내 삶을 지키는 정부 국민안전플러스

산업통상부 보도자료 대체불가 대한민국

보도시점 2026. 11. 27.(금) 06:00 < 11.27.(금) 석간 > 배포 2026. 11. 26.(목)

AI 로봇이 배우는 동작 데이터, 국가표준(KS)으로 체계화 추진

- 학습 데이터 공유·활용을 위한 데이터 구조·생성·저장에 대한 KS 제정예고 -

산업통상부 국가기술표준원(원장 김규성, 이하 국표원)은 휴머노이드 등 AI 로봇의 학습 데이터 공유·활용 촉진을 위해「로봇 조작 학습을 위한 데이터셋*」 KS 제정안을 11월 27일부터 60일간 예고고시하고 산업계 의견을 수렴한다고 밝혔다.

* 데이터셋(Dataset) : 로봇의 개발, 학습, 시험 및 평가에 활용하기 위하여 하나 이상의 센서 또는 외부 장치로부터 획득된 데이터 집합

AI 로봇이 산업 현장과 일상생활로 확대가 예상되면서 로봇 학습 데이터의 중요성이 커지고 있으나, 로봇마다 데이터 구조와 속성이 달라 기업 간 공유·재사용에 어려움이 있었다. 국표원은 아직 로봇 학습 데이터셋에 적용 가능한 국제표준이 확립되지 않았으나, 국내 AI 로봇 산업 발전과 표준화 시급성을 고려해 국가표준을 선제적으로 마련하기로 했다.

표준안은 데이터셋 구축을 위해 계획·수집·전처리·저장 등 생성 단계별 요구사항을 제시하고, 작업 수행 환경, 내용, 과정 등을 기록할 수 있도록 데이터 구조를 정의한다. 특히 작업 과정에서 수집한 영상·음향 등 센서 정보와 로봇의 상태·행동 데이터를 시간 순서와 일치하도록 최소 사항을 규정한다. 또한 데이터의 저장과 메타데이터* 형식을 체계화하여 로봇 간 학습 데이터를 공유하고 활용할 수 있게 지원한다.

* 메타데이터(Metadata) : 데이터 수집환경, 종류, 저장 위치 등 내용과 속성을 설명하는 데이터

이번 표준안은 로봇 분야 산학연 전문가로 구성된 지능형로봇표준포럼(KOROS)*에서 산업부 로봇 연구개발 과제 결과를 토대로 마련되었으며, 국표원이 AI 로봇 기업 및 전문가 협의, 기술 검토 등을 거쳐 구체화하였다.

* LG전자, 삼성전자, HD현대로보틱스, 두산로보틱스 등 38개 기업, 170여명 전문가 참여

국표원은 예고고시를 통해 산업계 의견을 반영하고 내년 상반기 내에 KS 제정을 완료할 계획이다. 제정된 표준은 향후 휴머노이드 실증 보급, 로봇 데이터팩토리 등에서 로봇 학습용 데이터를 일관되고 통합된 형태로 수집함으로써, AI 로봇 개발 성과 제고에 기여할 것으로 기대된다.

아울러 국표원은 동 표준안 외에 로봇 학습 데이터의 완전성, 정확도 등 품질 검증 및 평가 방법에 대한 표준을 내년 초에 개발하고, 이를 국제표준으로 제안할 예정이다.

김규성 국표원장은 “이번 표준안은 로봇 R&D 성과를 바탕으로 산업계가 제안한 수요를 KS 표준안으로 개발시킨 사례로, 현장에서 실질적으로 적용될 수 있다는 데 의미가 있다”며, “앞으로도 R&D와 산업계 수요를 연계해 현장에 필요한 표준을 지속적으로 발굴하겠다”고 밝혔다.

담당 부서	표준정책국 기계융합산업표준과	책임자	과 장	김용득 (043-870-5370)
		담당자	연구사	박종훈 (043-870-5374)
	산업인공지능정책관 인공지능기계로봇과	책임자	과 장	이동철 (044-203-4310)
		담당자	서기관	방규철 (044-203-3921)

대한민국 정책브리핑
대체불가 대한민국
OPEN 출처표시 공공누리 공공저작물 자유이용허락

첨부 **「로봇 조작 학습을 위한 데이터셋」 표준안 개요**

□ 로봇의 조작 학습에 필요한 데이터셋의 구조·생성·저장 형식을 표준화

데이터 생성 및 관리 절차 (데이터 생애주기)

- 학습 목적에 적합한 데이터셋 구축을 위해 계획 → 수집 → 전처리 → 저장 단계별 요구사항을 제시

계획	→	수집	→	전처리	→	저장
데이터셋 목적, 범위 설정		원시 데이터 확보		원시 데이터 가공		데이터셋 구조에 저장

학습 데이터셋 계층 구조 (데이터 구성 단위)

- 로봇의 조작 학습을 위한 데이터셋을 체계적으로 구성하기 위한 기본 계층 구조를 정의

데이터셋	→	시나리오	→	작업	→	에피소드	→	단계
특정 목적 학습을 위한 데이터 모음		작업 환경과 상황 (공장, 가정, 주방 등)		로봇 조작 작업 (컵 집기, 문열기 등)		작업을 수행한 전체 기록 (강화학습)		시간에 따른 개별 데이터 (관찰, 로봇상태 등)

데이터 유형 및 속성 정의 (데이터 내용)

- 로봇의 조작 학습에 필요한 다양한 데이터 유형과 속성을 정의하고, 데이터 간 시간 동기화를 규정

시계열 데이터	시각 데이터 (Videos)	음향 데이터 (Audios)	라이다 데이터 (Sensors)
관찰, 로봇 상태 등 (위치, 속도, 토크)	카메라별 영상	마이크 음성, 환경음	센서로 수집한 데이터

저장 구조 및 파일 형식 (데이터 표현)

- 학습 데이터셋의 상호 운용성과 재사용성을 위해 디렉터리 구조와 메타데이터 명세 정의

[데이터셋 디렉터리 구조 예시]

```
<dataset_root>/
├── meta/
│   ├── info.json                          # 전역 명세 (data_path, video_path, features)
│   ├── dataset.json                       # 데이터셋 확장 명세, 시나리오, 임무
│   ├── robot.json                         # 로봇, 센서, 좌표 프레임 명세
│   ├── episodes/chunk-000/file-000.parquet   # A.4 에피소드 메타데이터 + 통계
│   ├── tasks.parquet                  # 자연어 태스크 ↔ 정수 ID 매핑 (data_format을 따름)
│   └── stats.json                         # 데이터셋 전역 통계값
├── data/
│   └── chunk-000/
│       ├── file-000.parquet       # step 단위 (observation.state, action, reward, is_*)
│       └── file-001.parquet
├── videos/
│   ├── observation.images.cam_front_right/
│   │   └── chunk-000/file-000.mp4
│   └── observation.images.cam_wrist/
│       └── chunk-000/file-000.mp4
├── audios/
│   └── observation.audio.mic_contact_fixture/
│       └── chunk-000/file-000.flac
├── sensors/
│   └── observation.pointclouds.lidar_top_360/
│       └── chunk-000/file-000.parquet
└── README.md
```