



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

프로젝트기반

인공지능 윤리교육 프로그램 개발

제주대학교 교육대학교

컴퓨터교육전공

배진아

2020년 8월

프로젝트기반
인공지능 윤리교육 프로그램 개발

지도교수 조 정 원

배 진 아

이 논문을 교육학 석사학위 논문으로 제출함

2020 년 6 월

배진아의 교육학 석사학위 논문을 인준함

심사위원장 _____ 김 철 민



위 원 _____ 김 한 일



위 원 _____ 조 정 원



제주대학교 교육대학원

2020 년 8 월



<국문 초록>

프로젝트기반 인공지능 윤리교육 프로그램 개발

배진아

제주대학교 교육대학원 컴퓨터교육전공

지도교수 조 정 원

우리는 지능정보화 시대를 살아가고 있다. 빅데이터, 블록체인, 인공지능 등의 기술이 사회 변화를 주도하고 있고 그 중 인공지능과 로봇은 바이오 기술과 함께 영향력이 가장 큰 기술로 분류되고 있다. 또한 인공지능의 지능적이고 자율적인 특성으로 인해 발전 속도와 파급력이 상당히 크며 이러한 기술의 이면에는 역기능이 동반된다. 따라서 인공지능 기술이 안전하고 이롭게 사용되기 위해서는 그에 대응하는 윤리와 윤리교육이 함께 필요하다. 이러한 변화에 맞추어 교육계에서는 컴퓨팅사고력을 함양한 미래융합인재 양성을 위한 목적으로 다양한 소프트웨어 교육이 실시되고 있고 앞으로의 소프트웨어 교육은 인공지능을 포함하여 확대될 것이다. 기존의 소프트웨어 교육에서 정보윤리 교육은 인공지능의 특성이나 내용을 반영하기에는 한계가 있고, 문제 해결 과정의 시작이나 끝에 지식 전달 수준으로 잠시 언급되거나 생략하는 기존 수업방식에도 한계가 있다. 그러므로 본 논문에서는 인공지능의 특성을 반영한 새로운 인공지능 윤리교육을 개발하고, 문제 해결의 중요한 도구로서 인공지능 교육이 진행될 때 문제해결의 단계마다 알고 실천할 윤리를 적용한 ‘인공지능 윤리교육 모델’ 개발을 연구 목적으로 하였다.

본 논문의 연구 내용은 국내·외 인공지능 윤리 기준 및 기존 윤리를 바탕으로 인공지능 교육 내용 및 연구 동향을 분석하였다. 분석 내용을 토대로 개발한 내용은

다음과 같다.

첫째, 인공지능 윤리의 핵심가치와 기술의 핵심인 알고리즘과 데이터를 이해할 수 있는 윤리항목을 도출하였다.

둘째, 소프트웨어 교육의 목표인 컴퓨팅사고력 절차를 반영한 5단계 교육 프로세스를 개발하였다.

셋째, 윤리항목과 인공지능 교육 5단계 프로세스를 적용한 인공지능 윤리교육 프로그램을 개발하였다.

마지막으로 전문가의 검증을 거쳐 중학생에게 적용 가능한 인공지능 윤리교육 프로그램을 개발하였다.

개발된 모델은 인공지능 기술의 특성과 내용을 반영한 윤리교육 프로그램이고 SW를 통한 문제 해결 과정마다 윤리항목을 고민하고 적용하도록 하였다.

본 논문에서 개발한 ‘인공지능 윤리교육 모델’을 통해 실천할 수 있는 윤리교육으로 자리 잡는 기초가 되기를 기대한다.

〈제목 차례〉

I. 서론	1
1. 연구의 필요성 및 목적	1
2. 연구의 내용 및 방법	2
II. 이론적 배경	4
1. 인공지능 윤리 관련 국내·외 동향 분석	5
1.1 해외 인공지능 윤리 가이드라인	5
1.2 국내 인공지능 윤리 가이드라인	10
1.3 인공지능 윤리 연구 분석	12
2. 인공지능 교육 관련 국내·외 동향 분석	13
2.1 국내 인공지능 교육 현황	13
2.2 미국 AI4K12 인공지능 교육 현황	14
2.3 중국의 인공지능 교육 현황	15
2.4 MIT 인공지능 윤리교육	16
2.5 인공지능 윤리교육 연구 분석	18
2.6 인공지능 교육 플랫폼	19
3. 정보윤리 교육과 인공지능 윤리교육	20
3.1 정보윤리 교육	20
3.2 공학윤리 및 과학기술윤리	22
3.3 인공지능 윤리교육의 필요성	23
III. 인공지능 윤리교육 프로그램 개발	25
1. 프로그램 개발 과정	25

1.1 목적 및 개발방향	25
1.2 인공지능 윤리 핵심가치와 윤리 영역	28
1.3 인공지능 윤리교육 모델 개발	29
2. 인공지능 윤리교육 모델을 적용한 프로젝트	32
2.1 프로젝트 소개	32
2.2 세부 프로그램 내용	33
2.3 교수·학습 지도안	34
IV. 인공지능 윤리교육 프로그램 검증	37
1. 프로그램 대상 및 검증 방법	37
1.1 연구 가설	37
1.2 프로그램 적용 대상	37
1.3 연구 검증 방법	37
1.4 연구 검증 결과	39
2. 전문가 설문 내용	40
2.1 인공지능 윤리항목의 적절성	40
2.2 인공지능 윤리교육 모델의 적절성	41
2.3 교수·학습 지도안에 제시된 윤리 확인 문항의 적절성	42
2.4 사전·사후 인공지능 윤리 문항에 대한 적절성	45
V. 결론	47
[참고 문헌]	50
[ABSTRACT]	52
[부록]	54

〈표 차례〉

〈표 II-1〉 EU와 미국, 일본의 대응 동향	6
〈표 II-2〉 영국, 프랑스, 독일의 대응 동향	7
〈표 II-3〉 EURON의 로봇윤리 13원칙	7
〈표 II-4〉 아실로마 인공지능 23원칙 중 일부	8
〈표 II-5〉 미국 하원 결의안 531	9
〈표 II-6〉 일본 AI 개발 원칙	9
〈표 II-7〉 Ethically Aligned Design	10
〈표 II-8〉 공통원칙 내용과 의미	11
〈표 II-9〉 인공지능 윤리헌장 제 3장	11
〈표 II-10〉 차세대 SW교육 표준모델 중 ‘인공지능’ 부분	14
〈표 II-11〉 AI4K12 Big Idea 프레임워크	15
〈표 II-12〉 중국 초등학교 인공지능 교육목표 일부	16
〈표 II-13〉 인공지능 윤리교육 학습목표	17
〈표 II-14〉 인공지능 교육 플랫폼	19
〈표 II-15〉 정보윤리 교육 주요 영역	21
〈표 II-16〉 공학윤리 헌장	23
〈표 III-1〉 프로젝트 개요	33
〈표 III-2〉 세부 프로그램	33
〈표 III-3〉 1-2차시 지도안	34
〈표 IV-1〉 전문가 구성	38
〈표 IV-2〉 CVR 산출식과 응답인원에 따른 타당도	38
〈표 IV-3〉 델파이 분석 도구 기준 값	39
〈표 IV-4〉 전문가 설문 통계 결과	39

<표 IV-5> 윤리항목 및 정의 설문내용 최종	41
<표 IV-6> 인공지능 윤리교육 모델 설문내용 최종	41
<표 IV-7> 프로세스별 윤리 항목 설문내용 최종	42
<표 IV-8> 사전·사후 윤리항목 중심의 설문내용 최종	45

<그림 차례>

[그림 Ⅲ-1] 소프트웨어교육의 흐름과 목표	26
[그림 Ⅲ-2] 인공지능 윤리교육 진행 방안	27
[그림 Ⅲ-3] 인공지능 윤리교육에 필요한 핵심가치와 윤리항목	29
[그림 Ⅲ-4] 인공지능 윤리교육 모델	30

〈표 차례〉

〈표 1〉 EU와 미국, 일본의 대응 동향	6
〈표 2〉 영국, 프랑스, 독일의 대응 동향	6
〈표 3〉 EURON의 로봇윤리 13원칙	7
〈표 4〉 아실로마 인공지능 23원칙 중 일부	8
〈표 5〉 미국 하원 결의안 531	8
〈표 6〉 일본 AI 개발 원칙	9
〈표 7〉 Ethically Aligned Design	9
〈표 8〉 공통원칙 내용과 의미	10
〈표 9〉 인공지능 윤리헌장 제 3장	11
〈표 10〉 차세대 SW교육 표준모델 중 ‘인공지능’ 부분	13
〈표 11〉 AI4K12 Big Idea 프레임워크	15
〈표 12〉 중국 초등학교 인공지능 교육목표 일부	16
〈표 13〉 인공지능 윤리교육 학습목표	16
〈표 14〉 인공지능 교육 플랫폼	19
〈표 15〉 정보윤리 교육 주요 영역	21
〈표 16〉 공학윤리 헌장	22
〈표 16〉 프로젝트 개요	30
〈표 17〉 세부 프로그램	31
〈표 18〉 1-2차시 지도안	32
〈표 19〉 전문가 구성	36
〈표 20〉 CVR 산출식과 응답인원에 따른 타당도	36
〈표 21〉 델파이 분석 도구 기준 값	37

<표 22> 전문가 설문 통계 결과	37
<표 23> 윤리항목 및 정의 설문내용 최종	38
<표 24> 인공지능 윤리교육 모델 설문내용 최종	39
<표 25> 프로세스별 윤리 항목 설문내용 최종	40
<표 26> 사전·사후 윤리항목 중심의 설문내용 최종	42

<그림 차례>

<그림 1> 소프트웨어교육의 흐름과 목표	24
<그림 2> 인공지능 윤리교육 진행 방안	25
<그림 3> 인공지능 윤리교육에 필요한 핵심가치와 윤리항목	27
<그림 4> 인공지능 윤리교육 모델	28

I. 서론

1. 연구의 필요성 및 목적

4차 산업혁명을 주도하는 핵심 기술로 인공지능이 떠오르고 있다. 인공지능은 고도의 자율성과 우월한 지능을 가진 특이한 소프트웨어(이하 SW) [1]로, 자율주행 자동차, 인공지능 스피커, 챗봇 등 인류의 생활에 밀착된 형태로 근접하여 존재하고 있다. 인공지능은 고유 알고리즘으로 기계학습을 통해서 이루어지며 발전 속도가 빠르고 응용 분야가 방대해 사회 전반에 큰 영향을 미치고 있다.

인공지능 기술의 발전은 인간에게 편리함을 제공하고 인간의 행복을 추구하는 것을 목표로 진행되며 이로 인해 그동안 불가능으로 여겼던 많은 일을 이루어내고 있다. 그러나 인공지능 기술은 사용 목적과 방법을 달리할 때 인간에게 심각한 피해를 일으킬 수 있다. 인공지능의 역기능과 위험을 인지하고 그에 대응하는 방안을 모색하고 실천하는 것이 필요하다. 따라서 인공지능 기술은 인공지능 윤리와 함께 발전할 때 안전하고 편리한 기술이 되며 교육에서도 인공지능 교육과 인공지능 윤리교육은 함께 진행되어야 한다.

우리나라 정보윤리 교육은 2015개정 정보과 교육과정에서 ‘정보 문화’ 영역의 핵심개념으로 강조되고 있다. 하지만 기존의 정보윤리 교육은 인공지능의 개념과 특성을 반영하지 못하는 한계점을 가지고 있다. 인공지능의 특성은 예측과 통제가 가능한 기존의 알고리즘과 달리, 기계학습 과정에서 자체 규칙이 구축되어 그 과정을 인간이 통제하거나 알지 못한다. 그 예로 특정 네티즌들에 의하여 왜곡된 내용이 집중적으로 학습된 인공지능 ‘테이’는 ‘히틀러가 옳으며 유대인이 싫다’는 편견적 발언을 하여 세간의 비난 대상이 되기도 했다. 이는 인공지능을 개발하는 과정에서 인공지능의 특성과 윤리적 문제를 고려하지 않아 나타난 결과이다. 따라서 인공지능의 특성을 고려한 새로운 인공지능 윤리교육이 필요하다.

또한, 기존의 정보윤리 교육은 정보교과에서 시수 부족 등의 문제로 단순한 지식 전달에 그치거나 생략되는 경우가 대부분이다[14]. 지식 전달 위주의 교육은 인턴

넷 중독, 사이버 폭력, 사생활 침해 등의 부정적 변화를 초래하였으며 현재까지도 사회문제로 거론되고 있다. 따라서 윤리교육은 지식 전달의 교육이 아니라 윤리 실천 방안을 탐색하고 계획하여 실천하는 데 중점을 두는 교육으로 진행되어야 한다 [4]. 실천 위주의 교육이 되기 위해서는 SW교육과 윤리교육이 분리되어서는 안 된다. SW교육은 컴퓨팅사고력을 기반으로 실생활의 문제를 해결하는 과정을 담고 있다. 인공지능 또한 단순한 활용 교육을 넘어, 문제를 인식하여 아이디어를 도출하고, 설계 및 제작, 공유, 수정과 같은 [5] 메이커 활동에서도 문제해결을 위한 중요한 도구가 될 것이다. 따라서 올바른 윤리교육은 SW교육과 윤리교육이 함께 이루어져야 하며 문제를 해결하는 단계마다 윤리항목을 적용하여 역기능을 최소화하고, 실천력 강화에 도움이 되어야 한다.

본 논문은 기존의 정보윤리 교육이 인공지능 내용과 기술의 특성을 반영한 윤리 교육을 포함하지 못하는 한계를 극복하고, 실천력 중심의 인공지능 윤리교육을 위하여 메이커 교육의 단계마다 고려해야 하는 인공지능 윤리항목을 반영하고 적용하는 ‘인공지능 윤리교육 모델’ 개발을 목표로 한다.

2. 연구의 내용 및 방법

본 논문은 인공지능을 이용하여 문제를 해결하는 과정에서 윤리 문제를 고려하여 기획과 제작을 할 수 있도록 하는 데 목적을 두고, 인공지능 윤리교육에서 실천 가능한 프로그램을 개발하고자 하였다. 인공지능을 이용하여 문제를 해결하는 과정의 단계마다 실천적인 윤리의식을 함양할 수 있는 교육 프로그램과 함께 사전·사후 체크리스트 및 실천윤리 확인 항목을 동시에 반영하고자 하였으며 구체적인 연구의 내용과 방법은 다음과 같다.

첫째, 선행연구와 인공지능의 특성을 분석하여 인공지능 교육에 적용할 윤리 항목을 도출한다.

1) 인공지능의 특성과 사회적 영향력을 분석하여 인공지능 윤리의 필요성을 도출한다.

2) 기존의 정보윤리 교육과 공학윤리를 분석하여 한계점을 도출한다.

3) 국내외 인공지능 윤리 가이드라인과 최근 동향을 분석하여 교육적 차원에서 적용 가능한 인공지능 윤리의 핵심가치와 공통 윤리 항목을 도출한다.

둘째, 인공지능 윤리교육을 위한 모델을 개발한다.

1) 인공지능 교육 프로세스를 개발한다.

2) 인공지능 핵심 가치와 공통 윤리 항목과 관련한 사전·사후 체크리스트를 개발한다.

3) 인공지능 교육 프로세스에 윤리 항목을 적용한 인공지능 윤리교육 모델을 개발한다.

