

## 02 데이터와 학습

### 학습 목표

- 학습에 필요한 데이터를 수집하여 분류할 수 있다.
- 인공지능 문제 해결 과정을 살펴보고 인공지능의 학습에서 데이터의 중요성을 설명할 수 있다.



### 1 인공지능에서 데이터의 역할

인공지능은  
사람보다 많은  
양의 학습 데이터를  
필요로 해요.



사람은 지식과 다양한 경험을 통해 대상을 인식하고 결과를 예측한다. 그리고 이러한 경험이 누적되면 점점 더 정확한 인식과 예측이 가능해진다.

사람이 경험을 통해 대상을 인식하고 결과를 예측하는 것처럼 인공지능은 데이터 학습을 통해 대상을 인식하고 결과를 예측한다. 많은 양의 데이터로 학습을 반복하면 인공지능의 성능이 점차 개선되며, 이때 데이터는 인공지능이 제 역할을 하도록 하는 재료가 된다.

#### « 로봇 청소기 »



- ④ 사물 이미지 데이터를 학습하여 감지한 사물이 무엇인지 알 수 있다.

#### « 인공지능 스피커 »



- ④ 소리 데이터를 학습하여 감지한 음성이 어떤 내용을 뜻하는지 파악하여 응답할 수 있다.

#### « 온라인 쇼핑물 »



- ④ 텍스트 데이터를 학습하여 텍스트의 의미가 무엇인지 파악할 수 있다.

#### « 인공지능 카메라 »



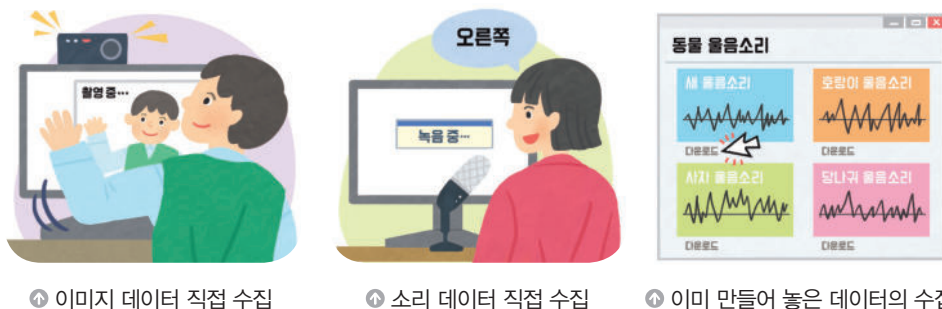
- ④ 움직임 데이터를 학습하여 움직임을 판단할 수 있다.

[그림 IV-3] 생활 속 인공지능에서 데이터의 역할

## 2 데이터의 수집과 분류

### 1 데이터 수집

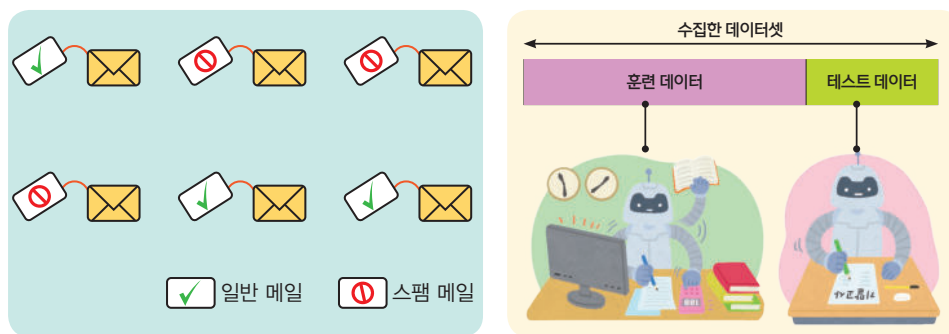
인공지능의 학습 데이터를 수집하는 방법에는 직접 수집하는 방법과 검색 엔진이나 데이터 전문 웹 사이트를 이용하는 방법이 있다. 이미지, 소리, 텍스트 등의 데이터는 카메라나 마이크 등을 통해 직접 수집하거나 미리 생성해 둔 데이터를 공유하는 웹 사이트, 검색 엔진을 통해 수집할 수도 있다.



