

04 인공지능의 윤리적 문제 해결

학습 목표

- 인공지능 시스템을 선택하여 문제를 해결하는 과정에서 인공지능의 사회적 역할과 가치를 판단할 수 있다.
- 데이터를 수집·활용할 때 발생할 수 있는 윤리적 문제의 해결 방안을 제시할 수 있다.



1 인공지능 프로그램 개발 시 주의할 점

인공지능 모델은 개발 및 활용하는 과정에서 사회적·법적·윤리적 문제가 생기기기도 한다. 인공지능 모델을 생성할 때 여러 가지 요소들을 고려하지 않으면 개인뿐만 아니라 사회에도 부정적인 영향을 미칠 수 있다. 다음의 문제 상황을 살펴보고 인공지능 프로그램을 개발하는 문제 해결 과정에서 단계별로 유의해야 할 사항을 점검해 보자.

● 문제 정의 단계의 고려 사항

문제의 해결 방안을 탐구하는 단계에서 사람에게 이로운 인공지능 프로그램인지 확인하고 개발 여부를 판단해야 한다. 다음의 문제 상황을 살펴보고 머리 모양 추천 인공지능 프로그램이 사람들에게 미칠 영향을 생각해 보자.

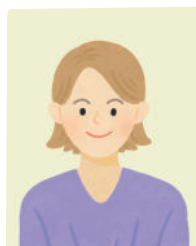
문제 상황

사람마다 얼굴형이 다르고, 이에 따라 어울리는 머리 모양도 다르다. 내가 어떤 얼굴형인지 구분하여 어울리는 머리 모양을 알려 주는 인공지능 시스템을 만들어 보자.

이 머리 모양은 어떤 얼굴형에 더 잘 어울릴지 궁금해.




나에게 어떤 머리 모양이 잘 어울릴지 궁금해.



● 데이터 수집 시 주의 사항

문제 해결을 위해 데이터를 수집할 때에는 다음과 같은 사항에 주의해야 한다.

① 개인 정보와 저작권

문제 해결에 필요한 데이터를 수집할 때에는 개인 정보나 저작권 침해 문제를 주의해야 한다. 수집하려고 하는 데이터가 개인 정보를 포함하고 있을 때는 개인 정보 수집·이용에 대한 동의서를 받은 후  비식별화 처리를 해야 하며, 저작물은 저작권자로부터 사용을 허락받아야 한다.

예를 들어 머리 모양 추천 인공지능 모델을 학습시키려면 얼굴 이미지 데이터가 필요한데, 이것은 개인 정보에 해당한다. 따라서 수집 목적과 항목 등을 설명하고 개인 정보 수집·이용 동의서를 받은 후 수집해야 한다.

비식별화

인공지능 모델을 만들 때 개인을 식별할 수 있는 정보를 제거하거나 변형하는 방법이다. 예를 들어, 이름을 가명으로 처리하거나 삭제하는 것 등이 있다.

해 보기

개인 정보 수집·이용 동의서 작성하기

조사 < 탐구 > 실습 < 토의 > 발표



다음은 머리 모양 추천 인공지능 모델의 학습에 필요한 데이터를 수집하기 위한 개인 정보 수집·이용 동의서이다. 수집 항목을 확인하고 수집해야 할 개인 정보 항목과 목적이 적절한지 이야기해 보자.

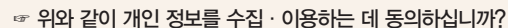
개인 정보 수집·이용 동의서

얼굴형에 맞는 머리 모양을 추천하는 인공지능을 개발하기 위하여 다음과 같이 개인 정보를 수집·이용하고자 합니다. 내용을 확인하신 후 동의 여부를 결정하여 주십시오.

□ 개인 정보 수집·이용 내역

| | |
|-------|---------------------------|
| 수집 항목 | 이름, 나이, 얼굴, 거주지, 연락처 |
| 수집 목적 | 얼굴형에 맞는 머리 모양 추천 및 이벤트 홍보 |
| 보유 기간 | 모델 학습 완료 후 이벤트 종료 시까지 |

위의 개인 정보 수집·이용 동의를 거부할 권리가 있습니다. 그러나 동의를 거부할 때에는 서비스 혜택에 제한을 받을 수 있습니다.



☐ 예 ☐ 아니요

년 월 일

본인 성명 (서명 또는 인)

보호자 성명 (서명 또는 인)