4) 개발된 인공지능 윤리교육 모델을 기반으로 교수·학습 과정을 설계한다.

셋째, 개발한 프로그램에 대한 타당도 검사를 시행한다.

II. 이론적 배경

인공지능에 관한 논의는 1945년 Vannevar Bush가 ‘As We May Think’에서 인간의 지식 및 이해를 증진 시키는 시스템을 제안한 것으로부터 출발하였다. 이어서 1950년 Alan Turing은 논문 ‘Computing Machinery and Intelligence’에서 인간을 모방하고, ‘체스’ 같은 지능화 능력을 갖춘 기계의 개념에 대하여 논의한 바 있다. 이 논문은 인공지능 연구의 시발점으로 일컬어지며 ‘기계도 생각할 수 있는가?’라는 주제 및 그 방법론으로서 ‘Turing Test’라는 유명한 이론이 제시되었다. 이후 인공지능은 컴퓨팅 환경과 기술의 발전에 따라 발전기와 침체기를 반복하며 현재에 이르렀으며 전 세계인들은 인공지능 시대를 살고 있다[6].

인공지능이 세계적인 사건으로 주목받기 시작한 것은 1997년 5월 체스 전용 인공지능인 ‘Deep Blue’가 체스 세계 챔피언을 이겼던 일화이다. 2005년에는 스탠퍼드대의 자율주행 자동차가 사막을 달리는 경주에서 우승했으며, 2007년에는 카네기멜런대의 자율주행 자동차가 교통법규를 지키며 55마일의 길을 주행하며 목적지에 도착했던 일도 있으며, 2011년에는 IBM의 인공지능 ‘왓슨’이 퀴즈쇼에서 챔피언들을 모두 이겼던 사건도 일어났다. 그리고 2013년 인간 고유의 전유물이라고 여겨지던 바둑마저도 인공지능 프로그램인 ‘알파고’가 인간을 능가하였다. 그 후 우리 생활 가까이에서 번역, 사진 인식, 음성 인식, 통역, 금융, 의료, 재난구조, 소셜로봇 등 인공지능은 다양한 형태로 등장하고 있다[7].

그러나 인공지능 기술의 발전이 가져오는 편리함과 더불어 부정적인 측면도 동시에 나타나고 있다는 것 또한 부인할 수 없다. 자율주행 자동차의 사고와 그에 따른 책임 주체의 문제, 로봇 어드바이저의 투자 실패로 인한 손실 문제, 학습 과정에서 발생한 오류로 인한 인종차별적 발언이나 왜곡된 정보의 제공 등 다양한 역기능이 보고되고 있기 때문이다. 최근에는 인공지능 로봇에게 사람과 유사한 감정이입으로 인한 정체성 혼란을 가져오고 또한 군사적으로 활용하는 등의 문제점도 심각하게 제기되어 왔다.

이 같은 역기능을 포함하고 있으나 인공지능 분야의 발전이 가속화되는 이유는

두 가지가 있다. 먼저 ‘인공지능은 지능적이다.’라는 것이다. 즉, 지능을 가진 사람 같으면서 전문적인 사람보다 더 전문적이기 때문이다. 다음은 ‘자율적이다.’라는 것으로 스스로 학습하고 판단할 수 있는 기술이라는 것이다. 이러한 두 가지 특징은 인공지능이 인간보다 뛰어난 지능을 가지고 스스로 학습할 수 있다는 것이다. 그러나 여전히 SW이기 때문에 여러 취약점이 있을 수밖에 없으며, 무엇보다 개발자의 주관적 의도에 따라 악용될 수 있는 점에서 커다란 단점을 동반한다. 이러한 맥락에서 “인공지능이 인류 사상 최대 성과인 동시에 최후의 성과이자 인류의 재앙이 될 수 있다”고 2014년 5월 영국 인디펜던트지의 기고문에서 문제로 제기되기도 했다. 더 나아가 구글의 기술사 레이 커즈와일은 2045년 즈음 인공지능이 인간을 뛰어넘는 시점인 ‘특이점’이 온다고 예견하였고 이후 전문가들에 의하면 그 시점은 더 단축될 것으로 예측되고 있다[2].

1. 인공지능 윤리 관련 국내·외 동향 분석

인공지능 교육과 함께하는 인공지능 윤리교육 프로그램 개발을 위해서는 ‘윤리’ 부분과 ‘교육’ 부분의 분석 및 연구 절차가 필요하며 먼저 인공지능의 윤리적인 측면에서의 연구 동향을 살펴보고자 한다.

1.1 해외 인공지능 윤리 가이드라인

세계 각국은 현재 인공지능 기술 발전에 따른 사회적 변화에 대비하기 위한 구체적인 방안을 마련하는 과정에 돌입해 있다. EU, 미국, 일본 등 AI 선도국을 중심으로 ‘AI 윤리원칙·가이드라인’이 마련되고 있으며, 지속적으로 수정·보완이 진행되고 있다. <표 II-1>을 살펴보면 유럽위원회는 이전부터 로봇의 윤리, 규제 가이드라인 마련 등을 위해 다양한 프로젝트 시행하고 있음을 알 수 있다. 또한 미국 정부에서도 국가과학기술위원회(NSTC) 분과에서 인공지능의 영향력을 예측하고 분석하는 연구를 진행해오고 있다. 일본 역시 2016년 말부터 인공지능의 윤리적 쟁점들을 조명하고, 국제적인 논의의 기초가 될 개발 가이드라인 및 사회 원칙[8] 등을 공개적으로 제시한 바 있다.

<표 II-1> EU와 미국, 일본의 대응 동향[8]

국가	대응 동향
EU	2004, 유럽로봇연구연합 (EURON)-로봇윤리 13 원칙 2005, 제 6차 다자간 공동기술개발연구프로그램 (레 6)-윤리로봇 프로젝트 2007, 로봇윤리 로드맵 (ICRA) 2012, 제 7차 다자간 공동기술개발연구프로그램 (FP6)-로봇법 프로젝트 2012, 로보틱스 규제 가이드라인 2017, 로봇 시민 규제 (의회) 2018, Coordinated Plan on AI(집행위원회) 2018, 신뢰할 수 있는 AI 윤리가이드라인 초안
미국	2016, 빅데이터 , 알고리즘 시스템 , 기회 , 시민권에 관한 리포트 2016, AI 미래에 관한 준비 2016, 국가 인공지능 R&D 전략 계획 2016, 인공지능 , 자동화 그리고 경계 2016, Ethically Aligned design(IEEE) 2016, 2030 년 인공지능과 삶 (스탠퍼드) 2017, 아실로마 인공지능 23원칙
일본	2016, 인공지능 R&D 안정성 확보 등 8대 원칙 2017, 인공지능 연구개발 가이드라인 2018, AI 활용 원칙 안 2019(예정), 인간중심 AI 사회원칙(안)

영국 역시 EU 탈퇴 후 독자적인 ‘데이터 윤리 프레임 워크 7원칙’을 발표하는 등 다양한 윤리 가이드라인을 개발하고 있음을 알 수 있으며, 프랑스는 2017년부터 다양한 대응 방안을 마련해오고 있으며 <표 II-2>에서와 같이 제시하였다. 특히 독일의 경우는 세계 최초로 ‘자율주행 자동차 윤리 가이드라인’[8]을 제시하였다.

<표 II-1>, <표 II-2>의 국가 또는 단체에서 발표한 가이드라인들의 내용을 분석하면 다음과 같다.

<표 II-2> 영국, 프랑스, 독일의 대응 동향[8]

국가	대응 동향
영국	2016, 로봇과 AI(UK House of Common) 2017, 기계학습 : The Power and Promise of Computer that Learn by Example (Royal Society) 2018. AI Sector deal(BEIS) 2018, 영국의 AI : ready, willing and able>(UK House of Lords) 2018, 데이터 윤리 프레임워크 7 원칙
프랑스	2017, 인공지능의 경제사회적 영향 전망 2017, How can humans keeps the Upper Hand?(CNIL) 2018, 프랑스와 유럽 전략을 향한 의미 있는 인공지능을 위하여 (Cedric villani)
독일	[독일 자율주행기술 윤리지침의 주요 내용] 인간의 생명 보호가 항상 최우선 순위이며, 사고가 예상된다면 다른 차량, 재산을 파괴하고 동식물 등 다른 생물에 피해를 주더라도 인간의 생명을 구하고 부상을 피해야 한다. 사고를 피할 수 없는 경우, 자율주행시스템은 생명에 대한 가치를 판단하고 선택하지 말아야 한다. 아이를 살리기 위해 노인을 죽이는 것과 같은 일은 할 수 없다. 연령, 성별, 인종, 장애 등에 대한 판단을 해서는 안 된다. 모든 것은 기록되어야 한다. 사고 발생 시 책임 소재가 확실할 수 있도록 블랙박스 같은 형태로 모든 데이터가 저장되어야 한다. 물론 운전자의 신원도 명확히 확인되어야 한다. 대기업이 사용자의 개인 정보를 불법으로 판매하거나 악용하는 것을 차단하기 위해서 운전자가 차량에서 수집되는 개인 정보를 완벽하게 통제할 수 있어야 한다.

○ EURON의 로봇윤리 13원칙은 유럽로봇연구연합(EURON)의 ‘로봇윤리 로드맵’(2007)에서 로봇 개발자를 위한 가이드로 <표 II-3>의 내용과 같이 사회적 책무, 차별과 낙인화 금지 등 총 13개의 원칙으로 구성되어 있다.

<표 II-3> EURON의 로봇윤리 13원칙[8][37]

로봇에 선행하는 윤리 원칙	
인간의 존엄과 인간의 권리	프라이버시
평등, 정의, 형평	기밀성
편익과 손해	연대와 협동
문화적 다양성을 위한 존중	사회적 책무
차별과 낙인화 금지	이익의 공유
자율성과 개인의 책무성	지구 상의 생물에 대한 책무
주지된 동의	

○ OECD는 2017년 ‘OECD Digital Economy Outlook 2017’ 보고서의 기술전망 항목에서 인공지능 기반 의사결정의 투명성과 감독 보장, 알고리즘 편향성 및 차별 방지, 책임과 보안, 안전에 관한 내용 등의 인공지능 개발에 관한 권고안을 발표하였다[9].

○ ‘아실로마 인공지능 23원칙’은 2016년에 연구 관련 쟁점, 윤리와 가치, 장기적 이슈 3가지 영역에서 23개의 원칙을 발표했다. <표 II-4>는 윤리와 가치 영역 13가지에 관한 내용으로서 인간의 가치와 투명성, 책임성, 프라이버시, 개인정보 보호, 공공의 이익 등을 제시한 부분이다.

<표 II-4> 아실로마 인공지능 23원칙 중 일부(IEEE, 2016)[8][38]

윤리와 가치 13가지	
6. 안전	13. 자유와 프라이버시
7. 실패의 투명성	14. 이익의 공유
8. 사법적 투명성	15. 번영의 공유
9. 책임성	16. 인간통제
10. 가치 일치	17. 사회 전복 방지
11. 인간의 가치	18. 인공지능 무기 경쟁
12. 개인정보 보호	

○ ‘하원 결의안 531’은 미국 하원의 발표로 인공지능의 윤리적 발전을 위한 인공지능 개발 10가지 목표를 제시하였다. 그 내용은 <표 II-5>와 같이 개인의 안전, 발전, 이익에 대하여 중점을 두고 결의안을 발표하였다.

<표 II-5> 미국 하원 결의안 531(2019.2.27.)[8]

미국 하원 인공지능의 윤리적 발전 결의안 : 인공지능 개발 목표
1. 산업계, 정부, 학계 및 시민사회가 참여
2. 인공지능 시스템, 프로세스의 투명성 및 설명 가능성
3. 소외된 계층이나 여성에게 권한을 부여하는 것을 도움
4. 개인정보보호와 정보 프라이버시 보장
5. 의미 있는 일과 삶을 찾을 수 있는 커리어 기회 확대
6. 모든 자동화된 의사결정을 위한 책임 및 감독
7. STEM, 사회 과학, 인문학의 평생 학습
8. 기술 서비스 및 혜택의 공정성
9. 안전하고 유익한 인공지능에 관한 학문적 연구
10. 현재와 미래의 안전, 보안 및 인공지능 시스템 제어

○ 일본의 ‘AI 개발 원칙’은 개발자가 인공지능 개발 시 지켜야 할 9원칙을 제시한 것이다. 편익증진, 위험 억제, 이용자 수용성 향상 등 3개의 기준에 따라 분류하였으며, <표 II-6>에서와 같이 투명성, 책임성, 프라이버시, 보안을 강조하였고 AI의 위험을 억제하는 데 중점을 두었다.

<표 II-6> 일본 AI 개발 원칙[8]

분류	원칙	세부 내용
AI의 편익증진 및 AI 네트워크화 촉진	연계	AI 시스템의 상호접속성, 상호운용성에 유의
AI의 위험 억제	투명성	AI 시스템 입출력의 검증 가능성과 판단 결과의 설명 가능성에 유의
	제어 가능성	AI 시스템의 제어 가능성에 유의
	안전	AI 시스템이 이용자와 제 3자의 생명·신체·재산에 위해를 미치는 것이 없는지 유의
	보안	AI 시스템의 보안에 유의
	프라이버시	이용자와 제 3자 프라이버시가 침해되지 않도록 유의
	윤리	인간의 존엄과 개인의 자율을 존중하도록 유의
이용자의 수용성 향상	이용자 지원	이용자에게 선택의 기회를 적절히 제공하는 것이 가능하도록 배려
	책임	개발자는 이용자를 포함한 이해관계자에 대한 책임을 완수하도록 노력

○ 미국 IEEE는 2019년 국제표준화를 목표로 하는 ‘Ethically Aligned Design’을 발표하였다. IEEE는 EAD1e에서 자율적이고 지능적인 시스템(autonomous and intelligent systems: A/IS)에 적용할 수 있는 8개의 일반원칙을 구성하여 제시하였다. 또한 인류학적, 정치적, 기술적 측면을 반영한 프레임워크를 제공하기 위해 보편적 인류 가치(Universal Human Values), 정치적 자기 결정과 데이터 에이전시(Political Self-Determination and Data Agency), 기술적인 신뢰도(Technical Dependability) 등 3가지로 설계하였다. 이 프레임워크와 8가지 원칙을 적용하여 제시된 EAD1e의 분석 및 권고 사항은 인류의 이익을 위한 기술 발전에 있어 정부, 기업 및 일반 대중이 고려해야 할 지침을 발표하였다[9].

<표 II-7> Ethically Aligned Design[9]

원칙	내용
인권	A/IS는 국제적으로 알려진 인권을 존경, 증진 그리고 보호하기 위해 만들어 지고 운영되어야 한다.
복지	A/IS 창조자는 개발의 주요한 성공 기준을 증가된 인간 복지로 채택해야 한다.
데이터 에이전시	A/IS 창조자는 개인이 그들의 정체성을 통제할 수 있는 능력을 갖추어 그들의 데이터에 접근하고 안전하게 공유할 수 있는 권한을 주어야한다.
효율	A/IS 창조자와 운영자는 A/IS의 목적에 대한 효과와 적합성의 증거를 제공해야 한다.
투명성	특별한 A/IS의 결정의 근거는 항상 쉽게 찾을 수 있어야 한다.
책임	A/IS는 모든 결정에 모호하지 않은 이유를 증거로 만들어지고 운영되어야 한다.
악용의 의식	A/IS 창조자는 A/IS의 운영에 있어 모든 잠재적인 악용과 위험에 대해 반대하여 감시하여야 한다.
역량	A/IS 작성자는 안전하고 효과적인 운영에 필요한 지식과 기술을 명시하고 운영자는 이를 준수해야 한다.