[그림 IV-4] 인공지능의 학습 데이터 수집 방법

### 2 데이터 분류

수집한 각각의 데이터가 어떤 결과를 예측하는 데 사용되는 데이터인지 +레이블링 작업을 한다. 그리고 인공지능 모델의 학습에 사용할 훈련 데이터와 학습이 잘 되었는지 평가할 때 사용할 테스트 데이터로 분류한다.



① 일반 메일과 스팸 메일 레이블링 예시

② 모델 학습에 사용할 훈련 데이터와 테스트 데이터의 분류

[그림 IV-5] 인공지능에 사용할 데이터 분류 예시

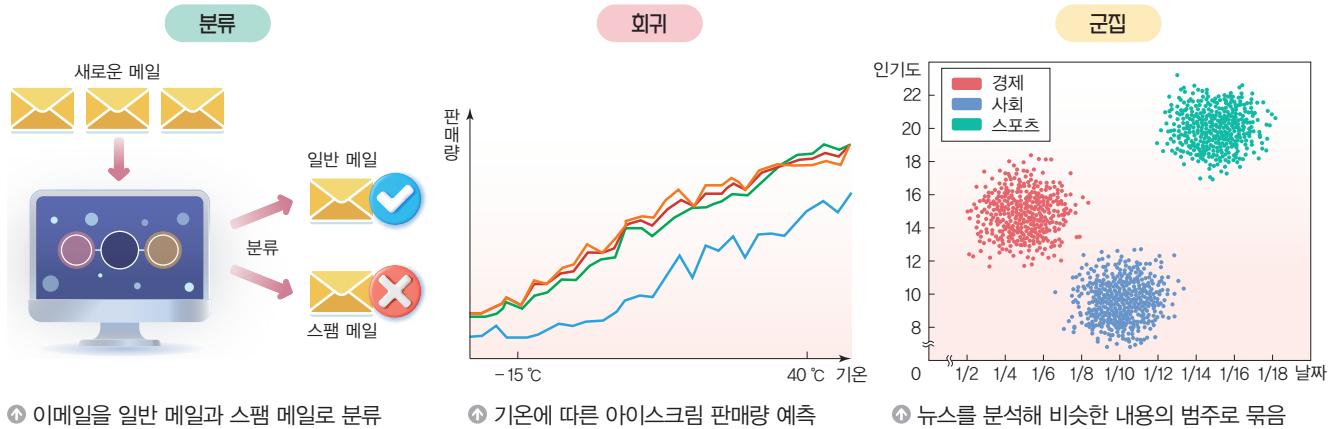
이때 테스트 데이터는 학습에 사용하는 훈련 데이터와 명확히 구분해야 한다. 학습에 사용한 데이터를 이용하여 인공지능 모델을 평가한다면 이미 정답을 학습한 상태이므로 올바르게 평가할 수 없다. 테스트 데이터로 결과를 제대로 예측하지 못한다면 학습이 부족했을 가능성이 있으므로 훈련 데이터를 추가하여 다시 학습시켜 성능을 향상시키도록 한다.

#### + 레이블링

데이터에 부여되는 특정한 태그 또는 정답을 레이블이라고 하며, 인공지능 모델을 생성할 때 인공지능이 예측하게 되는 결과값을 의미한다. 이러한 레이블을 부여하는 과정을 레이블링이라고 한다.

### 3 인공지능으로 해결할 수 있는 문제

인공지능은 분류, 회귀, 군집 등의 문제를 해결하는 데 주로 활용된다. 분류(classification)는 입력 데이터가 미리 정의된 범주 중 어느 것에 속하는지 결정하는 것을 말한다. 회귀(regression)는 입력 데이터에 따른 수치 형태의 결과값을 예상하는 것이고, 군집(clustering)은 입력 데이터를 유사한 것끼리 모아 동일 집단으로 만드는 것을 뜻한다.



[그림 IV-6] 인공지능으로 문제를 해결하는 방법

## 해 보기

### 인공지능으로 해결할 수 있는 문제 유형 파악하기

조사 < 탐구 > 실습 토의 발표



1 다음의 문제를 해결할 때 사용할 수 있는 방법은 무엇인지 보기에서 찾아보자.

