q2·Q3)	16
8장	DQ Score 산출 방법	17
9장	캘리브레이션 — 평가 전 준비	18
10장	평가 중 인지 편향 통제	19
11장	문제 발견 시 피드백 방법	20
12장	역할별 체크리스트	21
부록	용어 사전 · 참고 문서	23

1장. 이 가이드는 무엇인가?

이 가이드는 AI 학습용 디자인 데이터를 직접 검수하는 현장 담당자를 위한 실무 참조 문서입니다. 코드 오류나 파일 구조 같은 기술적 문제는 시스템이 자동으로 걸러줍니다. 이 가이드는 그 다음 단계, 즉 **사람이 봐야만 알 수 있는 것들**을 어떻게 판단할지를 다룹니다.

이 가이드에서 다루는 것

- 무엇을 보고 판단해야 하는가 — 10가지 판단 기준 (K-01~K-10)
- 어떻게 점수를 매기는가 — 리커트 5점 척도 운영 방법
- 5단계 평가 프로세스와 각 단계에서 담당자가 할 일
- 데이터 유형(텍스트·사진·2D·3D)별 구체적 점검 항목
- 빠르게 확인하는 약식 검수 3가지 기준
- 평가 전 준비(캘리브레이션)와 인지 편향 통제 방법

이 가이드에서 다루지 않는 것

- 코드 오류, JSON/XML 구조 오류, 번역 문법 정확성
- 시스템 코드 검증, 알고리즘 성능 검증
- 통계 분석 방법 (관리자 및 운영위원회 담당 영역)

2장. KODIA 프레임워크 개요

KODIA(Heuristic-based Expert Design Data Quality Assessment) 프레임워크는 디자인 데이터를 단순한 파일이 아니라 디자인 AI의 학습 품질을 결정하는 핵심 학습 자원으로 보고, 학습 가능성·재사용성·라벨 신뢰성·편향 위험·실사용 적합성까지 함께 진단합니다.

기존 기술 검수 vs. KODIA 휴리스틱 검수

구분	기존 기술적 검수	KODIA 품질진단
검수 방식	자동화 시스템	전문가 직접 판단
판단 기준	코드·수치·규칙 기반	사전 합의된 10가지 원칙 기반
잡을 수 있는 것	포맷 오류, 누락, 중복	감성 불일치, 맥락 부적합, 편향
결론	기술적 정확성 확보	디자인 품질·감성 가치 확보

핵심 원칙: 예쁜가보다 학습 가능한 구조로 잘 정리됐는가를 중심으로 판단합니다. 파일 자체의 품질뿐 아니라 메타데이터와 라벨 품질이 AI 학습 결과를 크게 좌우합니다.

3장. 10가지 평가 기준 상세

아래 10가지가 평가의 기준입니다. 각 기준마다 **평가 질문**과 **이러면 문제** 사례가 함께 제시됩니다. 평가 기준은 4개 차원에 걸쳐 배치됩니다.

의미·정합성	K-01, K-03, K-05	데이터 구성 요소들이 동일한 의도와 메시지를 향하는가
데이터 무결성	K-04, K-06, K-07	데이터가 완전하고 재사용 가능한 구조로 정리됐는가
실사용 전환성	K-10	실제 서비스·브랜드 산출물로 전환될 수 있는가
학습 적합성	K-02, K-08, K-09	AI가 유효한 패턴을 학습할 수 있는 다양성·공정성을 갖추었는가

K-01 정합성

의미·정합성

평가 질문: 텍스트·이미지·2D·3D가 같은 콘셉트와 메시지를 향하는가?

데이터 구성 요소들(텍스트·이미지·태그)이 동일한 디자인 의도와 메시지를 향하는지 평가합니다. 멀티모달 데이터 간 의미 일치 판단에 특히 중요합니다.

이러면 문제: 이미지는 따뜻한 분위기인데 태그에는 '차가움'이라고 붙어 있음

K-02 대표성

학습 적합성

평가 질문: 특정 스타일만 과도하게 많지 않고, 학습 목적에 맞는 다양성을 갖는가?