1.2 국내 인공지능 윤리 가이드라인

국내에서는 2017년 정보화진흥원을 중심으로 지능정보사회 윤리가이드라인(안)을 제시하였고 계속해서 2018년 ‘지능정보사회 윤리가이드라인’[10]이 발표되었다. 이후 한국인공지능윤리협회에서 ‘인공지능 윤리헌장’을[26] 발표 하는 등 국내에서도 윤리와 관련하여 다양하게 거론되고 있다.

○ ‘지능정보사회 윤리가이드라인’에서는 <표 II-8>과 같이 지능정보기술의 특성에 기반을 둔 공통적인 원칙으로 공공성, 책무성, 통제성, 투명성과 같은 4가지 원칙에 대하여 정의하고 있다. 이는 해외 윤리지침들이 개발자와 공급자를 대상으로 기술개발 과정에서 발생하는 책임윤리를 강화하는 부분에만 집중된 것과 비교하였을 때 개발자, 공급자, 이용자를 모두 고려한 윤리 지침이다.

4개의 공통원칙 아래에 개발자, 공급자, 사용자 등 모두가 지켜야 하는 원칙을 정의하고 대상별로 지켜야 하는 세부지침도 명시하고 있다. 이러한 대상별 윤리의

적용은 모든 이해관계자에 대한 이해와 배려가 고려되어야 한다는 점에서 본 연구와 방향을 같이 한다고 볼 수 있다.

<표 II -8> 공통원칙 내용과 의미[10]

공통원칙	내용
P : 공공성(Publicness)	지능정보기술은 가능한 많은 사람에게 도움을 주어야 하며, 지능정보기술에 의해 창출된 경제적 번영은 모든 인류의 혜택을 위해 광범위하게 공유되어야 한다. *공정성, 차별배제, 접근성 보장
A : 책임성(Accountability)	지능정보기술 및 서비스에 의한 사고 등의 책임 분배를 명확히 하고, 안전과 관련한 정보 공유, 이용자 권익보호 등 사회적 의무를 충실히 수행해야 한다. *책임성, 윤리적 절차 이행, 위험 예방
C : 통제성(Controllability)	지능정보기술 및 서비스에 대한 인간의 제어 가능성 및 오작동에 대한 대비책을 미리 마련하고, 이용자의 이용선택권을 최대한 보장하여야 한다. *제어 가능성, 위험관리, 이용자 주도성
T : 투명성(Transparency)	기술개발, 서비스설계, 제품기획 등 의사결정 과정에서 이용자·소비자·시민 등의 의견을 반영하도록 노력해야 하며, 이용 단계에서 예상되는 위험과 관련한 정보를 공개·공유하고, 개인정보 처리의 전 과정은 적절하게 이루어져야 한다. *설명 가능성, 위험정보 공유, 이용자/시민참여

○ ‘인공지능 윤리현장’에서는 각국의 정부, 기업, 단체, 개인에게 적용되는 공통된 기준을 공포하고 있다. 1장에서는 인간과 인공지능의 관계에 대한 원칙 7가지를 정의하고 있고, 2장에서는 선하고 안전한 인공지능이라는 원칙으로 알고리즘이나 데이터에 대한 투명성을 강조하고 있다. 3장의 인공지능 개발자 윤리는 <표 II -9>처럼 개발자의 책임의식에 관한 내용이 주를 이루고 있다.

4장은 인공지능 소비자의 윤리에 대해 언급하고 있고, 5장은 공동의 책임과 이익의 공유 원칙을 제시하고 있다.

〈표 II-9〉 인공지능 윤리헌장 제 3장 [26]

구분	인공지능 개발자의 윤리
제21조	인공지능 개발자는 강화된 윤리적 책임의식을 가져야 한다.
제22조	인공지능 개발자는 인간에게 해를 끼치는 인공지능을 만들어서는 안 된다.
제23조	인공지능 개발자는 합의된 안전 개발 지침에 의거하여 인공지능 제품과 서비스를 만들어야 한다.
제24조	개발자는 인공지능에 자체적인 의사결정 능력 부여시 고도의 주의를 기울여야 한다.
제25조	개발자는 머신러닝 알고리즘을 적용할 경우, 제13조에 의한 빅데이터를 선별하기 위해 노력해야 한다.
제26조	개발자는 머신러닝 알고리즘을 적용할 경우, 출시 전 충분한 시뮬레이션을 통해 결과값에 대한 오류를 최소화하고, 결과값에 대한 충분한 예측 정보를 확보해야 한다.
제27조	개발자 또는 개발회사는 소비자에게 인공지능 제품과 서비스에 대한 충분한 정보제공 및 주의사항을 고지할 의무가 있다.
제28조	개발자 또는 개발회사는 인공지능 제품과 서비스에 본래 목적 외의 기술이나 기능이 내장되는 경우 이를 소비자에게 고지할 의무가 있다.
제29조	개발자 또는 개발회사는 소비자에게 제품과 서비스에 인공지능 기술이 적용되어 있음을 표지나 문서, 음성 등으로 사전에 고지해야 한다.
제30조	개발자 또는 개발회사는 소비자로부터 데이터를 수집 받을 경우 사전에 동의를 구하거나 해당 내용을 고지해야 한다.
제31조	인공지능 개발자는 인공지능의 안전과 윤리에 관해 다양한 관계자들과 협력하고 지속적인 교육을 받아야 한다.
제32조	인공지능 개발회사는 소속 개발자들이 안전하고 윤리적인 인공지능을 만들 수 있도록 관련 환경을 조성해주고 적극 지원해주어야 한다.
제38조	인공지능 개발자는 인공지능을 이용한 신기술을 개발할 경우, 순기능뿐만 아니라 사회에 미칠 부작용을 고려해야 하며 역기능이 우려될 경우 개발에 보다 신중해야한다.

이처럼 국내·외 인공지능 윤리와 관련한 지침들의 공통적인 키워드는 인간 존중 및 가치, 책임, 안전, 공공성, 투명성, 다양성, 개인정보보호 이다. 반대로, 공통적인 키워드를 제외하면 윤리의 주체를 어디에 두느냐에 따라 국가나 기관별로 차이가 있는 것을 알 수 있다.

1.3 인공지능 윤리 연구 분석

다음은 인공지능 윤리와 윤리교육에 관한 연구 분석으로 인공지능의 특성을 반영한 윤리의 필요성과 대학이나 해외에서 진행되는 윤리교육 사례들도 소개하고

있다.

한상기는 “인공 지능 기술의 발전이 우리 사회의 이익을 극대화하고 공동체 전체의 혜택으로 이루어지게 해야 하지만 이런 필요성과 함께 위협받을 가능성 또한 잠재적으로 갖고 있다. 이런 문제를 대처하기 위해 준비해야 하는 우선 과제는 인공지능과 관련한 연구자, 개발자 등이 윤리문제에 관심을 갖게 하는 것이다”[7] 라고 하면서 실제적인 관련자들의 인공지능 윤리에 대한 문제를 인식하게 하는 것이 중요하다고 하였다.

김명주는 “인공지능은 인간과 수평적으로 공존할 대상으로 보아야 하며 이미 발생한 사건이나 발생 가능한 상상을 근거로 보았을 때 인공지능 윤리가 반드시 필요하다. 또한 인공지능 윤리를 인터넷 윤리의 연장선에서 보아야 한다.”라고 주장하였다.

이러한 선행 연구들을 정리하면, 인공지능 윤리교육의 필요성에 대해서는 대부분 공감하고 있지만 현재 일부 대학에서만 윤리교육이 진행되고 있고, 학교에서 진행되는 구체적인 교육 사례나 방법은 부족한 상황이라고 주장하였다.

2. 인공지능 교육 관련 국내·외 동향 분석

2.1 국내 인공지능 교육 현황

우리나라는 초등부터 단계별 ‘AI 교육 기본계획’[19]을 만든다고 발표하였다. 이 계획에는 초등부터 성인 대상의 평생교육까지 교육단계별로 중장기 비전을 담을 예정이라고 하였다. 유치원과 초등학교에서는 AI와 친숙해지도록 하는 과정에 교육의 중심을 두고 있으며 더 나아가 중학교와 고등학교에서는 AI 기본 소양을 갖추도록 교육하는 데 중점을 두고 있다. 최종적으로 대학 교육에서는 전문적인 AI기술 인력을 양성하는 것을 핵심적인 골자로 하고 있다. 더불어 누구나 AI관련 기본 소양과 올바른 윤리·가치관을 갖추도록 하는 방안도 중점적으로 다룰 방침인 것으로 발표되었다[18].

또한 한국정보과학교육연합회에서 제시한 「차세대 소프트웨어(SW)교육 표준모델 개발」에 따르면, SW교육에 인공지능 교육 내용[19]이 포함되어 있다. 현재 정보

교과의 학교 급별 내용체계를 보면 정보 문화, 자료와 정보, 알고리즘과 프로그래밍, 컴퓨팅 시스템의 영역으로 구성되어있다. 그러나 인공지능을 중심으로 급변하는 사회의 변화와 발전 상황을 반영한 SW교육이 필요하며 그 방안으로 인공지능을 포함하여 데이터 과학, 로봇틱스를 포괄하는 상위 수준으로서 인공지능과 융합이라는 영역이 추가로 제시된 바 있다. <표 II-10>은 초등학교부터 고등학교까지 학령별 인공지능 교육의 필수 내용을 제시하고 있다.

<표 II-10> 차세대 SW교육 표준모델 중 '인공지능' 부분[19]

학교 급	학년군	인공지능과 융합		
		데이터 과학	인공지능	로봇틱스
초등 학교	초등 2단계		인공지능의 활용1(음성인식 등)	생활 속 로봇의 종류 이해 로봇의 영향 탐색 올바른 로봇의 사용
	초등 3단계	빅데이터의 이해	인공지능의 개념 이해 인공지능의 활용2(머신러닝 등)	로봇의 기본 구조 이해
중 학 교	중등 1단계	빅데이터의 적용 올바른 빅데이터 사용	인공지능 지식표현의 이해(if then, tree) 머신러닝의 학습 방법 이해	완성형 로봇의 체험 완성형 로봇의 제어 완성형 로봇의 활용
	중등 2단계	데이터 수집 및 가공(제공데이터) 제공 데이터 처리 및 분석	인공지능 추론 방법 이해 머신러닝 접근 방법 이해(지도, 비지도, 강화 등)	기초로봇의 구성 기초로봇의 제어 이해 및 적용
	중등 3단계	데이터 수집 및 가공(웹데이터) 웹 데이터 처리 및 분석	여러 분야의 지식추론 여러 분야의 인공지능 학습(도메인)	로봇의 제작(필요한) 로봇제어 프로그래밍1
고등 학교	고등 1단계	다양한 데이터 수집 및 가공방 법 데이터 분석 및 시각화	머신러닝의 개념과 알고리즘 머신러닝의 데이터 분석방법(확률, 통계)	다양한 문제해결을 위한 로봇 분석 다양한 문제해결을 위한 로봇 설계 및 구성 로봇을 활용한 실생활 문제해결
	고등 2단계	사회문제 데이터 수집 및 가공 사회문제 데이터 분석 및 시각 화	인공신경망의 개념 및 프로그 램 작성 인공신경망 실생활 문제 적용	복잡한 문제해결을 위한 로봇 설계 및 제작 로봇 제어 프로그래밍 최적화

2.2 미국 AI4K12 인공지능 교육 현황

국내 인공지능 교육 현황과 달리 미국의 경우 카네기멜런대학, 플로리다대학, CSTA는 미국과학재단 후원으로 AAAI를 결성하여 K-12 대상으로 AI 교육을 위한 5개의 빅아이디어를 발표하였다. 이 자료는 <표 II-11>과 같이 주제를 정하고 개념, 교육목표, 학령별 교육내용을 정의하였음을 알 수 있다. 5가지 주제는 인식, 표현&추론, 학습, 자연스러운 상호작용, 사회적 영향으로 구성되어 있고 학년 군마다 학생들이 무엇을 할 수 있어야 하는지 제시하였다. 또한 K-12에서 인공지능을 가르쳐야 하는 이유와 함께 5개의 빅아이디어를 중심으로 학년별 교육 내용도 제시되어 있다[20]. 5가지 주제 중에서 ‘사회적 영향’에서 인공지능 윤리와 관련한 내용을 주로 다루고 있다.

<표 II-11> AI4K12 Big Idea 프레임워크[20]

주제	1. 인식	2. 표현·추론	3. 학습	4. 상호작용	5. 사회적 영향
개념	컴퓨터는 센서를 사용하여 세계를 인식	에이전트는 세상을 표현하고 구조화하여 추론에 사용	컴퓨터는 데이터로부터 학습	지능형에이전트는 인간과 상호작용을 위해 많은 지식을 요구	AI는 긍정·부정적 효과를 동시에 야기
교육 목표	-인간감각과 센서 차이 이해 -컴퓨터 인식 작동 방식과 한계 이해 -시각,음성 등 인식의 유형 파악 -지능형, 비지능형 기계 특징 파악	-표현의 유형 파악 -추론 알고리즘 유형 파악 -추론 알고리즘 작동 원리 이해 -추론 알고리즘의 한계 이해	-기계학습 접근방법 이해 -학습 알고리즘 유형 파악 -인공신경망 기초 개념 이해 -데이터가 학습에 미치는 영향 -기계학습의 한계	-자연어의 이해 -감성 컴퓨팅 이해 -상식 추론 이해 -인간과 로봇의 자연스런 인터페이스 방식 -인터페이스의 한계 이해	-AI가 산업, 의료, 교육, 정부 등에 미치는 영향 이해 -AI가 야기하는 윤리 딜레마 이해 -AI의 윤리적 표준 마련 필요 이해 -AI에 대한 일자리, 업무 변화 이해
학령별 교육 내용	(K-12)Alexa나 Siri와 소통하기 (3-5)센서입력이 디지털신호로 전환방법 이해 (6-8)센서 사용한 프로그램 개발 (9-12)다양한 컴퓨터 시각에 대한 배경 지식 습득	(k-12)의사결정 트리 활용법 습득 (3-5)트리를 이용한 동물 분류 시스템 표현 (6-8)집 위치를 그래프 모델로 설계, 최소 경로 추론 (9-12)택토 게임을 검색 트리로 표현	(k-12)이미지 인식을 사용하기 (3-5)기계학습 방법 3가지 이해 (6-8)훈련 데이터 편차 해결 (9-12)신경망 훈련 및 간단한 알고리즘 경험	(k-12)공부정 단어 식별 (3-5)억양, 표정, 자세를 통한 감정 이해 (6-8)단순 챗봇 구성 (9-12)문장 분석기로 애매한 표현 다루기	(k-12)실생활의 AI 사례 토론 (3-5)AI편향성에 의한 영향 설명 (6-8)AI의 예상치 못한 결과 이유 설명 (9-12)사회문제 해결을 위한 AI활용 방안 연구

2.3 중국의 인공지능 교육 현황

중국은 이미 학령별 인공지능 교과서를 만드는 등 2017년 중국 국무원에서 발간한 “차세대 인공지능 개발 계획”에서 인공지능 교육의 필요성을 강조하였다. 계획에 따르면 초등부터 고등학교의 인공지능관련 교육내용을 추가하도록 제안하고 이에 따라 2018년 4월 최초로 초·중학교용 인공지능 교과서가 출간되었다[21]. 해당 교과서에는 유치원의 경우 신체 놀이 활동을 통해 사물과 로봇에 대한 동작 원리를 이해하는 과정으로 구성되어 있다. 초등은 스크래치, 파이썬으로 프로그래밍 기초 개념을 익히고, 아두이노를 이용하여 주변 환경을 센싱하는 과정을 체험하도록 고안되었다. 중학교부터는 인공지능이 주변 환경을 어떻게 인식하는지와 데이터를 통한 기계 학습 원리를 익히고, 더 나아가 고등학교 교육에서는 인공지능 응용분야를 소개하고 기계학습 기초 모델의 원리를 익히고 구현하도록 설계되었다[22].