보기

㉠ 분류

㉡ 회귀

㉢ 군집

- (1) 광고비의 지출에 따른 물건 판매량을 예상한다. ( )
- (2) 영화 감상평을 분석하여 긍정과 부정을 파악한다. ( )
- (3) 동영상 시청 기록 정보로 비슷한 성향의 사람들을 묶어 새로운 동영상상을 추천한다. ( )

2 다음 조건을 만족하는 게임을 만들려고 할 때 어떤 문제 해결 방법을 이용하는 것이 좋을지 생각해 보고, 그 이유를 적어 보자.

조건

‘고요 속의 외침’이라는 게임처럼 해당 단어를 말 없이 입 모양으로 표현하면 컴퓨터가 맞추는 게임을 만들려고 한다.

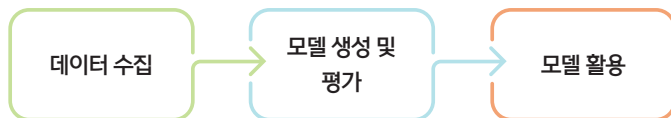
---



---

## 4 인공지능을 활용한 문제 해결 과정

인공지능을 활용하여 문제를 해결할 때에는 데이터 수집, 모델 생성 및 평가, 모델 활용의 과정을 거친다. +티처블 머신을 활용하여 문제 해결 과정에 따라 의자에 앉았을 때 바른 자세를 유지하도록 돕는 +인공지능 모델을 생성해 보자.



### 1 데이터 수집

데이터 수집 단계에서는 문제 해결에 필요한 이미지, 소리, 문자 등의 데이터를 수집하고, 데이터가 인공지능 모델 학습에 활용할 만한 가치가 있는지 분석한다. 바른 자세 유지 시스템을 만들기 위해 사람이 의자에 앉아 있는 이미지 데이터를 수집하고, 바른 자세와 바르지 못한 자세로 레이블링한 후 훈련 데이터와 테스트 데이터로 분류한다. 의자에 앉은 자세와 관련이 없거나 모델이 인식하기 어려운 이미지는 삭제한다.

#### + 티처블 머신

이미지, 자세, 소리를 학습시켜 인공지능 모델을 만들 수 있는 웹 기반 도구를 말한다.

#### + 인공지능 모델

데이터에서 패턴을 학습하고 이를 바탕으로 예측 또는 결정하는 프로그램을 인공지능 모델이라고 한다.

#### 활동 도구미

인공지능 모델의 성능은 학습에 사용되는 데이터가 중요한 역할을 하므로 양질의 데이터를 수집해야 한다.



