

NIA 한국지능정보사회진흥원	보 도 자 료	 디지털뉴스 한국판뉴딜
보 도 일 시	온라인 배포시점부터 보도하여 주시기 바랍니다.	
배 포 일 시	담당부서	인공지능데이터사업1팀
2021. 3. 5.(금) 09:00	담 당 자	박현우 수석(02-6747-2179)

고품질의 인공지능 학습용 데이터 확보를 위한 품질관리 가이드라인 및 구축안내서 개발·보급

- 인공지능 학습용 데이터 구축 수행기관의 품질관리 역량 강화 -
- 21년 인공지능 학습용 데이터 구축사업부터 본격 적용 -

- 한국지능정보사회진흥원(원장 문용식, 이하 NIA)은 과학기술정보통신부(장관 최기영, 이하 과기정통부)가 추진하는 인공지능(AI) 학습용 데이터 구축사업의 품질 확보를 위해 인공지능 학습용 데이터 품질관리 가이드라인 v1.0(이하 가이드라인)과 인공지능 학습용 데이터셋 구축안내서(이하 구축안내서)를 공개했다고 밝혔다.
- 이번에 공개하는 가이드라인과 구축안내서는 '21년도 인공지능 학습용 데이터 구축사업부터 본격적으로 적용하여, 구축사업 참여기관의 품질관리 및 구축사업 수행 역량을 제고하는데 기여할 것으로 기대된다.
- 2017년부터 과기정통부와 NIA가 '인공지능 학습용 데이터 구축사업'을 통해 약 21종 460만 건의 인공지능(AI) 학습용 데이터를 구축하여 민간에 개방을 하였으나, 데이터의 품질이 사용자가 원하는 수준에 미치지 못한다는 지적이 있었다.

○ 이를 해결하기 위해 NIA는 데이터 생애주기를 기반으로 '계획-구축-운영-활용' 영역별 데이터 품질관리 활동을 정의한 가이드라인과 인공지능(AI) 학습용 데이터셋 구축안내서를 개발하였다.

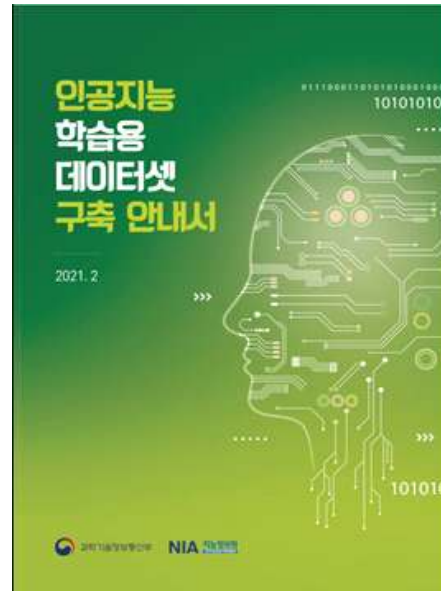
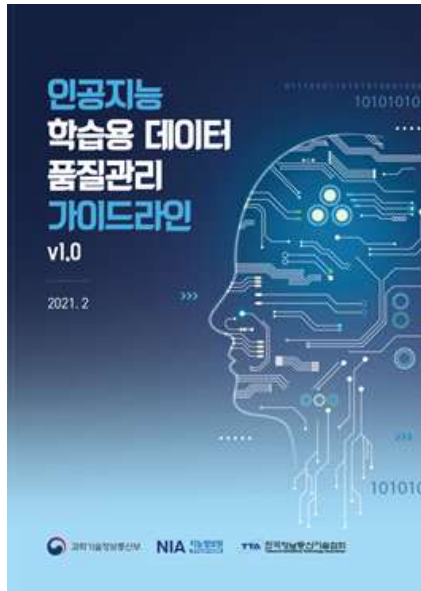
□ NIA 지능데이터본부 고윤석 본부장은 "개발된 가이드라인과 구축 안내서를 올해 인공지능(AI) 학습용 데이터 구축사업에 적용하고, 산·학·연 전문가 및 구축사업 수행기관, 참여기관의 의견을 적극 수렴하여 지속적으로 보완해 나갈 예정이다."라고 밝혔다.

○ 본 가이드라인과 구축안내서는 3월 5일(금)부터 NIA 누리집 (<https://www.nia.or.kr/>) 및 인공지능 학습용 데이터 통합 플랫폼 AI 허브(<https://aihub.or.kr/>)를 통해 내려받을 수 있다. <끝>

붙임. 인공지능 학습용 데이터 품질관리 가이드라인 및 데이터셋 구축 안내서 소개

인공지능 학습용 데이터 품질관리 가이드라인 및 데이터셋 구축 안내서 소개

《 인공지능 학습용 데이터 품질관리 가이드라인 및 데이터셋 구축 안내서 》



- 수행기관 및 참여기관이 품질관리 체계, 품질관리 기준, 품질검사 방법을 수립하고 수행할 수 있는 ‘인공지능 학습용 데이터 품질관리 가이드라인 v1.0’ 마련
 - (총론) 인공지능 학습용 데이터의 품질관리의 기본 개념, 가이드라인 개발 배경 및 목적과 전반적 구성체계와 가이드라인 활용방법 설명
 - (품질관리체계) 인공지능 학습용 데이터 생애주기에 따라 계획, 구축, 운영, 활용 등으로 품질관리 영역을 구분하고, 각 영역별 주요 품질관리 활동 설명
 - (품질관리기준) 데이터 및 프로세스 관점에서 인공지능 학습용 데이터가 적절한 수준의 품질을 유지하고, 구축 및 운영과정에서 품질관리 활동이 적절한지 판단할 수 있는 품질기준 제시
 - (품질검사방법) 인공지능 학습용 데이터 구축단계와 운영단계를 구별하여 제시한 품질관리기준을 이용한 품질검사 방법 설명

- 인공지능 학습용 데이터 구축사업에 참여하는 수행 및 참여기관이 사업 초기에 수립하는 '인공지능 학습용 데이터 구축·활용 가이드라인' 작성시 반드시 고려해야 할 내용을 수행할 수 있는 '인공지능 학습용 데이터셋 구축 안내서' 마련
 - '인공지능 학습용 데이터셋 구축 안내서'는 크게 텍스트, 음성, OCR(광학문자인식) 이미지, 영상(동적/정적 이미지) 등 4개의 데이터 유형별로 구분하여 제공
 - 또한, 인공지능 학습용 데이터셋 구축 시 고려해야 할 내용으로 텍스트, OCR 이미지, 자율주행, 영상 데이터 유형별 라벨링 및 어노테이션 공통참조 항목을 도출하여 '공통참조기준' 제시