<표 II-12> 중국 초등학교 인공지능 교육목표 일부[21]

학교급	목표	내용
1학년	AI 찾기	유치원에서 초등학교로 전환으로 습관의 발달에 중점을 둠
2학년	AI 느끼기	특정 학습 능력이 있어 일부 습관이 점차 형성되지만 불안정
3학년	AI 활용하기	대인 관계 기술을 논리적으로 생각하고 향상시킴
4학년	혁신적인 AI	실습 능력과 성격 논리 사고 능력이 매우 뛰어나야 함
5학년	AI 제조	컴퓨터 프로그래밍 게임, 강의실 실험 등을 중심으로 진행
6학년	AI 발전	가장 활력이 넘치는 시기

<표 II-12>는 초등학교의 인공지능 목표와 내용 일부로 학년별로 인공지능 경험, 활용과 제조 및 발전으로 제시하여 정리한 것이다. 초등학교 1학년부터 6학년 까지 초등학교 시리즈를 대상으로 진행하며 학년별 목표는 '인공지능 찾기', '인공지능을 느끼는 것', '인공지능을 활용한 것', '혁신적인 인공지능', '제조된 인공지능', '인공지능의 발전' 순으로 구성되어 있다. 저학년에서는 인공지능과 관련한 습관의 발달 및 형성을 하고 고학년에 도달함에 따라 실습 및 논리와 사고 및 실습 실험으로 발전시키는 것이 중국 초등학교 인공지능 교육 목표라고 할 수 있다.

2.4 MIT 인공지능 윤리교육

MIT공대는 이전의 AI4K12와 달리 대학에서 중학생을 대상으로 한 인공지능 윤리 과정을 개설하였으며 윤리 문제를 중심에 두고 만들어졌다. 중학교 과정임에도 기계학습 알고리즘을 직접 해보면서 동시에 윤리개념 또한 자연스럽게 익히도록 하고 있다. <표 II-13>[17]과 같이 5가지 학습목표 아래 세부 성취 기준을 만들어 학습활동에 적용하였다[23][24]. 하지만, 윤리적인 문제에 접근하는 방법 면에서 참고할 만하나 일회성의 윤리교육으로 끝나는 경향이 있다. 또한 인공지능 활용 교육에서 문제 해결 과정에 적용하여 확대 될 때 기획, 설계, 제작의 단계마다 윤리 요소를 적용하는 데 한계가 있다.

<표 II-13> 인공지능 윤리교육 학습목표[23]

<p>1. 인공지능 시스템의 기본적인 작동 원리(메커니즘)를 이해한다.</p>
<p>a. 전 세계의 다양한 알고리즘을 인식하고, 일상생활 속에 있는 알고리즘과 컴퓨터 알고리즘의 예를 제시할 수 있다.(예, 케이크 만들기)</p> <p>b. 알고리즘의 세 단계(입력, 처리, 출력)를 알 수 있다.</p> <p>c. 인공지능은 알고리즘의 특정 유형임을 알고, 데이터셋(dataset), 학습 알고리즘(learning algorithm) 그리고 예측(prediction)으로 구성되어 있다는 것을 알 수 있다.</p> <p style="padding-left: 20px;">i. 지도 기반 머신러닝 환경에서 분류 문제를 이해할 수 있다.</p> <p style="padding-left: 20px;">ii. 훈련(training) 데이터의 양이 어떻게 지도 기반 머신러닝 모델의 정확성(accuracy)과 안정성(robustness)에 영향을 미치는지 이해할 수 있다.</p> <p>d. 일상생활 속에 있는 인공지능 시스템을 이해하고 인공지능 시스템이 예측하는 것과 인공지능 시스템이 사용하는 잠재적 데이터셋을 추론할 수 있다.</p>
<p>2. 모든 기술 시스템은 사회-기술 시스템이라는 것을 이해할 수 있다. 사회-기술 시스템은 가치 중립적이지 않으며, 정치적 쟁점을 가진다는 것을 이해할 수 있다.</p>
<p>a. 최적화(optimization)라는 용어를 이해하고 인간이 만든 사회-기술 시스템의 목표는 인간이 결정할 수 있다는 것을 인식할 수 있다.</p> <p>b. 일상생활 속에 있는 사회-기술 시스템을 알고, 사회-기술 시스템의 외재적 목표(advertised goals)와 실제적 목표(true goals)를 구분할 수 있다. (예를 들어, Youtube 추천 알고리즘은 실제로는 회사의 이익 창출을 추구 그러나 외부적으로는 사용자를 즐겁게 하기 위한 것으로 홍보한다.)</p> <p style="padding-left: 20px;">i. 현재 사용하고 있는 사회-기술 시스템의 특징과 확인된 목표(identified goals)를 연결할 수 있다.</p> <p>c. 분류 과정 속에서 발생할 수 있는 알고리즘 편향성(Bias)라는 용어를 알 수 있다.</p> <p style="padding-left: 20px;">i. 훈련 데이터(training data)가 머신러닝 시스템의 정확성(accuracy)에 미치는 영향력을 이해할 수 있다.</p>

<ul style="list-style-type: none"> ii. 인간이 훈련 데이터를 수집하고 구성할 수 있다는 것을 인식할 수 있다. ii. 훈련 데이터의 구성이 '지도 기반 머신러닝 시스템의 결과'에 어떻게 영향을 미치는지 이해할 수 있다.
<p>3. 사회-기술 시스템에는 많은 이해관계자가 있으며 시스템은 이러한 이해관계자들에게 각각 다른 방식으로 영향을 미칠 수 있다는 것을 인식할 수 있다.</p>
<ul style="list-style-type: none"> a. 사회-기술 시스템과 관련된 이해관계자들을 식별 b. 사회-기술 시스템의 결과에 대해 이해관계자가 관심 갖는 이유를 합리적으로 설명 c. 사회-기술 시스템에서 각각의 이해관계자가 가지는 가치 식별(예, 사용자의 요구를 충족하기 위해 시스템이 어떤 목표를 가져야 하는지?) d. 사회-기술 시스템을 중심으로 윤리 매트릭스 작성
<p>4. 사회-기술 시스템의 올바른 목표를 결정하기 위해 인공지능에 대한 기술적 이해와 이해관계자에 대한 지식을 적용할 수 있다.</p>
<ul style="list-style-type: none"> a. 윤리 매트릭스 분석하고 사회-기술 시스템의 새로운 목표를 위해 분석결과를 활용 b. 목표 달성을 위해 인공지능 시스템 훈련에 필요한 데이터셋을 식별할 수 있다. c. 사회-기술 시스템의 식별된 목표를 반영하거나 이해관계자의 가치를 반영한 기능을 설계할 수 있다.
<p>5. 기술이 세계에 미치는 영향력을 고려할 수 있다.</p>
<ul style="list-style-type: none"> a. 기술로 인해 발생하는 2,3차적 영향력을 이해하고, 기술이 다양한 이해관계자를 위해 만들어진다는 것을 알 수 있다.

2.5 인공지능 윤리교육 연구 분석

김정민은 “하버드대학에 개설된 ‘Embedded ethics’ 과목은 컴퓨터 과학 전공자를 대상으로 윤리적 합리성에 대한 교육을 진행하고 있으며 중국과학원대학교에서는 ‘인공지능의 철학과 윤리’ 과목에서 철학과 윤리학 관점에서 인공지능의 역사, 현재, 미래를 탐구하는 등 대학의 인공지능 윤리 사례를 소개하였다. 또한 인공지능 윤리의 중요성은 점차 커지고 있으며, 지금까지 동향으로 볼 때 사회적 확산은 잘 진행되고 있는 것으로 보이지만 인공지능 윤리가 다루는 내용은 원론에 가까운 윤리적 정의가 다수 포함되어 있어 내용의 보완이 필요하다”고 하였다[11].

김효은은 “인공지능은 자율시스템의 특성으로 인해 과학기술윤리에서 다루어지지 않은 내용이 도출되므로 인공지능윤리는 독자적 내용을 가지고 이에 따른 별도의 교과가 필요하다. 인공지능이라는 기술도 중요하나 그 재료가 되는 데이터를 수집, 분류, 선정하는 단계에서 이미 윤리적 시각이 핵심적 요소가 된다. 인공지능 설계 단계에서의 윤리적 요소들이 가져오는 결과의 차이를 생각한다면 관련 기관

과 기업에 연구자, 창업자, 알고리즘 설계자에게 이에 대응하는 인공지능 윤리교육을 해야만 한다”고 하며 인공지능 윤리과목의 필요성을 강조하였다[12].

마지막으로 김한성은 “AI가 2022년까지 5천 5백만 개의 일자리가 창출된다고 예측하며, 미래 사회를 살아갈 우리 아이들에게 미치는 영향력은 아주 크다는 것을 알 수 있고 이를 통해 볼 때, 경제학적 관점뿐만 아니라 교육학적 관심에서도 AI 기반 사회를 살아갈 아이들에게 AI와 관련된 기초 역량을 길러주는 것은 필수적이다. AI 시스템은 객관적이지 않을 수 있고 의도하지 않았던 결과를 초래할 위험이 있으므로 기술적 이해와 사회적 영향력에 대한 이해를 연계해서 가르쳐야 한다”며 인공지능 교육과 윤리교육 병행의 필요성을 강조했다[13].

2.6 인공지능 교육 플랫폼

지금까지 국내외의 인공지능 교육과 관련한 교육 목표 및 내용을 보았다. 인공지능 학습을 위해서는 수많은 데이터와 그 데이터를 학습시킬 수 있는 컴퓨팅 능력이 매우 중요하다. 그러나 인공지능 학습을 위해서 개별적으로 딥러닝 훈련 장비를 가지고 있기는 거의 불가능하다. 이 같은 이유로 해외에는 학생들이 인공지능 학습을 쉽게 할 수 있도록 온·오프라인에서 다양한 플랫폼을 제공하고 있다. <표 II-14>는 대표적인 플랫폼으로 보다 쉽게 인공지능 교육을 할 수 있도록 제공하고 있다.

<표 II-14> 인공지능 교육 플랫폼

플랫폼	훈련 모델
Quick draw	그림 인식
teachable Machine	이미지, 오디오, 동작
Machine Learning for kids	텍스트, 숫자, 이미지 생성된 모델을 스크래치에서 호환 가능
Cognimates	음성인식, 감정분석, 시각적 패턴 감지, 로봇 조정 생성된 모델을 스크래치에서 호환 가능

‘Quick draw’는 사용자가 직접 기계 학습에 참여하는 것으로 인공지능이 제시하는 그림을 그리는 활동을 통해 대답과 학습을 동시에 하는 플랫폼이다. ‘teachable Machine’은 구글에서 제공하는 플랫폼으로 이미지, 오디오, 동작 인식이 가능하며 보유하고 있는 데이터나 온라인 데이터를 이용해서 훈련 모

델을 만들 수 있다. ‘Machine Learning for kids’는 IBM에서 제공하는 플랫폼으로 텍스트, 숫자, 이미지를 인식할 수 있고 훈련된 모델과 스크래치를 이용해서 다양한 프로젝트를 진행할 수 있다. 마지막으로 ‘Cognimates’는 음성인식, 감정분석, 시각적 패턴 감지, 로봇 조종이 가능하며 역시 훈련된 모델을 이용해서 스크래치로 확장이 가능하므로 조금 더 쉽게 인공지능 기술 및 서비스 구현이 가능하다.

이상과 같이 인공지능 교육은 대상이 유아~고등까지 전 학년이 대상이고 인공지능의 이해를 시작으로 빅데이터, 머신러닝, 로봇까지 다양한 활동으로 구성되어있는 것을 알 수 있다. 또한 모델 훈련을 위한 다양한 인공지능 플랫폼도 이미 제공되고 있는 것을 알 수 있다.

3. 정보윤리 교육과 인공지능 윤리교육

3.1 정보윤리 교육

정보윤리 교육이란 정보화 사회에서의 바람직한 정보통신 생활과 관련된 건전한 인격을 함양하는 교육을 말한다. 학교에서 진행되는 전반적인 교육 활동에서 동시에 이루어질 수 있는 생활 교육의 성격을 지니고 있다. 예를 들어 존중, 책임감, 자율적인 행동 통제, 공동체 의식의 함양될 수 있는 내용으로 이루어지고 있다 [14]. 지금은 소프트웨어 교육의 한 영역으로 포함되어있으며 교육 내용은 사이버 폭력, 저작권 보호, 정보 검색과 활용, 개인정보 보호 등 8개의 내용[15]으로 구성되어있다. 주요 영역과 내용[16]은 <표 II-15>와 같다. 정보 검색과 활용 영역에서 ‘정보를 교환하거나 올바르게 수집하는 방법을 이해하는 교육’이나 저작권 보호 영역에서 ‘저작권의 이해와 보호’와 관련된 내용을 인공지능의 핵심 기술인 데이터와 관련했을 때 인공지능 윤리에서 가장 중요하게 다루어져야 하는 부분이다. 이처럼 정보윤리 교육은 인공지능 윤리와 분리된 듯 보이지만 실제로 새로운 기술이 나온다고 해서 관련 윤리도 새롭게 나오는 것은 아니므로 기존 윤리의 확장 선에서 추가되고 보완하는 측면으로 접근해야 한다. 예를 들어, 환자의 진료데이터를 이용해서 병증을 빨리 알아내기 위한 인공지능 프로그램을 개발하는 과정에서 환자의 진료데이터를 이용할 것으로 예상되며, 이 데이터에는 많은 분량의 개인정보가 포함

되어 있을 것이다.