[그림 IV-7] 데이터 수집 단계

## ➤ 포즈 프로젝트



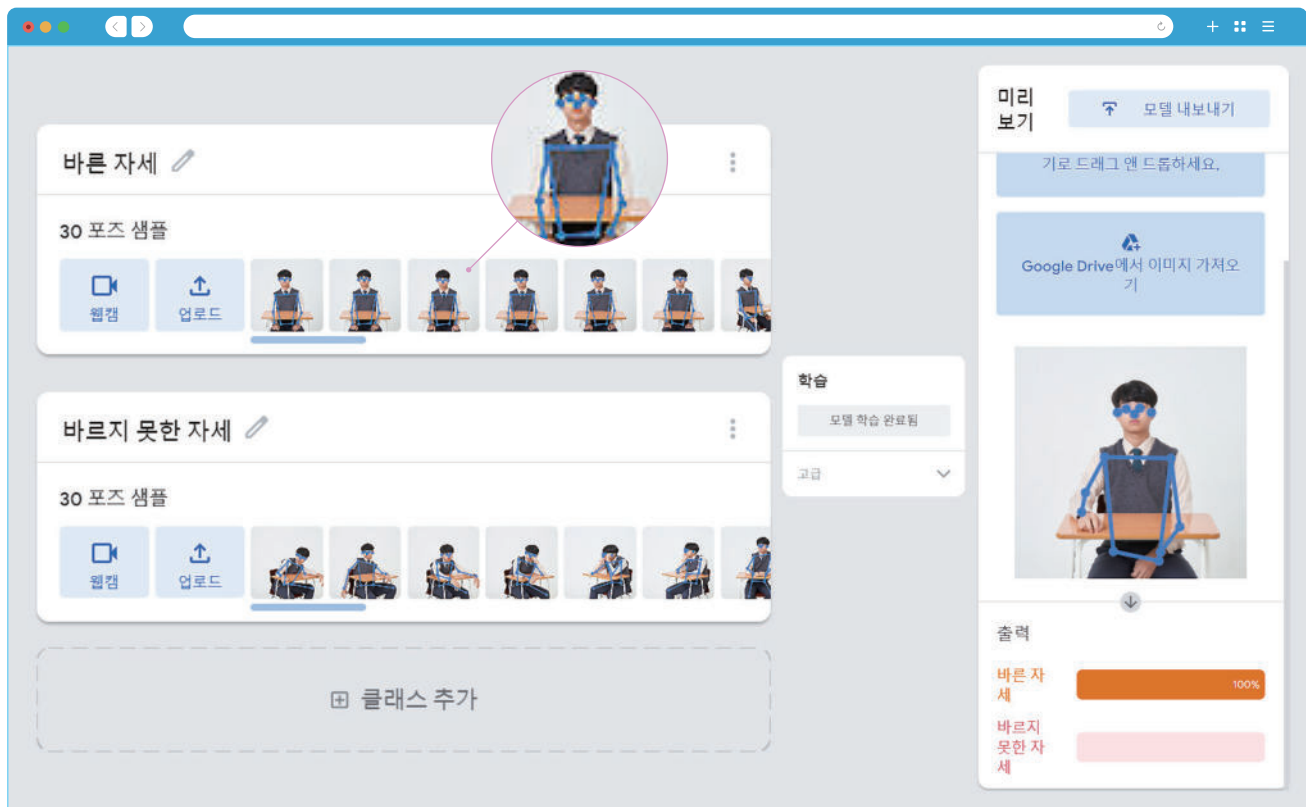
티처블 머신의 포즈 프로젝트는 눈, 코, 입뿐만 아니라 각 관절의 위치를 점으로, 관절 간의 관계를 선으로 나타낸다. 관절의 위치와 거리를 학습하여 모션을 분류하는 인공지능 모델을 만든다.

## ② 모델 생성 및 평가

모델 생성 및 평가 단계에서는 수집한 데이터를 활용하여 인공지능 모델을 학습시키고, 학습이 잘 되었는지 평가한다.

바른 자세를 유지하도록 돕는 인공지능 모델을 생성하기 위해 티처블 머신(<https://teachablemachine.withgoogle.com>)에 접속한 후 ➤‘포즈 프로젝트’를 선택해 보자.

클래스명을 ‘바른 자세’, ‘바르지 못한 자세’로 설정하고 수집한 훈련 데이터를 각 클래스에 맞게 업로드한 후 모델을 학습시킨다. 학습이 완료되면 학습에 사용하지 않은 테스트 데이터를 이용하여 모델이 이미지를 정확하게 인식하는지 평가한다. 모델이 올바르게 동작하지 않는다면 데이터를 추가로 학습시켜 성능을 높인다.



[그림 IV-8] 모델 학습(티처블 머신)

## ③ 모델 활용

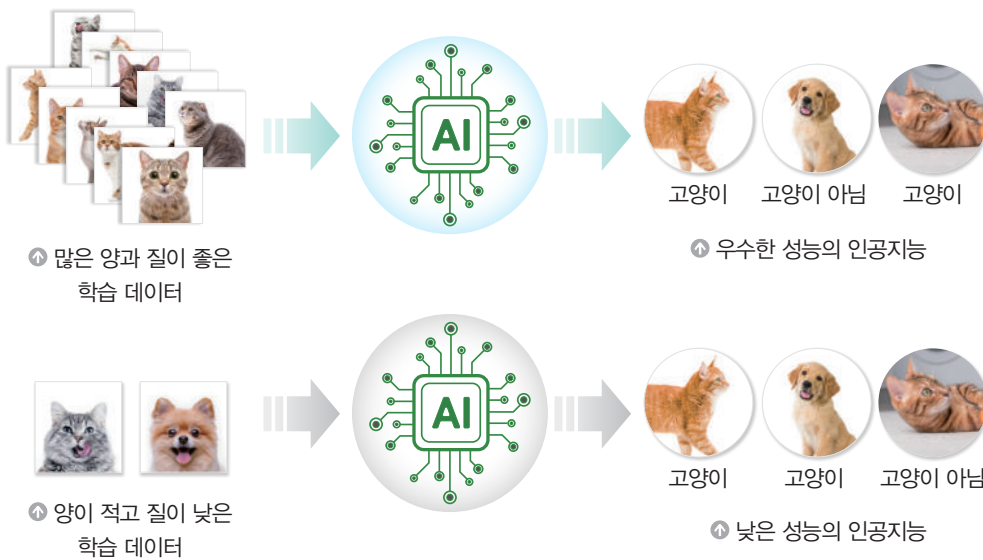
모델 활용 단계에서는 완성된 인공지능 모델을 프로그래밍을 통해 다양한 용도로 활용할 수 있다. 바른 자세 유지 시스템은 사람의 자세를 계속 관찰하다가 바르지 못한 자세가 되면 바른 자세로 앉으라는 신호를 출력하는 프로그래밍을 통해 구현할 수 있다.