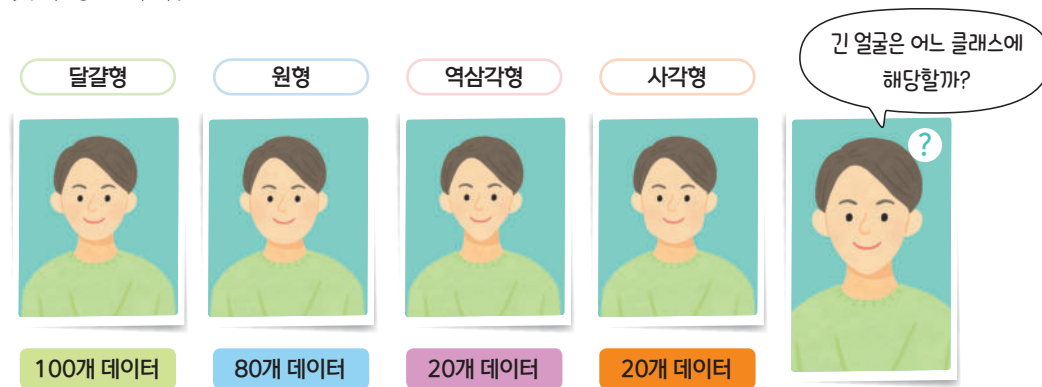
2 데이터의 편향성

데이터를 수집할 때에는 데이터가 편향되지 않도록 주의해야 한다. 머리 모양을 추천하는 프로그램에 달걀형, 원형, 역삼각형, 사각형의 클래스로 얼굴형을 구분하였다고 가정해 보자.

긴 얼굴을 어느 클래스에 포함하여 학습시킬지는 사람의 판단에 따라 달라지기 때문에 여기서 데이터의 편향이 발생할 수 있다. 이러한 때에는 긴 얼굴 클래스를 새로 생성하는 등과 같이 명확한 구분을 제시함으로써 편향을 최소화해야 한다. 하나의 클래스에만 많은 데이터를 학습시키면 모델 활용 시 데이터가 많았던 클래스로 예측할 가능성이 높으므로 클래스별로 데이터 양의 균형을 맞추는 것이 중요하다.

▶ 데이터 편향성

인공지능이 학습한 데이터가 고르지 못하고 한쪽으로 치우치는 것을 말한다.



[그림 IV-14] 데이터 편향 가능성 예시

● 인공지능 모델 활용 시 주의 사항

인공지능 모델을 활용하는 과정에서 개인 정보 유출과 사생활 노출로 피해를 줄 가능성은 없는지 다양한 요소를 검토해야 한다. 또한, 모델 활용 과정에서 예상하지 못한 문제가 생길 수도 있으니 다양한 관점에서 모델 활용 방법을 점검해야 한다.

해 보기1

인공지능의 데이터 편향과 사회적 영향 고려하기

< 조사 > 탐구 실습 토의 < 발표 >



1 다음 중 데이터 편향이 발생할 가능성이 높은 것은 ○, 그렇지 않은 것은 ×로 표시해 보자.

- (1) 인공지능 스피커에게 사람의 질문이 어떤 의미인지 파악할 수 있도록 표준어를 중심으로 학습시켰다. ()
- (2) 자율 주행 자동차의 교통 신호등 인식 학습에 세로 모양의 외국 신호등 이미지 데이터를 활용하였다. ()

2 머리 모양 추천 인공지능 프로그램으로 피해를 보는 사람이 있거나 프로그램의 혜택을 받지 못하는 사람이 있다면 어떻게 해결해야 할지 내 생각을 이야기해 보자.

다음과 같은 논제 중 하나를 골라 짝과 함께 토의·토론해 보자.

논제

- 인공지능 개발을 위해 개인 정보를 많이 수집해도 된다. vs 개인 정보 보호를 위해 인공지능을 개발하지 않아야 한다.
- 인공지능의 의사결정은 항상 공정하다. vs 공정하지 않을 수도 있다.
- 인공지능을 활용하여 예술 작품을 창작할 때 저작권은 인공지능을 개발한 개발자가 가져야 한다. vs 인공지능을 사용한 사용자가 가져야 한다.

1 논제에 대해 찬성과 반대 중 자신의 입장을 주장하고, 그 근거를 작성해 보자.



2 1에서 작성한 내용을 바탕으로 찬성 입장, 반대 입장인 친구와 5분간 토의·토론해 보자.



3 찬성과 반대의 입장인 친구와 역할을 바꾸어 5분간 다시 토의해 보자.
(이 때, 앞에서 상대방이 주장했던 의견과 근거를 똑같이 발언하지 않도록 유의한다.)



4 지금까지 토의한 내용을 바탕으로 찬성과 반대의 입장을 종합하여 자신의 의견을 정리해 보자.

| 구분 | 찬성 입장 | 반대 입장 |
|----------|-------|-------|
| 주장과 근거 1 | | |
| 주장과 근거 2 | | |
| 주장과 근거 3 | | |
| 종합 의견 | | |

우리나라는 과학기술정보통신부에서 사람이 중심이 되는 인공지능 윤리 기준을 크게 3대 원칙과 10대 핵심 요건으로 정하고 있다. 이 윤리 기준에서는 사람이 중심이 되는 인공지능을 구현할 것을 강조하였으며, 인공지능이 인간에게 유용함과 동시에 인간의 성품을 보존하고 함양하는 방향으로 개발·활용되어야 함을 강조하고 있다.