<표 II-15> 정보윤리 교육 주요 영역[16]

영역	내용
사이버 폭력	사이버 폭력은 온라인 공간에서 발생하는 건전하지 않은 정보나 채팅을 통하여 이루어지는 욕설, 비방 등을 말하며 기타 온라인 공간에서 이루어지는 언어적 폭력 외에도 모든 폭력을 말한다.
(사이버) 네티켓	네티켓은 네트워크와 에티켓의 합성어로서 인터넷 공간에서 지켜야 할 예의를 의미한다. 네티켓은 인터넷을 사용하는 네티즌들이 스스로 사이버 공간에서의 예의를 지키고 건전한 사이버 공간을 사용하기 위해 반드시 요구되는 분야이다.
저작권 보호	저작권은 저작자가 창작한 저작물에 대해서 갖는 권리이다. 다양한 멀티미디어 자료 수집 및 편집, 공유가 활발하게 이루어지는 환경에서 저작권 이해와 보호는 중요하다. 저작권 보호에서 다루는 내용은 정보 공개, 정보 해킹, 자료 유형, 저작물, 인터넷 실명제 등이다.
스마트폰 · 인터넷 중독	스마트폰 · 인터넷 중독은 스마트폰이나 인터넷 활용에 있어 자기 조절이 어려운 상태에서 과몰입하여 일상생활을 유지하는 것이 힘든 상황을 의미한다. 스마트폰 · 인터넷 중독은 일상생활 등에서 올바른 판단을 막고 학생의 행동과 심리적 측면에서 여러 가지 문제들을 발생시킬 수 있다.
정보 검색과 활용	필요한 정보는 점점 더 증가하고 있으며 정보를 수집하고 가공하는 능력의 중요성이 더욱 강조되고 있다. 학생들이 정보를 교환하거나 올바르게 수집하는 방법을 이해하는 교육의 필요성이 커지고 있다.
건전한 정보기기 활용	정보기기의 발달은 사회적으로 많은 변화를 가져왔고 각 교과 학습에서 또한 정보기기의 사용이 흔한 모습이 되었다. 많은 학생이 스마트폰이나 정보기기를 잘못 활용함으로써 각종 중독 상황에 노출되는 상황이므로 건전하게 정보기기를 활용할 수 있는 교육이 필요하다.
진로 탐색	진로 탐색은 교과 내용 속에 포함된 진로교육 요소를 추출하고 진로 목표가 교과 목표와 연계되도록 구성한다. 진로 탐색 과정이 교육과정과 연계됨으로써 학생은 학교에서 익히는 교과 내용과 생활 속에서 자연스럽게 자신의 진로를 탐색할 기회를 얻게 된다. 정보윤리 교육을 통해 직업을 올바르게 이해하고 직업에 대한 건전한 의식을 함양할 수 있다.
개인정보 보호	현대가 개인 사회가 되면서 개인을 식별, 구별할 수 있는 개인정보 보호가 매우 중요해졌다. 개인 정보가 보호되지 않으면 한 개인의 안전과 재산에 문제를 일으키게 된다. 정보윤리 교육에서 다루는 개인정보는 보통 성명, 주민등록번호 등의 인적사항 정보가 가장 많은 비중을 차지하는데 이는 사회적으로 경제생활을 하지 않는 학생에게 생활 속에서 접하기 가장 쉬운 개인정보이기 때문이다.

이처럼 데이터를 이용해서 인공지능 서비스를 개발하는 경우 개인정보의 유출 문제가 발생할 수 있으므로 두 가지 측면에서 신중하게 고려해야 한다. 첫 번째는 개인정보 보호와 관련된 정보윤리 측면이며, 두 번째는 어떻게 하면 개인정보 유출 없이 유의미한 결과를 가져올 수 있는가 하는 측면이다. 즉 어떻게 하면 데이터의 유출없이 유의미한 데이터로 재가공할 수 있는지에 대하여 고민하는 과정과 그러한 방법 즉, 알고리즘을 찾는 과정에서는 인공지능 윤리 측면이 논의되어야 할 것이다. 기존에는 의미 없는 데이터라 하더라도 인공지능 알고리즘을 통해 가공되는 과정에서 이전에는 중요하지 않던 데이터도 새로운 정보로 가공되면서 개인정보의 유출 등 새로운 윤리문제를 일으킬 수 있다. 이러한 이유로 인공지능 윤리에 대한 접근은 기존의 정보윤리 교육의 토대 위에 추가되고 확장되어야 한다.

3.2 공학윤리 및 과학기술윤리

공학윤리 또한 정보윤리와 마찬가지로 새로운 기술이 생긴다고 기존의 관련 윤리가 무시되고 새로운 윤리가 출현되어야 하는 것은 아니다. 인공지능 윤리를 논할 때 거론되는 공학윤리나 과학기술윤리도 마찬가지이다. 즉 개발자, 기술자, 공급자를 위한 윤리라는 점에서 기존 윤리를 포함하고 확대·보완할 필요가 있다. 기존 윤리는 자동화 시스템 혹은 자동화 기계가 주가 된다면 인공지능 윤리는 지능적이고 자율적인 시스템으로 인간이 예측 불가능한 상황이 발생할 수 있으므로 개발자 윤리 측면에서도 좀 더 확장된 윤리가 필요하다.

<표 II-16>은 국제전자전기공학협회(IEEE), 미국기계공학협회(ASME), 미국토목공학협회(ASCE)가 제정한 공학윤리 헌장으로 공공의 안전과 복지라는 공통적인 목표를 두고 상세 내용을 기술하고 있다[17].

<표 II-16> 공학윤리 헌장[17]

○ IEEE <공학윤리헌장> 제1조
 모든 공학적 결정을 내림에 있어서 **공공의 안전과 건강과 복지에 부합하도록 하고, 공공이나 환경에 위해를 줄 요소들에 대해서는 즉각 공개한다.**

○ ASME <공학자의 윤리헌장> I. 기본규범

공학자는 전문가로서 자신의 의무를 수행할 때 다음과 같이 해야 한다.

공공의 안전, 건강 복지를 가장 중요한 사항으로 고려한다.

자신이 **감당할 능력이 있는 영역**의 서비스만 수행한다.

공적 발언은 오로지 **객관적이고 진실된 방식**으로 한다.

고용주나 고객에게 충실한 대리인 혹은 수탁자로 행동한다.

기만적인 행위를 하지 않는다.

명예롭게, **책임감을 가지고**, 윤리적으로, 그리고 법률에 부합하도록 행동함으로써 전문직의 명예와 평판과 유용성을 높인다.

○ ASCE <공학윤리헌장> 제1조

공학자는 **공공의 안전과 건강과 복지를 최고의 가치**로 여기며 전문가적 임무를 수행함에 있어 지속가능한 발전의 원칙을 준수하도록 노력해야 한다.

3.3 인공지능 윤리교육의 필요성

앞서 제시된 인공지능 윤리의 필요성 및 인공지능 교육의 현황 외에도 인공지능 윤리교육의 필요성을 근거할 몇 가지 사례를 보자.

먼저 인공지능 로봇을 하나의 인격체로 보는 경우이다. 구체적인 사례로 보스턴 다이내믹스사의 개발자가 개발한 사족 보행 로봇 ‘스팟’이다. 스팟은 다른 로봇들 보다 뛰어난 균형 능력이 있음을 강조하기 위해 스팟을 발로 차도 넘어지지 않는 모습을 담은 동영상을 올렸을 때 영상을 본 시청자들이 보여준 분노이다. 또 다른 사례는 미국 MIT에서 제작한 어린이 교육로봇 겸 소셜로봇 ‘허거블’이다. 허거블은 곰 모양의 로봇으로 어린 아이들에게 엄마보다 친근한 친구다. 거실을 청소하던 엄마가 단지 로봇인형에 불과한 허거블을 발로 툭 차며 밀어 넣는 모습을 본 아이는 이내 울음을 터뜨리면 자기 친구를 학대한 엄마에 대한 분노를 표현하기도 한다 [1]. 이 경우 단순히 인공지능 로봇에 인격을 부여하는 차원이 아니라 인격을 부여함으로써 인간에게 피해를 가하거나 인간보다 높은 위치에 두는 등의 윤리문제 발생이다. 따라서 이러한 문제를 예방하고 해결하기 위한 인공지능 윤리교육이 선행되어야 한다. 알고리즘 설계에서부터 로봇이 지켜야 할 윤리와 관련한 개발자의 윤

리교육과 인공지능 로봇을 대하는 사용자로의 윤리교육이 필요하다.

다음 사례는 온라인에서 자주 듣는 ‘추천’이라는 단어이다. 온라인 쇼핑몰 아마존의 상품 추천, Youtube의 영상 추천, 광고성 추천 메일 등 사용자에게 특화된 추천이 다양하게 사용되고 있고 여기에도 인공지능 알고리즘이 그 역할을 한다. 아마존을 예로 들면, 이 ‘추천’에 무엇이 표시될 것인가를 결정하는 것은 인간이 아닌 컴퓨터이다. 이용자의 조회 이력, 구매 이력 등을 기반으로 향후 이용자가 구입할 만한 상품이 무엇인지 판단하는 시스템에 인공지능 알고리즘이 활용되고 있다. 하지만, 이러한 추천과 관련한 알고리즘 개발에 개발자의 윤리교육이 부재한다면 상업적 목적을 위해 인간의 선택을 지배당하게 될 것이다.

2017년 일본에서 발표한 ‘AI 네트워크화 검토 회의 중간 보고서’에 의하면 인공지능이 우리 생활 곳곳에 자리하면서 발생하는 새로운 위험들이 생기고 있다고 발표하였다. 자율주행이나 로봇에 의한 ‘사고 위험’, 인공지능의 악용이나 무기가 테러에 사용되는 등의 ‘범죄 위험’, 다수를 위한 정책으로 인한 ‘일부 사람의 권리 이익이 손상될 위험’, 개인 정보 수집과 활용의 활성화와 지나친 추론을 통한 ‘프라이버시나 개인정보 침해 위험’, 인공지능이 인간의 의사 결정에 지나친 영향을 미치는 ‘인간의 존엄과 개인의 자율에 관한 위험’[40]등을 예상할 때에도 인공지능 윤리교육의 중요성은 더욱 강조되어야 한다.

III. 인공지능 윤리교육 프로그램 개발

1. 프로그램 개발 과정

1.1 목적 및 개발방향

인공지능 윤리교육은 윤리적인 측면에서의 접근과 소프트웨어 교육의 방향, 인공지능 기술 교육 등 다양한 관점에서의 접근과 내용 반영이 필요하다. 따라서 본문에서는 윤리교육의 접근 방향과 소프트웨어 교육 변화를 고려한 인공지능 윤리교육의 목적 설정을 통해 윤리의식 함양을 위한 인공지능 윤리교육 프로그램 개발을 진행하였다.

1.1.1 인공지능 윤리교육 방향

윤리는 인간 관계와 사회적 관계 전반에서 질서 유지를 위한 규범으로, 전통적 윤리학이 학문의 한 영역으로 자리 잡고 있다. 윤리와 유사한 의미로 도덕성은 개인의 행동원칙과 규범으로 인간은 누구나 도덕적인 사고를 하지만 반드시 도덕적 행동으로 연결되지는 않는다. 또한 사회 변화에 따라 그에 맞는 새로운 윤리가 필요하며 그런 이유로 윤리학은 사회 윤리, 의료 윤리, 문화 윤리, 과학 윤리 등의 이름으로 다양한 분야에서 다루어지고 있다[25]. 특히 정보윤리 분야는 응용윤리학이라는 이름으로 연구되고 있으며 인공지능 윤리도 응용윤리학의 범주로 보고 있다[25]. 응용윤리학은 실제 우리 삶의 여러 가지 영역에서 발생하는 구체적인 도덕적 문제들을 해결하고 실천하는 것을 목표로 삼는 학문이다.

본 논문에서 다루고자 하는 인공지능 윤리는 기존의 공학 윤리 혹은 과학기술윤리와 목적 면에서는 크게 벗어나지 않는다. 즉 공학윤리 또는 과학기술윤리 역시 과학적 창조활동을 함에 있어서 그 과정과 결과에 수반 되는 도덕적 위험을 판단하고 올바른 결정을 내릴 수 있도록 하는 윤리적 행위의 차원에서 접근되어야 한다. 이는 개발자 중심의 윤리로 그 결과물이 인간을 해하지 않고 이롭게 하는 기술로 사용되어야 한다는 것을 의미한다. 공학윤리는 자동화 시스템 또는 자동화 기계

에 관한 윤리로서 인간이 예측 가능한 범위의 윤리에 해당한다. 그러나 인공지능은 지능을 가진 자율적인 존재를 대상으로 하므로, 이전의 공학윤리나 과학기술윤리만으로는 인공지능 관련 윤리에 대한 설명이 충분하지는 않다[12]. 따라서 앞으로 인공지능 윤리는 지금까지의 일반적인 공학윤리의 연장 선상에서 포괄적으로 다루어지기보다는 확장하고 구분하여 접근할 필요가 있다.

특히 인공지능은 기존의 기술과는 달리 인류에 미치는 영향력이 막대하다. 지금까지 불가능했던 많은 일이 인공지능 기술을 통해 실현되고 있으며 이러한 특징은 놀라운 발전을 가지고 왔던 것은 사실이나, 그만큼 부수적으로 발생하는 부정적인 영향은 적지 않을 것으로 예상된다. 그러므로 인공지능 기술은 인공지능 윤리와 함께 나아갈 때 안전한 이용이 담보될 수 있으며, 우리는 인공지능의 안전과 윤리 문제를 해결해 나가면서 인공지능 기술을 발전시켜야 할 것이다[26].

1.1.2 컴퓨터 교육의 변화



[그림 Ⅲ-1] 소프트웨어교육의 흐름과 목표

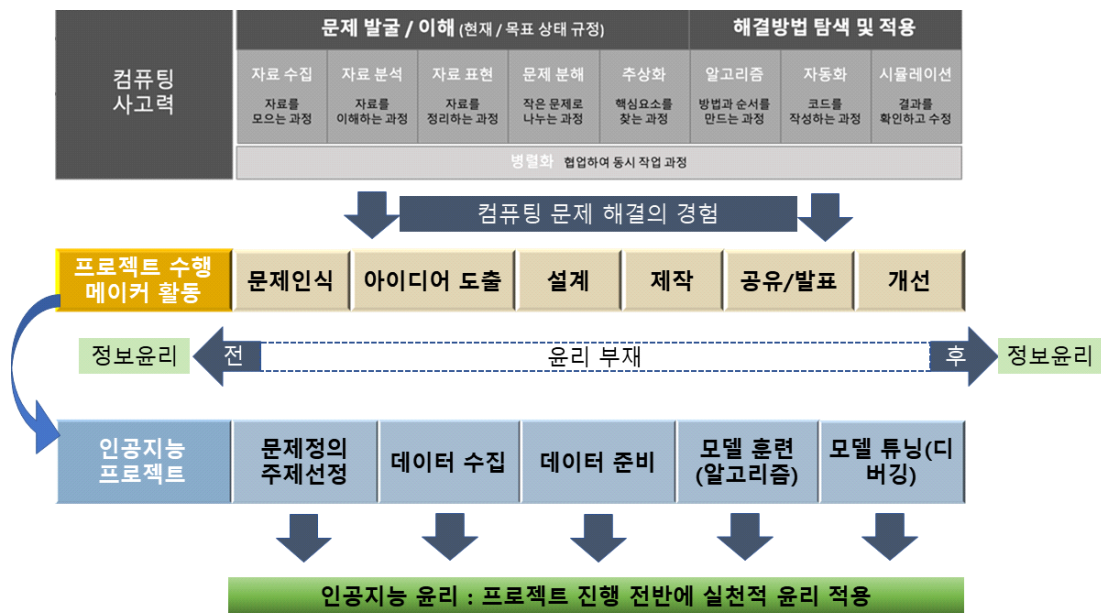
우리나라 컴퓨터 교육 즉, 정보 교육은 컴퓨터 활용교육을 시작으로 코딩교육, 알고리즘 교육으로 변화했고 이후 인공지능을 포함한 교육으로의 확대가 예고된다. 컴퓨터 활용 교육은 문서처리, 자료관리, 그래픽 작업 등 상용 프로그램을 이용하여 생활 과제를 처리하는 교육으로 시작하였다[29]. 이후 컴퓨팅사고력을 함양한 미래인재 양성이라는 시대 흐름에 따라 SW교육이라는 이름으로 확대되었으며 공교육에도 도입 되었다[4]. 컴퓨팅사고력은 컴퓨팅의 기본적인 개념과 원리를 기반으로 문제를 효율적으로 해결할 수 있는 사고능력으로 현대인이 갖추어야 할 기본

능력이다.