배치 내 데이터가 AI 학습 목적에 맞게 다양한 스타일·색·주제로 균형 있게 분포했는지 평가합니다. 과도한 편중은 AI 과적합의 원인이 됩니다.

이러면 문제: 배치 1,000건 중 800건이 미니멀 스타일 — 다른 스타일 학습 불가

K-03 일관성

의미·정합성

평가 질문: 같은 브랜드/서비스/카테고리에서 스타일 규칙이 흔들리지 않는가?

같은 브랜드나 카테고리 내에서 스타일·톤·색상 규칙이 균질하게 유지되는지 평가합니다. 배치 전체를 보고 판단합니다.

이러면 문제: 같은 배치 안에 미니멀 이미지와 화려한 이미지가 혼재

K-04 표현 충실도

데이터 무결성

평가 질문: 디자인 의도가 저해상도·왜곡·누락 없이 잘 담겼는가?

이미지·2D·3D·영상 데이터의 표현 품질을 판단합니다. 디자인 의도가 파일에 잘 구현됐는지, 기술적 품질 저하가 없는지 확인합니다.

이러면 문제: 저해상도로 인해 세부 디자인 요소가 뭉개져 AI 학습에 부적합

K-05 주석 신뢰성	의미·정합성
평가 질문: 라벨·태그·설명이 사람이 봐도 재현 가능하게 정의됐는가? 라벨·태그·설명이 처음 보는 사람도 동일하게 이해하고 재현할 수 있도록 구체적으로 정의됐는지 평가합니다. 모호한 라벨은 AI 학습 재현성을 해칩니다. 이러면 문제: '따뜻한 느낌'처럼 모호한 라벨 — 재현 불가능. 'Warm 톤, 2700K 조명'처럼 구체적이어야 함	
K-06 구조화 가능성	데이터 무결성
평가 질문: 나중에 검색·필터링·재사용하기 쉬운 메타데이터가 붙어 있는가? 검색·필터링·재사용·학습 파이프라인에 필요한 메타데이터가 체계적으로 붙어 있는지 평가합니다. 메타데이터 없이는 데이터 재사용과 파이프라인 자동화가 불가능합니다. 이러면 문제: 파일명만 있고 카테고리·해상도·용도 메타데이터가 전혀 없음	
K-07 중복·변형 관리	데이터 무결성
평가 질문: 거의 같은 샘플이 과도하게 반복되지 않는가? 거의 동일한 샘플이 과도하게 반복되면 AI가 특정 패턴에 과적합될 위험이 높아집니다. 의미 없는 변형(크롭·밝기 미세 조정)도 확인합니다. 이러면 문제: 동일 제품 사진을 밝기만 5단계로 변형한 5장 — 실질적으로 1장	
K-08 편향 위험	학습 적합성
평가 질문: 특정 색·레이아웃·문화권·성별 표현이 과도하게 치우치지 않는가? 데이터 내에서 특정 색·레이아웃·문화권·성별 표현이 과도하게 편중되지 않았는지 평가합니다. AI 데이터 품질에서 필수 점검 항목입니다. 이러면 문제: 글로벌 브랜드 데이터인데 동아시아 문화권 이미지만 900건 — 서구권 표현 부재	
K-09 확장성	학습 적합성
평가 질문: 다른 해상도·포맷·채널로 변환해도 의미가 유지되는가? 다른 해상도·포맷·채널로 변환해도 디자인 의미와 품질이 유지되는지 평가합니다. 멀티채널·멀티포맷 활용성 판단에 유효합니다. 이러면 문제: 데스크탑용 배너를 모바일로 변환하면 텍스트가 잘려 메시지 전달 불가	

K-10 실사용 적합성

실사용 전환성

평가 질문: 실제 서비스 화면이나 브랜드 산출물로 무리 없이 쓰이는가?