3대 원칙

인간 존엄성 원칙

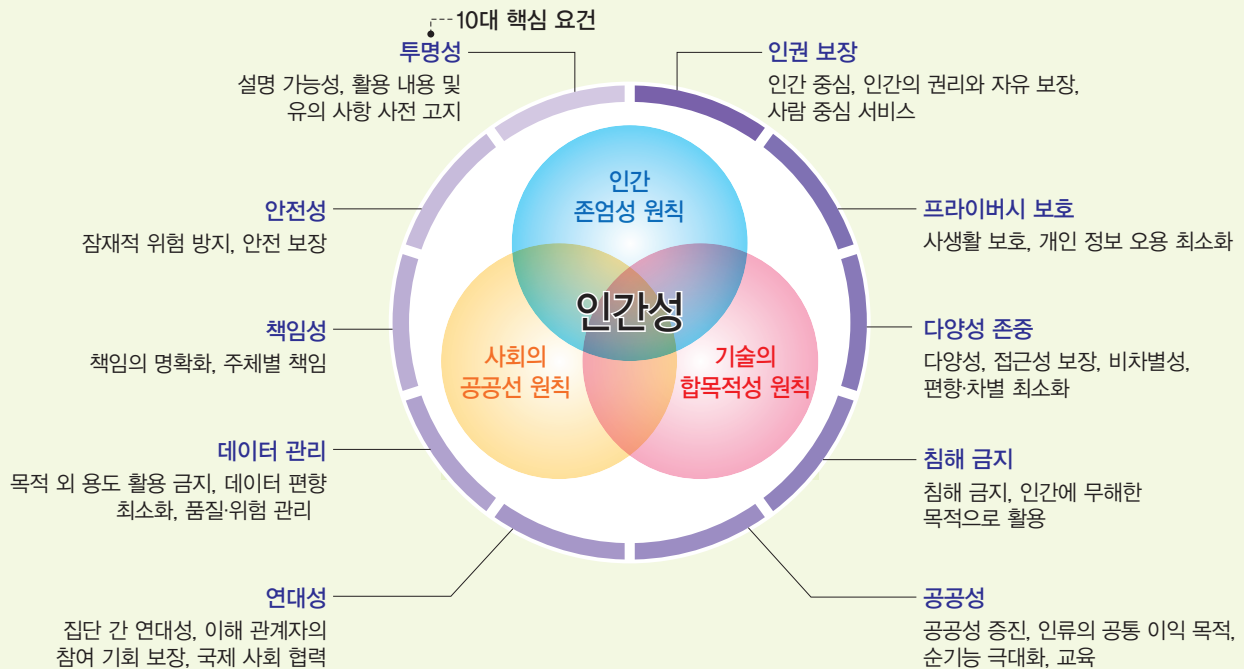
인간의 생명은 물론 정신적 건강에 해가 되지 않는 범위에서 개발 및 활용되어야 한다.

기술의 합목적성 원칙

인류의 삶에 필요한 도구라는 목적과 의도에 맞게 개발 및 활용되어야 하며 그 과정도 윤리적이어야 한다.

사회의 공공선 원칙

공동체로서 사회는 가능한 많은 사람의 안녕과 행복이라는 가치를 추구한다.



▲ 인공지능 윤리 기준 10대 핵심 요건과 3대 원칙



이미지 분류 모델

(은)는 이미지의 특성을 분석하여 특정 이미지가 어떤 범주나 클래스에 해당하는지 예측하는 모델이다.

179쪽

소리 분류 모델

소리 분류 모델은 (을)의 특성을 분석하여 특정 소리가 어떤 범주나 클래스에 해당하는지 예측하는 모델이다.

186쪽

텍스트 분류 모델

텍스트 분류 모델은 (을)를 학습하는 과정에서 텍스트의 패턴을 분석하여 특정 텍스트가 어떤 범주나 클래스에 해당하는지 분류하는 모델이다.

192쪽



데이터 수집 시 주의 사항

문제 해결에 필요한 데이터를 수집할 때에는 개인 정보나 (을) 침해 문제를 주의해야 한다.

200쪽

인공지능 모델 활용 시 주의 사항

인공지능 모델을 활용하는 과정에서 (이)가 유출되지는 않는지, 사생활이 노출되어 피해를 줄 가능성은 없는지 다양한 요소를 검토해야 한다.

201쪽

나의 성취 수준

| 교과 역량 | 평가 항목 | 성취 수준 |
|---------|--|-------|
| 인공지능 소양 | 인공지능 시스템으로 해결 가능한 문제를 탐색할 수 있는가? | ☆☆☆☆☆ |
| | 알맞은 데이터를 활용하여 인공지능 시스템을 구성할 수 있는가? | ☆☆☆☆☆ |
| | 문제 해결에 적합한 인공지능 시스템을 적용할 수 있는가? | ☆☆☆☆☆ |
| | 인공지능 학습에서 발생하는 윤리적인 문제의 해결 방안을 제시할 수 있는가? | ☆☆☆☆☆ |
| | 인공지능 윤리 문제에 대해 근거를 가지고 논리적으로 자신의 주장을 이야기할 수 있는가? | ☆☆☆☆☆ |