인공지능 교육은 기존의 SW교육에 포함되어 활용 교육부터 메이커 교육까지 다양하게 진행될 것이다[30]. SW교육에서 인공지능 교육을 위해서는 코딩교육, 알고리즘 교육이 선행된 바탕에서 이루어져야 하며 더불어 다양한 응용 프로그램을 활용할 수 있는 교육도 빠져서는 안 된다.

1.1.3 인공지능 윤리교육의 목적

2015년 개정된 교육과정을 기점으로 국내 학생들은 컴퓨팅사고력 향상이라는 목표 아래 메이커 교육이 명시적으로 제시되고 있다. 메이커 활동은 문제를 해결하는 과정에서 이루어지는 과정으로 누구나 하는 보편적 활동이다. 또한 정보나 과학에 국한되지 않고 모든 교과나 체험 활동으로도 확장될 수 있도록 고안되었다. 또한 앞으로 인공지능 기술을 경험하고 활용하는 교육에서 메이커 교육은 학생들에게 우선시되는 교육방법 중에 하나로써 사용 될 것이다. 이러한 측면에서 [그림 III-2]와 같이 기존 메이커 활동의 프로세스와 인공지능 프로젝트 수행 프로세스를 비교 제시하였다.



[그림 III-2] 인공지능 윤리교육 진행 방안[27]

컴퓨팅사고력 향상을 목표[4]로 한다는 점에서는 기존의 SW교육이나 인공지능 교육이 차이를 보이지는 않는다. 그러나 문제를 해결하는 과정으로의 메이커 활동을 기준으로 프로젝트 수행 프로세스를 비교해 보면 윤리 교육적 측면은 다소 부족하다는 점을 알 수 있다. 실제로 교육 가능한 시간의 부족이라는 물리적인 문제도 있고, 교육을 진행한다 하더라도 프로젝트의 시작에서 정보윤리의 중요성을 전달하거나 프로젝트 마지막에 결론과 함께 잠시 언급하는 정도에 그쳐있음을 알 수 있다. 물론 이 같은 형태도 정보윤리를 올바르게 인식하는 과정에서 도움이 되는 측면도 존재하지만, 알고 있는 것과 알고 있는 것을 실천하는 과정에는 큰 차이가 있을 수 있다[28].

본 연구의 목적은 첫째, 인공지능의 엄청난 파급력과 함께 인간이 예측 불가능한 상황이 발생할 수 있고 개발자의 의도에 따라 악용될 수 있으므로 새로운 윤리교육을 통해 역기능을 최소화해야 한다. 둘째, 인공지능 윤리교육이 실천으로 이어질 수 있도록 프로젝트 수행의 모든 단계에 걸쳐 윤리문제를 확인하는 과정 및 실천 과제를 제공해야 한다. 이러한 두 가지 목적을 두고 개발을 진행하였다.

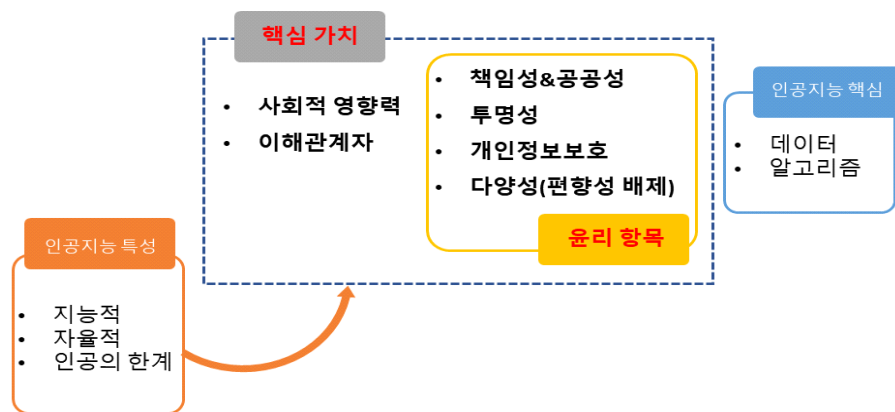
1.2 인공지능 윤리 핵심가치와 윤리 영역

인공지능을 이용한 기술이나 서비스의 핵심 요소는 알고리즘과 데이터라고 할 수 있다. 알고리즘은 어떠한 문제 해결을 위한 절차나 방법으로 알려져 있으나 인공지능 알고리즘은 규칙이 생성되는 방식에서 기존 알고리즘과 차이가 존재한다. 기존 알고리즘은 예측 가능하고 통제 가능한 결과를 가져오는 반면에 인공지능 알고리즘은 기계 학습을 진행하는 과정에서 자체 규칙이 구축되고 그 과정은 인간이 알지 못하는 내용과 과정이 발생하게 된다. 이처럼 알고리즘의 적용과 함께 기계 학습에 사용되는 데이터에 따라 전혀 다른 결과를 출력할 수 있으므로 알고리즘과 데이터는 인공지능 기술의 핵심이라고 할 수 있다.

인류가 발명한 대부분의 발명품의 최우선 가치는 인간이었고, 앞으로도 그럴 것이다. 개발자들은 개발된 제품이나 서비스가 사회에 이로운 영향력을 끼치고 관련된 여러 이해관계자를 염두에 두고 사용되기를 바라는 것이 현실이다. 일부 잘못된 의도를 가지고 개발하거나 오용 및 남용으로 인한 피해도 있지만 인공지능 기술의 특성이나 사회적 영향력을 고려한다면 최소화할 수 있는 노력이 시급한 시점이라

고 할 수 있다.

이상과 같은 논의에 따라 [그림 Ⅲ-3]에 제시한 인공지능 윤리 항목은 국내·외의 인공지능 윤리 가이드라인과 개발자 윤리를 분석하여 공통요소를 도출하고 중고등학교 수준에서의 프로젝트 개발에 꼭 필요한 영역으로 선정한 결과이다 [31][32][33].



[그림 Ⅲ-3] 인공지능 윤리교육에 필요한 핵심가치와 윤리항목

선정된 윤리 항목에 대하여 본 논문에서는 다음과 같이 정의하였다. 「책임·공공성」은 인공지능 시스템이 객관적으로 평가 가능해야 하고 이용자를 포함한 모든 이해관계와 영향력에 대한 책임 및 책임 소재가 명확해야 한다는 것으로 정의한다. 이때 이해관계자는 사용자, 개발 및 연구자, 공급자를 포함하는 것이다[34]. 「투명성」은 인공지능 시스템에 의한 결정은 사람이 이해할 수 있고 추적 가능해야 하며 위험과 관련한 정보를 공개·공유할 수 있고 개인정보 처리의 전 과정은 적절하게 이루어져야 한다는 것을 의미한다. 「개인정보보호」는 인공지능 시스템에 의해 수집되고 활용되는 모든 개인정보가 안전하고 비공개적이어야 하며 침해당하거나 개방되어서는 안 된다는 것을 강조한 것이다[35]. 마지막으로 「다양성」은 인공지능 시스템은 연령, 성별, 인종 등의 특성과 무관하게 모든 사람이 사용할 수 있어야 하고 공정하게 다루어져야 하며 불공정한 편향성을 지녀서는 안 된다는 것을 가리킨다.

1.3 인공지능 윤리교육 모델 개발

인공지능 프로젝트를 수행할 프로세스는 일반적으로 데이터와 훈련 알고리즘 중심으로 진행된다. [그림 Ⅲ-4]를 구체적으로 설명하자면, 문제 정의 및 주제를 선정한다. 선정된 주제에 맞는 데이터를 수집, 훈련에 적합한 형태로 수정한다. 이러한 데이터셋으로 목적에 맞는 알고리즘으로 훈련시키고 튜닝 하는 과정으로 진행되는 프로세스이다. 프로젝트의 전·후에 앞서 도출된 인공지능 핵심가치와 윤리 항목을 바탕으로 사전·사후 진단 문항을 도출하여 적용하였다. 다음으로 프로젝트를 구현하는 프로세스 5단계에 걸쳐 각 단계에서 가장 비중을 두고 실천해야 하는 윤리 항목을 중심으로 실천 가능한 윤리 항목을 개발하였다. 인공지능 윤리의 경우 공통으로 다루어야 할 윤리 영역 80%와 수행하는 프로젝트별로 다른 윤리 영역을 20%를 분리하여 개발하였다. 20%의 경우는 실제 현장에서 수행하는 프로젝트에 맞추어 적용하고 이 논문에서 제시하는 윤리 영역은 공통적인 80% 부분이다.



[그림 Ⅲ-4] 인공지능 윤리교육 모델

SW에서 기존 알고리즘은 예측이나 통제가 가능한 결과를 가져오지만 인공지능 알고리즘은 기계학습 과정에서 자체 규칙이 구축되고 그 과정은 인간이 알지 못한다. 따라서 알고리즘과 함께 기계학습 과정의 재료로 사용되는 데이터는 중요한 요소로 작용하며 인공지능 프로세스는 데이터 수집 및 처리와 알고리즘을 중심으로 진행되어야 한다. [그림 III-4]의 인공지능 윤리교육 모델은 인공지능 교육 프로세스, 프로세스의 주제선정 단계에서 사회적 영향력과 이해관계자에 대한 이해, 프로세스 모든 단계에서 적용한 윤리항목으로 구성되어 있으며 단계별 상세 내용은 아래와 같다.

1.3.1 문제정의 및 주제선정

프로세스의 1단계는 문제정의 및 주제선정 단계로 우리 주변의 문제를 인식하고 인공지능 기술을 이용해서 해결 방안을 찾는 과정이다. 인공지능 기술 자체는 객관적이고 중립적이다. 기획하는 단계에서 사회적 영향력에 대한 파급효과와 이해관계자를 충분히 고려한 주제를 선정 및 해결 방안을 모색하여야 한다. 또한 만들어지는 기술이나 서비스에 대한 책임과 공공의 이익을 중심에 두고 기획되어야 한다.

1.3.2 데이터 수집

2단계 데이터 수집은 문제해결에 필요한 데이터를 수집하는 단계로 알고리즘 요건에 맞는 데이터를 충분히 수집해야 한다. 수집된 모든 데이터는 개인정보들이 안전하게 관리 되어야 하며 개인정보 처리 과정은 투명하게 이루어져야 한다. 데이터는 결과가 왜곡되지 않을 만큼 충분한 양이 준비되어야 하며 다양하게 구성되어야 한다. 알고리즘에 따라 이미지, 음성, 텍스트, 모션 등 다양한 데이터가 있다.

1.3.3 데이터 준비

데이터의 준비는 수집된 데이터를 훈련에 적합한 형태로 수정·가공하는 과정을 말한다. 이렇게 준비된 데이터가 알고리즘에 직접 사용될 데이터로 5단계의 과정 중 인공지능 윤리가 필요한 가장 핵심적인 단계이다. 편향된 데이터는 의도하지 않는 결과를 초래하기도 하고 자칫 개발자의 잘못된 의도가 반영되어 인류에 해가 되는 결과물이 도출될 수도 있다. 따라서 의도한 결과를 위해서 다양한 데이터의

준비와 투명성이 무엇보다 중요하다.

1.3.4 모델 훈련 및 알고리즘

모델 훈련은 준비된 데이터를 가지고 훈련시킬 최적의 알고리즘을 선택하는 단계이다. 학생을 대상으로 하는 인공지능 교육에서는 알고리즘 직접 만드는 과정은 생략되고 이미 만들어져 있는 오픈 소스 알고리즘을 사용하게 된다. 사용하는 알고리즘은 최초 개발자의 사용 허가 범위에서 사용하여야 하며 허락 없이 수정하거나 재배포해서는 안 된다. 훈련된 결과는 조작 없이 공개하여야 하며 결과가 개인정보를 침해하거나 공공의 이익에 반하여서는 안 된다.

1.3.5 모델 튜닝 및 디버깅

훈련을 통해서 만들어진 모델은 다른 하드웨어나 소프트웨어와 결합하여 새로운 기술이나 서비스로 만들어진다. 자연어처리 기술과 음성을 인식하는 모델을 이용하여 인공지능 스피커를 만들고 더 나아가 개인 비서 로봇을 만드는 것이 이에 해당한다. 이런 과정으로 만들어진 완성된 서비스에서 예상되는 부작용에 대한 충분한 토론이나 해결 방안도 함께 진행되어야 한다.

2. 인공지능 윤리교육 모델을 적용한 프로젝트

인공지능 윤리교육을 위해 개발한 인공지능 교육 모델을 바탕으로 단계별 세부 프로그램의 내용과 학습자의 인공지능 교육 내용, 실천 윤리 내용, 사전·사후 체크리스트[36]에 대하여 논의하면 다음과 같다.

2.1 프로젝트 소개

- 프로젝트의 주제 : 우리 생활 주변에서 불편한 문제를 찾아내고 인공지능이라는 기술을 이용해서 긍정적인 해결방법을 모색하자. 그 예로 ‘우리 동네 Hero!’ 라는 주제로 진행하였다.

- 프로젝트 목적 : 공중화장실에서의 범죄 예방을 목표로 하며, 성별과 반대의 화장실에 들어가거나 장애인이나 노약자가 성별에 맞는 화장실 입구를 찾지 못하는 상황 등을 고려하여 범죄 예방, 불편 해소 등을 목적으로 한다. <표 Ⅲ-1>은 ‘우리 동네 Hero!’ 의 예시 프로젝트로 입력-처리-출력 형태로 표현하여 제시한 것이다.

<표 Ⅲ-1> 프로젝트 개요

입력	처리	출력
화장실 출입구 : 이미지분류기 작동	-입력된 남/여 값에 따라 유형 정의 -C형일 경우 : 추천 프로그램 적용 -추천 프로그램은 스크래치로 프로그래밍	-A형. 올바른 입장 : 출력 없음 -B형. 입장 대기 시간이 길 경우 : 남녀 안내 -C형. 잘못 입장할A 경우 : 추천 알고리즘에 의해 이미지나 사운드 출력

2.2 세부 프로그램 내용

<표 Ⅲ-2>의 세부 프로그램은 인공지능 프로젝트를 개발하는 과정 중심으로 문제정의 및 주제선정-데이터 수집-데이터 준비-모델 훈련-모델 튜닝의 5가지 프로세스에 맞춘 학습내용과 활동내용을 정리하여 제시한 것이다. 학습내용은 인공지능 기술이나 교육을 중심으로 기술하고 활동내용은 인공지능 교육과 함께 인공지능 윤리교육을 포함하는 내용으로 구성하였다. 활동내용과 수집 데이터의 경우 어떤 프로젝트를 진행하느냐에 따라 내용이 다를 수 있으나, 진단 내용은 단계별로 필수적으로 점검해야 하는 내용으로서, 어떠한 형태의 프로젝트를 기획하더라도 공통으로 수행해야 하는 과정으로 설정하였다.