학습 데이터가 실제 서비스 환경이나 브랜드 산출물로 전환될 때 무리 없이 쓰일 수 있는지 평가합니다. 이론적으로는 맞아도 실제 적용 시 어색하면 낮은 점수입니다.

이러면 문제: AI가 생성한 UI 컴포넌트가 실제 앱에 배치하면 다른 컴포넌트와 스타일 충돌

4장. 리커트 5점 척도 판정 방법

각 기준마다 OK / 부분OK / 문제를 먼저 판정하고, 이어서 1~5점 점수를 부여합니다.

점수	판정	기준	처리 방법
5점	OK	기준을 완벽히 충족. 추가 설명 없이 즉시 승인 가능	승인
4점	OK	기준을 잘 충족. 경미한 개선 여지가 있으나 승인 가능	승인
3점	부분OK	보통 수준. 큰 문제는 없으나 개선 가능. 조건부 수용	조건부
2점	문제	미흡. 수정 후 재검수 필요	수정
1점	문제	매우 미흡. 재작업 수준의 문제	재작업

3점(부분OK)을 남용하지 마세요

매매할 때마다 3점을 주면 평가의 변별력이 없어집니다. 판단이 어려우면 3점으로 표시하되 반드시 '판단 어려움, 추가 검토 필요'라고 근거를 기록하세요. 세션당 3점 비율이 40%를 초과하면 중심화 편향을 의심해야 합니다.

5장. 5단계 평가 프로세스

KODIA 평가는 아래 5단계 순서로 진행합니다. 단계를 건너뛰면 신뢰도가 떨어질 수 있습니다.

STEP 1 데이터 유형 분석 및 평가 기준 설정

담당: 데이터 관리자

1. 데이터를 유형별로 분류합니다 (텍스트 / 사진 / 2D / 3D).
2. 데이터의 목적·브랜드·타겟·학습 목적을 파악합니다.
3. 데이터 ID를 부여하고 특성에 따라 적용할 평가 기준(K-01~K-10)을 확인합니다.
4. 위험도를 분류합니다: 신규 유형·AI 이상 신호·레이블 모호 데이터 → 고위험(전수검사).

→ 산출물: 검수 계획서, 데이터 분류표, 위험도 목록

STEP 2 전문가 선정 및 캘리브레이션

담당: 운영 담당자 + 전문가
전원

1. 데이터 유형·산업 도메인에 맞는 전문가를 배정합니다. 범위 외 배정은 금지입니다.
2. 캘리브레이션 세션을 수행합니다: 고·중·저 품질 샘플 각 5건(총 15건)을 독립 평가합니다.
3. 전체 공개 토론을 통해 점수 기준을 합의하고 기준 문서를 확정합니다.
4. 모든 평가자가 기준 문서에 서명합니다.

→ 산출물: 평가 위원 배정표, 서명된 기준 문서, 캘리브레이션 합의 점수표

STEP 3 AI 사전 검수 + 전문가 휴리스틱 평가

담당: AI 시스템(1차) +
전문가(2차)

1. AI가 먼저 기술적 문제(중복·저품질·태그 오류)를 자동 탐지합니다.
2. 이상 신호 데이터를 우선 검수 대상으로 분류합니다.
3. 전문가가 K-01~K-10 기준으로 각 데이터를 평가합니다.
4. 원칙별로 OK/부분OK/문제를 판정하고 1~5점 점수와 판단 근거를 기록합니다.
5. 세션당 최대 50건으로 제한하고 중간에 충분히 휴식합니다.

→ 산출물: AI 검수 리포트, 전문가 평가 시트

STEP 4 교차 검증 및 통계적 신뢰도 분석

담당: 분석 시스템 + 데이터
관리자

1. 배분 데이터의 10~20% 수준에서 중첩 데이터 교차 검증을 진행합니다.
2. 평가자 간 일치도(ICC)를 확인합니다: 0.75 이상 양호, 0.60 미만 재검수.
3. 이상하게 높거나 낮은 점수 패턴을 탐지합니다.
4. 5개 차원별 가중 합산으로 DQ Score를 산출합니다.