## 5 인공지능 학습에서 데이터의 중요성

인공지능에서 데이터는 인공지능이 올바른 기능을 수행할 수 있도록 하는 재료와 같다. 만약 앞에서 살펴본 자세를 구분하는 모델의 학습에 바른 자세 클래스에 잘못된 자세 이미지나 관절의 위치가 정확히 드러나지 않는 이미지가 섞여 있다면 결과는 어떻게 될까? 또 너무 적은 양의 이미지로 학습하였다면 결과를 제대로 예측할 수 있을까?

많은 양의 정확한 데이터로 학습하였다면 우수한 성능의 인공지능을 생성할 가능성이 높지만, 양이 적고 정확하지 않은 데이터로 학습한 인공지능은 정확한 결과를 산출할 수 없다.



[그림 IV-9] 인공지능 학습에서 양과 질에 따른 데이터의 중요성을 나타낸 예시

### 지식 충전소

#### 인공지능으로 복원한 독립운동가

오른쪽은 독립운동가였던 백범 김구 선생의 사진을 이용하여 움직임을 구현한 영상의 한 장면이다. 이 영상에서 백범 김구 선생은 눈을 깜빡이기도 하고 고개를 돌리기도 해 마치 살아 있는 듯한 생생함을 느끼게 해 준다.

인공지능 딥페이크 기술은 컴퓨터 비전을 활용하여 실존했던 인물의 얼굴이나 목소리를 사실적으로 재현해 낼 수 있다. 이 기술을 이용한 역사적 인물의 복원은 역사적 사건으로의 관심을 불러일으키고 당시 상황의 이해를 높이는 데 사용될 수 있다. 또한, 역사적 인물들의 삶을 새로운 시각에서 접근할 수 있게 해 준다.

(<https://www.moneys.co.kr/article/2021030415068043679?outlink=1>)



인공지능으로 구현한  
백범 김구 선생의 모습

다음과 같은 상황의 문제를 해결하는 과정을 적어 보자.

건물 안에서 운동화를 신고 다니면 실내가 더러워지므로 깨끗한 환경을 위하여 실내화 신기를 권장한다. 건물 안에서 운동화를 신고 다니는 학생이 있으면 자동으로 신호를 출력해 주는 인공지능 모델을 만들어 보자.



1 문제 해결에 필요한 데이터를 수집하여 레이블링해 보자.