<표 Ⅲ-2> 세부 프로그램

단계	프로 세스	학습 내용	활동 내용	수집 및 진단 항목
1 단 계	문제 정의 및 주제	-인공지능 알고리즘 경험하기	-이미지 · 텍스트 인식 알고리즘 -자율주행 시스템 -추천 알고리즘	

	선정 (2차시)	-인공지능 기술 이해 -인공지능과 사회적 영향력 이해 -이해관계자 의미의 이해	-인공지능 기술을 이용해서 작지만 중요한 우리 주변의 문제를 찾아서 해결하기 위한 프로젝트 구상 -구상한 프로젝트의 사회적 영향력과 이해관계자 찾기	'우리 동네 Hero' 프로젝트 -공중화장실 성별 구분 -추천프로그램 : 음악, 이미지, 비디오 등
		-인공지능 윤리교육의 필요성 인식하기	-인공지능 프로젝트를 구상하는 과정에서 윤리 문제가 배제되었을 때를 가정하고 토론하기	-인공지능 윤리의 핵심가치 : 사회적 영향력, 이해관계자 배려 ※ 윤리 핵심 가치, 공공성, 책임성, 투명성
2 단 계	데이터 수집 (1차시)	-현대 사회에서 데이터의 의미 -올바른 데이터를 수집 하는 방법 -적절한 데이터의 크기 알기	-우리 주변의 빅데이터 활용 사례 알아보기 -저작권에 대한 올바른 인식과 이해하기 -개인정보보호의 중요성에 대한 인식과 이해하기	-인물 사진, 이미지 수집 -사람의 감정에 영향을 주는 이미지, 사운드, 비디오 수집 ※ 개인정보보호, 투명성
3 단 계	데이터 준비 (2차시)	-수집한 데이터를 훈련에 적합한 형태로 수정, 변형 -최종 데이터셋 분석	-이미지 사이즈, 위치에 맞게 라벨링 하기 -프라이버시나 보안 문제가 발생할 수 있는 데이터 찾고 수정하기 -분류 과정에 연령, 인종, 성별 등 데이터 편향성이 없는지 검증하기 -다양한 데이터가 존재하는지 검증하기	-이미지 라벨링 -감정과 관련된 자료 분류하고 태그 달기 ※ 데이터 편향성, 다양성, 투명성
4 단 계	모델 훈련 (2차시)	-이미지 인식 알고리즘 사용하기 -추천 알고리즘 만들기 -알고리즘 사용 관련 투명성 이해하기	-훈련 데이터로 남녀 이미지 인식 학습시키기 -테스트 데이터로 남녀 인식을 확인하기 -인식률이 낮다면 이유를 확인하고 정리 -최종 데이터로 다시 학습시키기	-machine learning for kids 플랫폼 -스크래치 -알고리즘 개발보다 기존 알고리즘 활용 ※ 알고리즘 투명성, 개인정보보호
5 단 계	모델 튜닝 (1차시)	-완성된 분류기를 이용하여 인공지능 서비스 만들기 -검증 후 오류 수정하기	-남녀 인식을 확인하고 문제점 발견하기 -잘못된 입장을 했을 경우 추천음악을 플레이해주는 프로그램 만들기 -이미지 분류기와 음악프로그램 연결하기	-성찰과 비판적 사고 ※ 투명성, 책임·공공성

2.3 교수·학습 지도안

<표 III-3> 1-2차시 지도안

< 1~2차시 지도안 : 1단계 문제정의 및 주제선정 >

학습 목표	<ul style="list-style-type: none"> ○ 인공지능 알고리즘을 체험하고 이해할 수 있다. ○ 인공지능 기술을 이용한 아이디어를 찾을 수 있다. ○ 인공지능 윤리의 필요성을 알 수 있다.
윤리 항목	<ul style="list-style-type: none"> ○ 인공지능 윤리 핵심 가치 : 사회적 영향력, 이해관계자 배려 ○ 윤리 항목 : 공공성&책임성, 투명성

단계	교수·학습 활동	자료 및 유의점
도입	<ul style="list-style-type: none"> ■ 우리 주변에서 인공지능 기술이나 시스템 알아보기 ○ EBS 프로그램 시청 : '인공지능 - 지능 만들기' 편집 ○ '지능을 만든다.'라는 기술에 대한 의미에 대해 이야기해 보기 ○ 내가 알고 있는 인공지능 기술이나 시스템 나열해보기 ○ 우리도 만들 수 있을까? 누구나 만들 수 있어야 하나? ○ 인공지능 기술은 인간을 위해 꼭 필요한 기술인가? 	<ul style="list-style-type: none"> ※ 영상 시청 후 전체 발표 및 팀별 토론으로 진행 인공지능에 대한 생각과 관심을 가 지는 시간
전개	<ul style="list-style-type: none"> ■ 인공지능 알고리즘 체험하기 ○ 머신러닝 기술 체험 <ul style="list-style-type: none"> - Quick Draw 플랫폼을 이용해서 내가 그리는 그림을 인공지능 시스 템이 이해하는지 확인 - 그동안 학습된 학습데이터 확인 ○ 이미지 인식 분류기 사용하기 <ul style="list-style-type: none"> - Congnimates 플랫폼을 이용하여 사과와 수박 이미지 훈련시키고 바 르게 인식하는지 확인 ■ 인공지능 알고리즘 이용한 간단한 사례 안내 <ul style="list-style-type: none"> ○ 음성 인식 기술 ○ 추천 알고리즘 ○ 자율주행 자동차 	<ul style="list-style-type: none"> ※ HTTP://quickdraw. withgoogle.com ※ http://cognimates. me ※ 사과와 수박 이 미지 수집은 직접 그린 그림으로 훈 련 ※ 여러 가지 알고 리즘 소개로 인공 지능기술이 다양한

■ '우리 동네 Hero! 프로젝트' 기획

- 체험한 인공지능 기술을 이용하여 우리주변의 문제를 해결할 수 있는 프로젝트 기획
- 제주도는 관광도시의 특성상 주변에 공중화장실이 많이 있다. 공중화장실 사용 시 안전에 도움을 주는 인공지능 시스템 구상해보기
- 선정 주제 : '우리 동네 Hero, 1탄! 안전한 공중화장실'

목적 : 공중화장실에서의 안전한 사용과 편리함을 제공
 - 범죄 예방 차원과 장애인·노약자에게 성별에 맞는 입장 유도
 - 잘못된 출입구로 입장하려고 할 때 강제적으로 막는다거나 기분이 상하지 않도록 이미지나 사운드 제공

■ 기획된 프로젝트 검증

- 완성된 서비스라고 가정
- 이 서비스가 사회에 주는 영향력은 무엇일까?
- 이 서비스를 사용할 사람들은 누구일까?
 - 서비스 사용자 나열
 - 사용자별로 어떤 이로운 점이 있는지 작성
- 이 서비스로 인해 의도치 않은 불편을 겪는 사람은 없을까?
 - 피해를 보는 사람 나열
 - 불편한 점 작성
- 사용자를 포함한 이해관계자를 표로 만들어 정리

구분	이해관계자1	이해관계자2	이해관계자3	이해관계자4
사용이유				
장점				
단점				

■ 완성된 '인공지능 핵심 가치' 매트릭스 발표

- 팀별로 작성된 매트릭스 발표
- 우리 팀과 비교하여 수정·보안

■ 인공지능 윤리의 중요성 정리

- 기획부터 검증까지 진행했던 활동을 기반으로 인공지능 윤리 항목과 연계하여 설명
 - 사회적 영향력과 이해관계자의 배려가 인공지능 개발의 핵심가치임을

분야에 활용되고 있는 부분 강조

※ 프로젝트 기획 시 유의점

초급자의 경우 같은 주제와 내용으로 진행할 수 있도록 유도

중급자 이상일 경우 팀별로 주제만 주고 세부 내용은 팀별로 기획할 수 있도록 유도

※ 사회적 영향력, 이해관계자에 대한 의미를 이해시키고 작업

※ 언플러그드 활동 전지와 포스트잇을 이용해서 작성

설명하고 프로젝트 전반에 걸쳐 고려해야 함

■ 실천 윤리 확인

중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	내가 기획한 프로젝트는 완성 후 사회에 어떠한 영향력을 끼칠 것인가에 대해 고려하고 기획하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	인공지능 기술이나 서비스를 개발하면서 서비스 사용자 입장에서 충분히 고려하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	기획한 프로젝트는 공공의 이익에 도움이 되는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
인공지능 기술이나 서비스를 기획하는 과정에서 안전이나 가격 등 여러 가지 고려해야 할 사항 중 자신이 생각하는 최우선 순위는 무엇인가요?		

정리

■ 바꾸어 생각해 보기

○ 인공지능 윤리가 배제된 상태에서 기획되고 만들어진 서비스가 사회에 미치는 영향력에 대해 생각해보기

■ 다음 차시 안내

- 데이터 수집
- 데이터 수집에도 윤리가 있다.

IV. 인공지능 윤리교육 프로그램 검증

1. 프로그램 대상 및 검증 방법

1.1 연구 가설

본 연구의 내용은 인공지능 윤리교육을 위한 인공지능 교육 모델을 개발하는 것을 목표로 연구를 진행하였다. 프로젝트를 수행하는 활동으로 1단계는 인공지능 윤리에 대한 인식도에 대한 사전 진단을 하는 과정을 논의하였다. 2단계는 프로젝트를 수행하는 5개의 프로세스에서 가장 중요하다고 선정된 윤리항목에 대해 실천 학습을 하고 3단계는 사후 진단을 하는 과정을 제시하였다. 이상의 과정을 통해 모든 프로젝트에서 인공지능 윤리의 중요성을 인식하고 기획부터 설계, 완성까지 인공지능 윤리를 실천하도록 하는 데 목적을 두고 연구가 진행되었다.

- 가설 : 개발한 인공지능 윤리교육 모델을 적용했을 때 인공지능 윤리 의식 함양 및 실천력 향상에 유의한 영향을 끼칠 것이다.

단, 사회 현상을 고려해서 학생들에게 교육을 직접 적용하는 대신에 전문가 집단을 구성하는 방식으로 프로그램의 타당도를 검증할 수 있도록 하였다.

1.2 프로그램 적용 대상

- 블록 코딩의 경험이 있고, 인공지능 기술에 대한 이해가 없는 중학생을 대상으로 교육용 플랫폼과 스크래치 사이트 활용
- 프로그래밍 경험이 많고 인공지능 프로젝트에 관심이 많은 중학생을 대상으로 머신러닝 플랫폼, 파이선 라이브러리 활용
- 팀 프로젝트 수업으로 팀별 4~5명, 총 25명 내외
- 사회적 상황을 반영하여 전문가 설문을 통한 델파이 검사로 진행

1.3 연구 검증 방법

전문가 10인을 구성하여 1차 설문 후 설문내용을 반영하여 2차 설문을 진행하였

다. 2번의 설문과 수정을 통해 프로그램에 적용하였으며 <표 IV-1>은 설문에 참여해 주신 전문가의 구성 및 관련 경력 사항이다.

<표 IV-1> 전문가 구성

구분	직책	전공	경력	학위
전문가 1	교수	컴퓨터과학	16년	박사
전문가 2	교수	인공지능	10년	박사
전문가 3	교수	인공지능	10년	박사
전문가 4	교수	컴퓨터공학	8년	박사
전문가 5	교수	컴퓨터교육	10년	박사
전문가 6	교사	교육공학	18년	박사
전문가 7	교사	컴퓨터교육	7년	석사
전문가 8	교사	컴퓨터교육	2년	석사
전문가 9	교사	컴퓨터교육	9년	석사
전문가 10	연구원	교육학	6년	박사

전문가 집단은 교수와 교사 등 현직 종사자 비중이 대다수이며, 전공은 컴퓨터과학, 교육공학, 교육학 등으로 고루 분포하며 경력은 평균 9.6이며 과반수는 박사 학위 이상 전문가 집단이라고 할 수 있다.

<표 IV-2> CVR 산출식과 응답인원에 따른 타당도

산출 공식	응답자 수	CVR 최솟값
$CVR = \frac{N_e - \frac{N}{2}}{\frac{N}{2}}$ <p>N = 응답 사례 수 Ne = '중요하다'고 응답한 빈도수 (척도에서 '중요하다', '매우 중요하다'에 응답한 빈도수를 합한 것)</p>	5 ~ 7	.99
	8	.75
	9	.78
	10	.62
	11	.59
	12	.56
	13	.54
	14	.51
	15	.49
	20	.42
	25	.37
	30	.33
	35	.31
	40	.29

설문은 리커트 5점 척도를 적용하였고 각 문항에 대한 타당도 여부는 내용타당도로 판단하였으며, CVR은 Lawshe(1975)가 개발한 산출 공식을 적용하였으며 전문가 수에 따른 타당도 값과 산출 공식은 <표 IV-2>와 같다.

또한 도출된 의견들의 합의점 도달 여부를 판단하기 위해 합의도와 수렴도를 산출하였으며 통계에 필요한 중앙값과 사분위수 Q1, Q3을 사용하였고 델파이 조사 결과 자료 분석 기준은 <표 IV-3>과 같다.

<표 IV-3> 델파이 분석 도구 기준 값

구분	분석 기준	의미 또는 해석	선행연구
CVR(내용타당도비율)	0.62 이상	델파이 조사에 참여한 전문가의 수가 10명일 때 내용타당도 비율 값이 0.62이상이면 내용타당도가 있다고 판단	Lawshe(1975)
수렴도 = $\frac{Q_3 - Q_1}{2}$ ※ Q1 : 1사분위수 Q3 : 3사분위수	0.5 이하	델파이 전문가의 의견 일치도를 보기 위한 것으로 완전 합의 되었을 때 값은 0이며, 의견의 편차가 클수록 수렴도 수치는 증가	이종성(2001)
합의도 = $1 - \frac{Q_3 - Q_1}{M_d}$ ※ Md : 중앙값	0.75 이상	델파이 전문가의 의견 일치도를 보기 위한 것으로 완전 합의 되었을 때 값은 1이며, 의견의 편차가 클수록 합의도 수치는 감소	이종성(2001)
안정도 = $\frac{SD}{M}$	0.5 이하	응답의 일치성이 높은 경우 안정도가 확보됨 0.5~0.8 비교적 안정적 0.8 이상은 추가 설문이 필요	

1.4 연구 검증 결과

<표 IV-4>의 델파이 분석 결과 CVR은 1차 0.74, 2차 0.90으로 내용 타당도가 있다고 판단되므로 타당성이 검증되었다고 할 수 있다. 수렴도와 합의도는 각각 0.33, 0.85로 의견 일치도 만족하여 합의 되었으며 안정도는 0.12로 높은 일치성을 보였다.

<표 IV-4> 전문가 설문 통계 결과

구분	기준 값	1 차	2 차
평균(5점 척도)	5 점 (최댓값)	4.44	4.62
CVR(내용타당도)	0.62 이상	0.74	0.90
수렴도	0.5 이하	0.43	0.33
합의도	0.75 이상	0.82	0.85
안정도	0.5 이하	0.17	0.12

2. 전문가 설문 내용

2.1 인공지능 윤리항목의 적절성

제시된 윤리항목 4가지를 학생들의 인공지능 윤리교육에 적용하는 경우 항목 선정 및 항목의 정의가 적절한가를 설문한 내용이다.