→ 산출물: 신뢰도 분석 리포트, DQ Score 목록

STEP 5 결과 보고 및 품질 개선 피드백

담당: 관리자 + 제작팀

1. 단순한 적합/부적합 판정이 아니라 실질적 수정 가이드를 제공합니다.
2. 결과 보고서 초안 작성 → 개선 방안 포함 공유 → 재검수 → 최종 보고서 생성.
3. 수정 완료된 데이터는 재검수를 거쳐야 합니다. 임의 승인은 금지입니다.
4. 반복되는 오류 패턴은 차기 배치 제작 기준에 반영합니다.

→ 산출물: 결과 보고서, 수정 가이드, 재검수 결과, 승인 데이터셋

6장. 파일별 체크포인트

데이터 유형에 따라 확인해야 할 구체적 항목이 다릅니다. 해당 유형의 체크포인트를 평가 전에 반드시 숙지하세요. 핵심은 예쁜가가 아니라, 학습 가능한 구조로 잘 정리됐는가입니다.

텍스트 데이터
의미·정합성
☐ 라벨-설명-컨셉트 간 의미 일치성 확인 (K-01)
☐ 같은 배치 내 톤 일관성 (K-03)
☐ 브랜드 용어 표준화 준수 여부
☐ CTA(행동 유도 문구) 강도의 적절성
구조·무결성
☐ 길이 분포 — 너무 짧거나 긴 텍스트 편중 여부 (K-07)
☐ 중복 문구 비율 — 유사 텍스트 과다 반복 여부 (K-07)
☐ 검색·필터링용 메타데이터 부착 여부 (K-06)
☐ 라벨이 사람이 봐도 재현 가능하게 정의됐는지 (K-05)
사진 데이터
표현 품질
☐ 구도 안정성 — 불안정한 기울기·프레이밍 여부 (K-04)
☐ 배경 노이즈 — 학습을 방해하는 배경 요소 여부 (K-04)
☐ 색감 일관성 — 배치 내 화이트 밸런스·채도 균질성 (K-03)
☐ 시각적 주의점의 명확성 — 주제가 명확히 부각되는가 (K-04)
학습 적합성
☐ 메시지 전달력 — 디자인 의도가 사진에 담겼는가 (K-01)
☐ 주제 적합성 — 라벨된 주제와 사진 내용 일치 (K-05)
☐ 객체 식별 가능성 — AI가 주요 객체를 인식 가능한가 (K-04)
☐ 학습 방해 요소 여부 — 워터마크·텍스트 오버레이 등 (K-04)
☐ 편향 위험 — 특정 인구 집단·문화권 이미지만 편중됐는가 (K-08)

2D 데이터

레이아웃 구조

- ☐ 레이아웃 그리드 일관성 — 그리드 시스템 준수 여부 (K-03)
- ☐ 간격 규칙 — 패딩·마진 일관성 (K-03)
- ☐ 시각 위계 — 헤드라인·본문·CTA 위계 명확성 (K-04)
- ☐ 컬러 팔레트 일관성 — 브랜드 컬러 가이드 준수 (K-03)

재사용성·확장성

- ☐ 컴포넌트 재사용성 — 반복 사용 가능한 구조인가 (K-06)
- ☐ 다른 해상도로 변환 시 의미 유지 여부 (K-09)
- ☐ 레이어 명칭·구조 정리 상태 (K-06)
- ☐ 메타데이터 및 파일 네이밍 규칙 준수 (K-06)

3D 데이터

형태·재질 품질

- ☐ 형태 인지성 — 주요 형태가 명확히 식별되는가 (K-04)
- ☐ 재질·조명 일관성 — 배치 내 렌더링 조건 균질성 (K-03)
- ☐ 모델 깨짐 여부 — 폴리곤 오류·UV 이상 등 (K-04)
- ☐ 스케일 기준 일관성 — 실세계 기준 스케일 정확성 (K-04)

