

# 015 데이터 라벨링

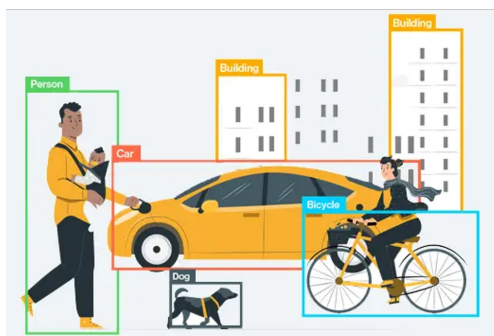
Data Labeling

## 데이터에 의미나 정답을 부여해 AI 학습이 가능하도록 만드는 과정

- AI가 이미지·음성·텍스트 등 데이터를 올바르게 인식하고 판단할 수 있도록 정보를 부여하는 절차
- AI 모델의 정확도와 신뢰성을 좌우하는 학습 데이터 품질 관리의 핵심 단계

### 데이터 라벨링의 개념

데이터 라벨링은 AI 학습에 활용되는 데이터에 사람이 의미나 정답 정보를 부여하는 과정입니다. 예를 들어 이미지에는 사물의 이름을, 음성에는 문장을, 텍스트에는 감정이나 의도 정보를 표시해 AI가 이를 학습하도록 돕습니다. 이 과정을 통해 AI는 단순한 데이터의 모양이나 패턴을 넘어서, 그 속에 담긴 의미를 인식하고 분류·예측 등의 작업을 수행할 수 있습니다. 라벨링은 AI가 세상을 이해하기 위한 언어를 가르치는 일과 같으며, 모델의 인식력과 판단력을 결정하는 출발점이 됩니다.



출처: LXT

### 데이터 라벨링의 과정

데이터 라벨링은 수집된 데이터를 분석하고, 기준에 따라 범주를 정의한 뒤, 각 데이터에 올바른 정보를 부여하는 순서로 진행됩니다. 이 작업은 사람이 직접 수행하기도 하지만, 최근에는 AI가 초기 분류를 제안하고 사람이 이를 검수하는 오토라벨링(auto labeling) 방식이 널리 쓰입니다. 특히 대규모 데이터셋에서는 품질 관리가 중요해, 여러 명이 같은 데이터를 반복 검토하는 다중 검증 절차나 표준화된 가이드라인이 함께 적용됩니다. 정확한 라벨링이 이루어져야 모델이 오류 없이 학습할 수 있으며, 잘못된 라벨은 학습 방향을 왜곡시켜 성능 저하를 초래할 수 있습니다.

### 데이터 라벨링의 역할

데이터 라벨링은 AI가 데이터의 숫자를 실세계의 개념과 연결하기 위한 과정으로, 단순한 데이터 정제 단계를 넘어 의미 구조를 형성하는 역할을 합니다. 전처리가 데이터의 품질과 형식을 다듬는 과정이라면, 라벨링은 그 데이터가 '무엇을 뜻하는가'를 정의하는 지식 부여 단계입니다. 올바르게 라벨링된 데이터는 AI가 맥락과 관계를 학습해 인간의 판단과 유사한 인식 능력을 갖추도록 만듭니다. 또한 체계적인 라벨링은 편향된 데이터 해석을 방지하고, AI가 사회적 맥락 속에서 공정하게 작동하도록 돕는 토대가 됩니다.