데이터 수집

데이터 레이블링

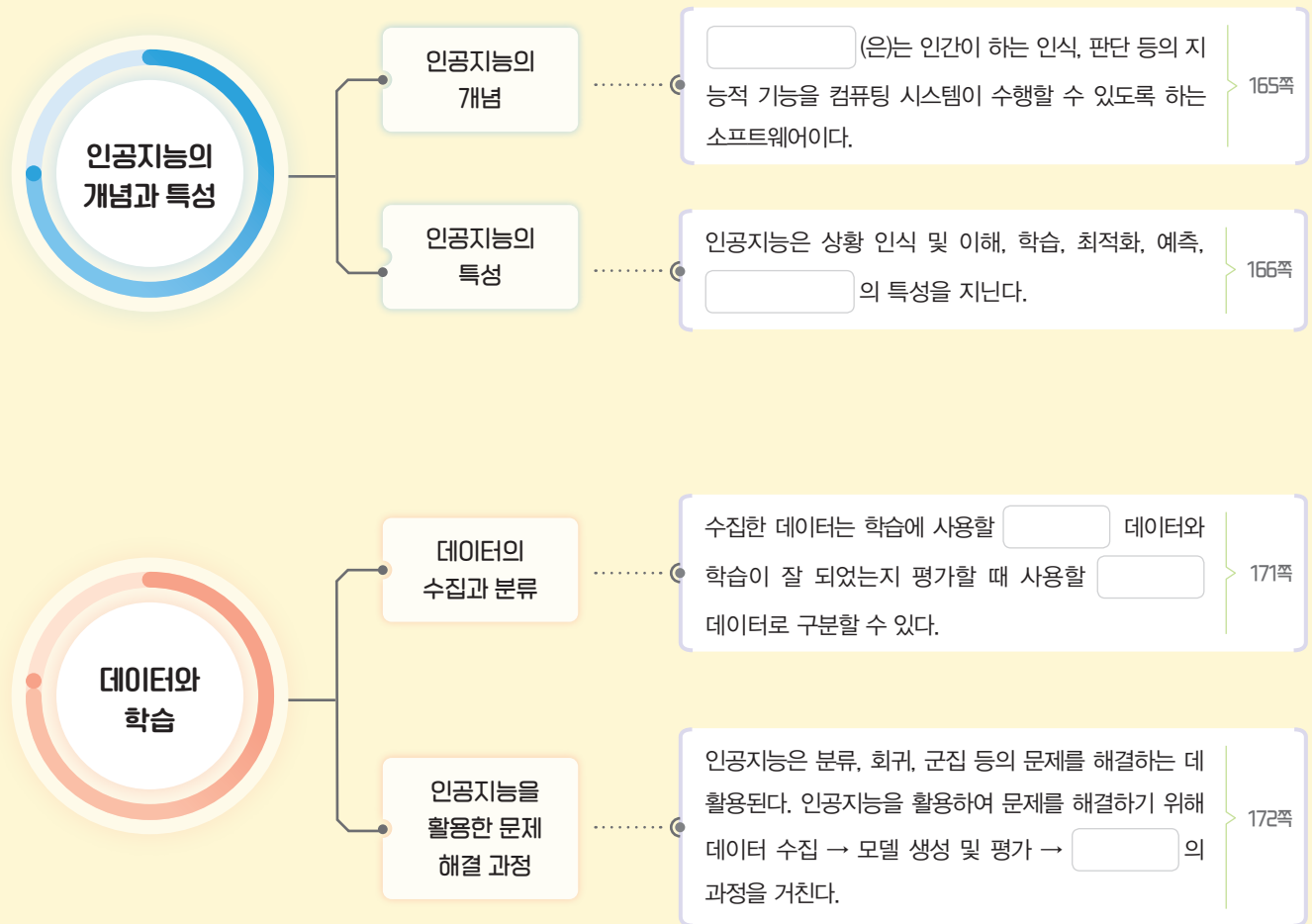
2 수집한 데이터를 훈련 데이터와 테스트 데이터로 분류해 보자.

훈련 데이터

테스트 데이터

3 훈련 데이터로 인공지능 모델을 학습시키고, 테스트 데이터로 인공지능 모델의 성능을 평가해 보자.

4 만약 인공지능 모델의 성능이 좋지 않다면 개선 방법을 찾아 성능을 개선해 보자.



## 나의 성취 수준



| 교과 역량   | 평가 항목                             | 성취 수준 |
|---------|-----------------------------------|-------|
| 인공지능 소양 | 인공지능의 개념과 특성을 설명할 수 있는가?          | ☆☆☆☆☆ |
|         | 인공지능과 인공지능이 아닌 소프트웨어를 구별할 수 있는가?  | ☆☆☆☆☆ |
|         | 인공지능 학습에서 데이터의 중요성을 설명할 수 있는가?    | ☆☆☆☆☆ |
|         | 인공지능 학습에 필요한 데이터를 수집하여 관리할 수 있는가? | ☆☆☆☆☆ |