메타데이터·맥락

- ☐ 실제 맥락 적합성 — 실제 환경에서 자연스러운가 (K-10)
- ☐ 다각도 일관성 — 여러 뷰에서 동일 품질 유지 (K-04)
- ☐ 렌더링 조건 메타데이터 포함 — 조명·카메라 설정 (K-05, K-06)
- ☐ 파일 포맷·해상도 표준 준수 여부 (K-06)

7장. 약식 검수 3가지 기준 (Q1·Q2·Q3)

약식 검수는 정식 검수를 대체할 수 없습니다

위험도 높음 데이터, 신규 유형 데이터, 최종 승인 전 데이터는 반드시 정식 10가지 기준으로 검수해야 합니다.
약식 기준은 위험도 낮음 데이터의 1차 필터링에만 사용합니다.

Q1 구성 요소들이 하나의 콘셉트와 메시지를 향하는가?

텍스트·이미지·2D·3D 등 구성 요소들이 같은 디자인 의도를 전달하는지 10초 안에 판단합니다.

커버 기준: K-01 정합성, K-03 일관성

통과	모든 구성 요소가 동일한 콘셉트를 향한다
재확인	대체로 맞으나 일부 요소가 어긋난다 → 정식 검수(K-01)
탈락	구성 요소 간 의미가 충돌한다 → 즉시 정식 검수

Q2 라벨과 메타데이터가 재현 가능하게 정의됐는가?

라벨·태그·설명을 처음 보는 사람도 동일하게 이해하고 재현할 수 있는지 확인합니다.

커버 기준: K-05 주석 신뢰성, K-06 구조화 가능성

통과	라벨과 설명이 명확하고 재현 가능하다
재확인	일부 모호한 표현이 있다 → 정식 검수(K-05, K-06)
탈락	라벨이 모호하거나 내용과 불일치 → 즉시 수정

Q3 배치 내 특정 스타일이나 속성이 과도하게 쏠려 있는가?

배치 전체를 훑어보며 특정 색·레이아웃·문화권이 지나치게 편중됐는지, 유사 샘플이 반복됐는지 확인합니다.

커버 기준: K-02 대표성, K-07 중복·변형 관리, K-08 편향 위험

통과	다양성이 확보되어 있고 반복이 적다
재확인	약간의 편중이 있다 → 정식 검수(K-02, K-08)
탈락	심각한 편중 또는 과다 중복 → 배치 전체 정식 검수

8장. DQ Score 산출 방법

4개 평가 차원별 점수에 가중치를 곱해 합산한 종합 점수입니다. 가중치는 프로젝트 착수 시 평가 위원 전체가 AHP 합의를 통해 확정합니다.

평가 차원	포함 기준	기본 가중치	이 차원이 낮으면
의미·정합성	K-01, K-03, K-05	30%	라벨 재정의, 멀티모달 요소 간 의미 일치 재확인
데이터 무결성	K-04, K-06, K-07	25%	파일 품질 보정, 메타데이터 정비, 중복 제거
실사용 전환성	K-10	20%	실제 환경 적용 테스트 후 수정
학습 적합성	K-02, K-08, K-09	25%	데이터 다양성 확보, 편향 제거, 포맷 확장성 검토

DQ Score 등급	판정	조치
4.5 ~ 5.0	매우 우수	즉시 승인 가능
3.5 ~ 4.4	우수	경미한 개선 권고 후 승인
2.5 ~ 3.4	보통	수정 후 재검수 필요
1.5 ~ 2.4	미흡	주요 수정 필요
1.0 ~ 1.4	불량	재작업 필요

※ 위 등급 기준은 참고용입니다. 프로젝트 착수 시 운영위원회가 학습 목적에 맞게 기준을 확정합니다.

