

우리는 디지털로 사회 현안을 해결하고, 국가 미래를 열어간다.

NIA 한국지능정보사회진흥원	보 도 자 료	다시 도약하는 대한민국 함께 잘사는 국민의 나라
보 도 일 시	온라인 배포시점부터 보도하여 주시기 바랍니다.	
배 포 일 시	담당팀	AI데이터사업팀
2025. 5. 26.(월) 09:00	담 당 자	유호진 팀장(053-230-4221) 윤주미 주임(053-230-4220)

한국지능정보사회진흥원, '인공지능(AI) 데이터 품질관리 가이드라인 v3.5' 발간

- 다중양식(멀티모달) 데이터, 합성데이터 등 고품질의 인공지능 데이터를 위한
품질관리 프레임워크 및 기준 제시 -

- 한국지능정보사회진흥원(원장 황종성, 이하 진흥원)은 과학기술정보통신부(장관 유상임, 이하 과기정통부) 와 함께 '인공지능(AI) 데이터 품질관리 가이드라인 v3.5(이하 가이드라인)'를 발간했다.
- 진흥원은 2021년 인공지능 학습용 데이터 품질관리 가이드라인 v1.0을 발간한 이후 매년 인공지능 기술 발전에 따라 고도화해 왔으며 2024년 공개된 가이드라인 v3.1은 6만건 이상의 다운로드를 기록하는 등 공공 및 민간에서 기준서로서 활발히 활용되고 있다.
- 가이드라인 v3.5는 2024년 '초거대AI 확산 생태계 조성사업'의 경험을 바탕으로 최신 인공지능 기술과 산업의 변화를 반영해 고도화하였다.
- 이는 총 3권으로 구성되어 있으며, ▲1권은 인공지능 데이터 품질관리 프레임워크 ▲2권은 데이터 구축 방법론 및 오류 대응 전략 실무사례 ▲3권은 생성형 인공지능 데이터에 특화된 품질관리 방법을 각각 제시하고 있다.

- 특히 이번 가이드라인은 언어모델(LLM) 뿐만 아니라 다중양식(멀티모달) 데이터*, 합성데이터 등 생성형 인공지능을 위한 데이터 특성에 맞춘 신규 지표와 구축 방법론을 별도 제시함으로써 빠르게 변화하는 인공지능 기술 동향에 유연하게 대응할 수 있도록 확장성 확보를 도모하였다.

* 다중양식(멀티모달) 데이터 : 텍스트, 음성, 동영상 등 두 가지 이상의 모달리티 (Modality)를 함께 활용하여 학습하는 데 필요한 데이터

< AI 데이터 품질관리 가이드라인 v3.5 구성 >

구분		내용
1권	AI 데이터 품질관리 가이드 v3.5	- 인공지능(AI) 데이터 품질관리 프레임워크 및 품질관리 지표 선정 절차
2권	AI 데이터 구축 가이드 v3.5	- 산업별·유형별 인공지능 데이터 구축방법 및 실무사례 제시
3권	생성형AI 데이터 품질관리 가이드 v2.0	- LLM, LMM, 합성데이터 등 생성형 인공지능(AI)을 위한 품질관리 지표 제시

- 본 가이드라인은 AI허브*(www.aihub.or.kr) 및 진흥원 누리집을 통해 누구나 내려받아 활용할 수 있으며, 권별로도 별도 제공하여 현장 실무자가 업무 및 관리 범위에 따라 편리하게 참조할 수 있도록 구성하였다.

* 과기정통부와 NIA가 운영하는 국내 최대 인공지능 학습용 데이터 제공 플랫폼

- 또한, 데이터 구성의 ‘잘된 사례 vs 부족한 사례’ 비교 분석과 체크리스트 제공 등을 통해 현장 중심의 데이터 품질관리에 실질적인 도움을 줄 것으로 기대된다.

- 진흥원 황중성 원장은 “인공지능 기술의 핵심은 데이터로 인공지능의 생명은 데이터 품질관리에 있다고 할 수 있다” 라며, “진흥원은 빠르게 발전하는 인공지능 기술에 대응하기 위하여 고품질 데이터 생태계 기반이 조성될 수 있도록 지속적으로 인공지능 데이터 품질 가이드라인 및 기준을 고도화하고 제시하겠다” 고 밝혔다. <끝>

<AI 데이터 품질관리 가이드라인 V3.5 표지>



[AI 데이터 품질관리 가이드라인 v3.5는 3권으로 구성되어 있다(왼쪽부터 1권, 2권, 3권)]

<인공지능 학습용 데이터 품질관리 프레임워크>



[품질관리 프레임워크는 AI 데이터의 구축 공정에 따른 프로세스, 산출물, 조직, 품질관리 활동을 제시한다]