9장. 캘리브레이션 — 평가 전 준비

캘리브레이션이란 평가자들이 같은 기준으로 판단하도록 맞추는 사전 연습입니다. 이것을 하지 않으면 평가자마다 점수가 크게 달라져 결과를 신뢰할 수 없습니다.

1단계	샘플 준비 — 고품질·중간·저품질 각 5건씩 총 15건. 데이터 유형별로 별도 구성합니다.
2단계	독립 평가 — 평가자들이 각자 독립적으로 평가합니다. 서로 점수를 보거나 상의하면 안 됩니다.
3단계	공개 토론 — 점수 차이가 큰 항목을 중심으로 왜 그 점수를 줬는지 판단 근거를 공유합니다.
4단계	기준 합의 — 토론 결과를 기준 문서에 반영합니다. 모든 평가자가 확인 서명합니다.
5단계	앵커 포인트 저장 — 캘리브레이션 샘플의 합의 점수를 기준점으로 저장합니다. 이후 모든 평가의 기준이 됩니다.

캘리브레이션은 신규 프로젝트 착수 시마다, 그리고 반기(6개월)마다 정기 실시를 권장합니다. ICC 수치가 0.60 미만으로 떨어지거나 평가자가 교체되면 즉시 재실시합니다.

10장. 평가 중 인지 편향 통제

평가자가 자신도 모르게 빠지는 인지 편향을 사전에 알고 적극적으로 통제해야 합니다.

편향 유형	어떤 현상인가?	통제 방법
앵커링 편향	첫 번째 샘플이 기준점이 되어 이후 평가에 영향을 줌	캘리브레이션 샘플 순서를 매번 무작위로 섞어서 제시
후광 효과	'이 색감이 좋다'고 생각하면 다른 항목도 높게 줌	항목별로 화면을 분리해서 평가 — 한 번에 하나씩만 보이게
중심화 경향	피로해질수록 애매한 데이터에 3점을 반복 부여	세션당 최대 50건 제한, 중간에 15분 이상 의무 휴식
관대화·엄격화	특정 평가자가 일관되게 높거나 낮은 점수를 부여	ICC 실시간 모니터링 — 이탈 시 관리자 알림 및 재교육
도메인 한계	UX 전문가가 CMF 데이터를 평가하면 오류 가능성 높음	전문가-데이터 유형 적합도 매트릭스 사전 수립, 범위 외 배정 금지

11장. 문제 발견 시 피드백 방법

'이 데이터는 문제있음'이라고만 표시하는 것으로는 부족합니다. 어떤 문제인지, 왜 문제인지, 어떻게 고쳐야 하는지까지 전달해야 합니다.

문제 유형	구체적인 상황	수정 가이드
정합성 오류	이미지는 따뜻한데 레이블이 '차갑다'로 되어 있음	레이블을 실제 이미지 감성에 맞게 수정하거나 이미지를 교체
메타데이터 부재	파일명만 있고 카테고리·해상도·용도 필드가 없음	메타데이터 템플릿에 맞춰 필수 필드를 채워서 재제출
편향 발생	글로벌 데이터인데 특정 문화권 이미지만 집중됨	다문화적으로 해석 가능한 표현으로 교체하거나 다양성 보완
과다 중복	유사 이미지를 밝기·크롭만 바꿔 5장으로 늘림	실질적으로 다른 이미지로 교체하거나 중복 제거 후 재제출
확장성 부재	데스크탑용 배너를 모바일로 변환하면 텍스트가 잘림	반응형 레이아웃으로 재제작하거나 모바일 전용 버전 추가

우선순위	기준	처리 방법
즉시 수정	AI가 잘못된 방향으로 학습할 가능성이 높음	학습 데이터에서 즉시 제외 후 수정
다음 배치 전 수정	품질이 낮아지지만 심각한 오류는 아님	수정 후 재검수 요청
개선 권고	경미한 표현 차이, 개선하면 더 좋아짐	다음 데이터 제작 시 반영

12장. 역할별 체크리스트

현장 검수 담당자

- ☐ 오늘 평가할 데이터 유형(텍스트·사진·2D·3D)을 확인했는가
- ☐ 해당 유형의 파일별 체크포인트를 숙지했는가
- ☐ 적용할 기준(K-01~K-10) 중 이번 유형에 해당하는 항목을 확인했는가
- ☐ 캘리브레이션 합의 점수를 기준으로 숙지했는가
- ☐ 1회 세션 최대 50건 제한을 확인했는가
- ☐ 판단 근거를 반드시 기록한다 — '보통', '무난함' 같은 모호한 표현 금지
- ☐ 50건 초과 시 15분 이상 휴식 후 재개한다
- ☐ 내 전문 분야가 아닌 데이터가 배정되면 관리자에게 즉시 알린다
- ☐ 수정 완료 후 관리자에게 '수정 완료 + 재검수 요청' 메시지를 보낸다

데이터 관리자

- ☐ 데이터 유형별 분류 및 수량 확인
- ☐ AI 사전 검수 실행 및 이상 신호 데이터 목록 추출
- ☐ 위험도 분류 기준 적용: 신규 유형·AI 이상·레이블 모호·문화권 포함·반복 오류 → 고위험
- ☐ 전문가-데이터 유형 적합도 매트릭스 기준으로 배정표 작성
- ☐ 파일별 체크포인트 문서를 검수자에게 배포
- ☐ ICC 실시간 모니터링 — 0.60 미만 시 재검수 시행
- ☐ 4개 차원별 DQ Score 산출 후 취약 차원 파악
- ☐ 결과 보고서 초안 → 개선 방안 포함 공유 → 재검수 → 최종 보고서

부록. 용어 사전

용어	정의
KODIA 프레임워크	Heuristic-based Expert Design Data Quality Assessment Framework. 한국디자인산업연합회 디자인데이터 품질연구소가 개발한 AI 학습용 디자인 데이터 품질진단 체계.
DQ Score	Design Data Quality Score. 4개 평가 차원에 가중치를 곱해 합산한 종합 품질 점수. 5점 만점.
캘리브레이션	평가자들이 같은 기준으로 판단하도록 맞추는 사전 연습. 고·중·저 품질 샘플 각 5건(총 15건) 공동 평가 후 기준 합의.
ICC	Intraclass Correlation Coefficient(급내상관계수). 평가자 간 점수 일치도 지표. 0.75 이상 양호, 0.60 미만 재검수.
주석 신뢰성 (K-05)	라벨·태그·설명이 사람이 봐도 재현 가능하게 정의됐는지를 평가. '따뜻함'보다 'Warm 톤, 2700K'가 높은 점수.
구조화 가능성 (K-06)	검색·필터링·재사용·학습 파이프라인에 필요한 메타데이터가 체계적으로 붙어 있는지 평가.
편향 위험 (K-08)	특정 색·레이아웃·문화권·성별 표현이 과도하게 치우치지 않는지 평가.
AHP	Analytic Hierarchy Process(분석적 계층화 프로세스). DQ Score 가중치를 도출하기 위한 의사결정 기법.
배치 (Batch)	한 번에 제출되어 함께 검수되는 데이터의 묶음 단위. K-03, K-07, K-08은 배치 단위로 판단.
앵커링 편향	첫 번째 샘플이 기준점이 되어 이후 평가에 영향을 주는 인지 편향. 샘플 순서 무작위화로 통제.
후광 효과	한 항목 고점이 다른 항목에도 영향을 주는 인지 편향. 항목별 화면 분리로 통제.
중심화 경향	피로해질수록 중간값(3점)을 반복 부여하는 인지 편향. 세션당 50건 제한으로 통제.

참고 문서

- 2024년 인공지능 학습용 데이터 품질관리 가이드라인 v3.1 (과학기술정보통신부)
- Nielsen, J. (1994). Heuristic Evaluation. In Nielsen & Mack (Eds.), Usability Inspection Methods.
- ISO/IEC 25012:2008 — Data Quality Model