<표 IV-5> 윤리항목 및 정의 설문내용 최종

1.1 윤리항목	응답					1.2 정의	응답				
	1	2	3	4	5		1	2	3	4	5
(1:전혀 그렇지 않다 ↔ 5:매우 그렇다)											
다양성(편향성 배제)						인공지능 시스템은 연령, 성별, 인종 등의 특성과 무관하게 모든 사람이 사용할 수 있어야 하고 공정하게 다루어져야 하며 불공정한 편향성을 지니서는 안 된다.					
수정 및 보완 의견											
투명성						인공지능 시스템에 의한 결정은 사람이 이해할 수 있고 추적 가능해야 하며 위험과 관련한 정보를 공개·공유할 수 있고 개인정보 처리의 전 과정은 적절하게 이루어져야 한다.					
수정 및 보완 의견											
책임 · 공공성						인공지능 시스템은 평가 가능해야 하고 이용자를 포함한 모든 이해관계와 영향력에 대한 책임 및 책임 소재가 명확해야 하며, 이해관계자는 사용자, 개발 및 연구자, 공급자를 포함한다.					
수정 및 보완 의견											
개인정보보호						인공지능 시스템에 의해 수집되고 활용되는 모든 개인정보는 안전하고 비공개적이어야 하며 침해당하거나 개방되어서는 안 된다.					
수정 및 보완 의견											

설문 결과 ‘다양성이라는 항목은 두 가지 단어를 모두 명시하면서 정의에서 주로 의도하고자 했던 편향성 배제가 명확히 드러나지 않는다.’, ‘책임에 대해 강조하는 만큼 추후 자율성과의 공존에 대해서 고려할 필요가 있다.’, ‘공공성에 관한 설명이 부족하다.’ ‘모두 책임을 지는 것이 아닌 책임 소재가 명확해야 하는 것으로 정의 하는 것도 고려해야 한다.’, ‘공정하게 다루어야 하며는 신뢰성이라는 항목으로 구분하여 다양한 정보의 신뢰성을 검토하고, 그 중 가장 신뢰성 있는 정보를 체계적으로 선택하는 것으로 윤리항목을 추가할 수 도 있을 것 같다.’, ‘현재의 인공지능 기술은 모든 과정을 사람이 이해하는 데 한계가 있다. 상황에 따라서는 입력과 출력의 윤리에 대해서만 고민하는 것이 옳바르다고 생각한다.’ 와 같은 의견을 반영하여 최종적으로 <표 IV-5>와 같은 윤리항목 선정 및 윤리항목별 정의를 내렸다.

2.2 인공지능 윤리교육 모델의 적절성

인공지능 교육을 위한 프로세스와 프로세스 단계마다 윤리항목을 적용한 모델이 적절한가를 설문한 내용이다.

<표 IV-6> 인공지능 윤리교육 모델 설문내용 최종

2. 프로세스별 인공지능 윤리 적용 내용		응답				
		1	2	3	4	5
1:전혀 그렇지 않다 ↔ 5:매우 그렇다)						
1단계. 문제정의 & 주제선정 (프로젝트 주제 결정)	윤리 핵심 가치, 공공성, 책임성에 높은 비중을 두고 다룬다.					
[수정 및 보완 의견]						
2단계. 데이터 수집 (데이터 요건 정의 및 데이터 수집)	개인정보보호, 투명성에 높은 비중을 두고 다룬다.					
[수정 및 보완 의견]						
3단계. 데이터 준비 (데이터를 훈련에 적합한 형태로 변환)	다양성, 편향성 배제에 높은 비중을 두고 다룬다.					
[수정 및 보완 의견]						
4단계. 모델 훈련, 알고리즘 (주제에 맞는 최적의 알고리즘 선택)	투명성, 개인정보보호에 높은 비중을 두고 다룬다.					
[수정 및 보완 의견]						
5단계. 모델 튜닝, 디버깅 (인공지능 서비스 적용, 결과 분석 및 공유)	투명성, 책임 · 공공성에 높은 비중을 두고 다룬다.					
[수정 및 보완 의견]						

설문 결과 ‘인공지능 윤리와 관련된 사례를 발견하고 문제를 정의하여 프로젝트 주제 결정하면 좋겠다.’, ‘저작권뿐만 아니라 초상권으로 인한 문제가 증가하고 있어 함께 다루었으면 한다. 특히 인공지능 기술의 도입으로 범죄에 악용될 가능성이 높다.’, ‘데이터 수집 단계에서도 편향성이 나타날 수 있다.’, ‘모든 단계에 모든 윤리항목이 필요한 것 같다.’ 등의 의견을 반영하여 최종적으로 <표 IV-6>의 인공지능 윤리교육 모델을 구성하였다.

2.3 교수·학습 지도안에 제시된 윤리 확인 문항의 적절성

인공지능 교육과 함께 이루어지는 프로세스별 실천윤리 확인 문항은 5단계의 프로세스 중심으로 작성되었으며 이해 대해서 적절한가에 대한 설문내용이다.

<표 IV-7> 프로세스별 윤리 항목 설문내용 최종

3. 프로세스별 실천윤리 확인 문항(프로세스 중심으로)						
3.1 문제정의 및 주제선정						
번호	문제정의 및 주제선정	응답				
		1	2	3	4	5
1	내가 기획한 프로젝트는 완성 후 사회에 어떠한 영향력을 끼칠 것인가에 대해 고려하고 기획하였는가? [수정 및 보완]	사회적 영향력 등				
2	인공지능 기술이나 서비스를 개발하면서 서비스 사용자 관점에서 충분히 고려하였는가? (예, 마스크를 생산하면서 필터의 성능을 높이는 것을 목표로 했을 때 마스크를 구매하고 사용해야 하는 사람의 입장을 생각해보자.) [수정 및 보완]	이해관계자 배려 등				
3	기획한 프로젝트는 공공의 이익에 도움이 되는가? [수정 및 보완]	공공성, 책임성 등				
4	인공지능 기술이나 서비스를 기획하는 과정에서 안전이나 가격 등 여러 가지 고려해야 할 사항 중 자신이 생각하는 최우선 순위는? [수정 및 보완]					
추가 문항 의견						
3.2 데이터 수집						
번호	데이터 수집	응답				
		1	2	3	4	5
1	타인의 개인정보가 포함되지 않고, 초상권을 침해하지 않은 데이터를 수집하였는가?	투명성 등				

	[수정 및 보완]					
2	저작권의 사용표기와 출처를 정확히 하였는가?	개인정보보호				
	[수정 및 보완]					
3	원하는 모델을 생성하기 위해 저작권이 부여된 데이터를 저작권 허락 없이 사용하여려는 욕망이 발생하였는가?	개인정보보호				
	[수정 및 보완]					
4	수집된 데이터가 객관적이고 신뢰성이 있는지에 대한 검증방법으로 아래 예시에서 제안하고 싶은 것이 있는가? (예를 들어, 남녀를 구분하는 방법으로 키나 헤어스타일 같은 신체 특징 이외의 방법)					
	추가 문항 의견					

3.3 데이터 수정

번호	데이터 수정	응답	12345				
			1	2	3	4	5
1	수집된 데이터 종류는 인종이나 지역, 성별 등으로 차별하거나 의도적으로 무시하지 않고 다양하게 구성되었는가?	편향성, 다양성 등					
	[수정 및 보완]						
2	데이터 가공 시 민감한 데이터나 개인정보를 해칠 우려가 있는 데이터는 제외하였는가?	편향성 등					
	[수정 및 보완]						
3	차별적 결과를 도출할 수 있는 데이터 쓸림 현상이 발생하지 않을 만큼 고르게 분포시켰는가? (예, 여자 사진 50장, 남자 사진 100장과 같은 경우 여자 사진을 제대로 인식할 확률이 낮아 질 수 있음)	다양성 등					
	[수정 및 보완]						
4	데이터를 수집하면서 개인적인 취향이나 선호도에 치우치지 않았는가?	다양성 등					
	[수정 및 보완]						
5	훈련된 결과에 대해 내가 원하는 결과가 나오지 않아도 결과를 조작 없이 정직하게 공개하고 훈련 데이터를 수정하여 사용하였는가?	책임·공공성 등					
	[수정 및 보완]						
6	내가 예측하는 결과를 유도하기 위해 '편향'되거나 '왜곡'시키는 편집 행위가 없었나요?						
	[수정 및 보완]						
	추가 문항 의견						

3.4 모델 훈련(알고리즘)

번호	모델 훈련(알고리즘)	응답	12345				
			1	2	3	4	5
1	체계적인 정보 공유 등 정직한 프로세스에 맞추어 모델링 하였는가?	책임성, 투명성 등					

	[수정 및 보완]							
2	오픈소스 알고리즘의 의도를 이해하고 용도에 맞게 사용하였는가?	투명성 등						
	[수정 및 보완]							
3	내가 개발해보고 싶은 인공지능 관련 알고리즘이 있나요? (정답은 없습니다. 여러분의 생각을 적어주세요)							
	[수정 및 보완]							
추가 문항 의견								
3.5 모델 튜닝(디버깅)								
번호	모델 튜닝(디버깅)		응답					
			1	2	3	4	5	
1	팀 내에서 사용자, 개발자, 공급자 등 역할을 나누고 완성된 서비스에서 예상되는 부작용 및 해결 방안에 대해서 토론하였는가?	책임성, 공공성 등						
	[수정 및 보완]							
2	완성된 서비스에 대해서 의도치 않게 발생하는 역기능에 대해 책임감을 가지고 공개하고 수정할 의무가 있다고 생각하는가?	책임성, 공공성 등						
	[수정 및 보완]							
3	본 서비스 관련자들과 해당 서비스의 기능, 역할, 부작용 등에 대해서 책임을 가지고 의견을 수렴할 통로를 마련하였는가?	책임성, 공공성 등						
	[수정 및 보완]							
4	'우리 동네 Hero 2탄!, 0000000'을 만든다면 어떤 프로젝트를 기획하고 싶은가? (사회에 미칠 영향력과, 누구를 위한 프로젝트인가를 중심으로 작성)							
	[수정 및 보완]							
추가 문항 의견								

설문 결과 ‘실제 모델 학습에 사용하는 데이터에 대하여 개인정보를 해칠 우려가 있는지 살펴보는 것이 그리 쉽지 않다. 그러므로 사후 조치보다는 사전 초지로 데이터 수집 시 배제하는 것이 더 효과적일 것 같다.’, ‘어쩔 수 없이 데이터 풀림 현상이 발생할 상황도 많으므로 전체 개수보다 고르게 분포시켰는지 중요하다.’, ‘오픈소스 알고리즘이 개인정보와 관련된 것이라 보기 합당한지 확인하자.’, ‘개발해본 경험자 대상의 질문은 아닌지 검토하자.’, ‘인공지능 윤리의 질문으로 합당한지 확인이 필요하다. 기술에 대한 질문인 것 같다.’, ‘구체적인 사례와 연결하여 중학생이 이해하기 쉬운 문항으로 수정해야 한다.’와 같은 다양한 의견을 반영하여 <표 25>의 프로세스별로 실천해야 할 윤리 문항을 완성하였다.

2.4 사전·사후 인공지능 윤리 문항에 대한 적절성

사전·사후 문항은 [그림 III-3]에 도출된 내용을 기반으로 인공지능 교육 전후에 진단하는 문항으로 인공지능 윤리 핵심가치와 윤리 항목 중심으로 구성된 설문 내용이다.

<표 IV-8> 사전·사후 윤리항목 중심의 설문내용 최종

4. 사전·사후 진단 문항(윤리 항목 중심으로)						
4.1 인공지능 윤리 핵심 가치						
번호	인공지능 윤리 핵심 가치	응답				
		1	2	3	4	5
1	인공지능 기술이나 서비스는 기본적으로 인간 삶의 질을 향상 시키는데 기여해야 한다고 생각 하는가? [수정 및 보완]					
2	인공지능 기술이나 서비스는 긍정적 혹은 부정적이든 사회 변화에 영향을 끼칠 수 있다고 생각 하는가? [수정 및 보완]					
3	인공지능 기술이나 서비스는 기본적으로 인간의 자유와 존엄성을 존중해야 하는가? [수정 및 보완]					
추가 문항 의견						
4.2 다양성&편향성 배제						
번호	다양성&편향성 배제	응답				
		1	2	3	4	5
1	개발의 전 과정에서 성, 인종, 종교, 지역, 민족을 내가 속한 집단의 이익을 이유로 차별하지 않았는가? [수정 및 보완]					
2	인공지능 기술이나 서비스를 이용함에 있어 사회적 약자 및 취약 계층 등을 고려한 별도의 서비스가 필요한가? [수정 및 보완]					
3	민감 데이터 또는 개인 데이터의 이용을 최소화 하거나 사용하지 않고 훈련 할 수 있는 방법을 고려했는가? [수정 및 보완]					
추가 문항 의견						
4.3 알고리즘 투명성						
번호	알고리즘 투명성	응답				
		1	2	3	4	5
1	오픈 알고리즘에 개발자(나)만 알고 있는 숨은 기능을 넣어도 괜찮은가? [수정 및 보완]					
2	최종적인 의사결정은 인공지능 시스템이 아니라 인간이 할 수 있어야 하는가? [수정 및 보완]					

3	훈련에 사용된 데이터의 양은 결과 도출에 필요한 만큼 적절한가? [수정 및 보완]						
4	알고리즘의 성능을 위해 의도적인 수정을 하였는가? [수정 및 보완]						
추가 문항 의견							
4.4 책임성&공공성							
번호	책임성&공공성	응답					
		1	2	3	4	5	
1	최초 개발부터 이용에 이르기까지 전 과정과 결과에 대한 책임을 공유해야 할 책임이 있는가? [수정 및 보완]						
2	위급상황 발생 시 해결을 위해서 필요 데이터를 공유해야 할 책임이 있는가? (예, 회사의 특허와 관련된 중요한 기술이지만 발생한 위급상황을 해결하기 위해 공개를 요구할 경우) [수정 및 보완]						
3	인공지능 기술이나 서비스에 사용된 데이터는 공공의 이익을 위한 경우라도 소수의 사람을 해롭게 하거나 차별하는데 사용되어서는 안 되는가? [수정 및 보완]						
추가 문항 의견							
4.5 개인정보보호							
번호	개인정보보호	응답					
		1	2	3	4	5	
1	다수의 사람들에게 이로운 서비스를 위해서 소수의 저작권, 초상권, 개인정보 는 희생되어도 괜찮은가? [수정 및 보완]						
2	나와 관련한 데이터의 사용에 대해 알고 있어야 하고 원치 않을 경우 사용을 거부할 수 있어야 하는가? [수정 및 보완]						
3	암호화, 익명화 등 개인정보보호 강화를 위한 조치는 필요한가? [수정 및 보완]						
추가 문항 의견							

설문 결과 ‘가치가 충돌하는 상황에 관한 내용 등이 추가된다면 설문의 진단 효과가 높아진다.’, ‘윤리항목별로 구분된 질문이 중복되는 것 같다.’와 같은 의견을 반영하여 <표 IV-8>의 윤리항목 중심으로 한 사전·사후 진단 문항을 완성하였다.

이상과 같이 인공지능 윤리교육 모델을 개발하고 전문가 설문으로 타당도 검사를 진행하여 중학생에게 적용 가능한 인공지능 윤리교육 모델을 완성하였다.

V. 결론

우리는 지능정보사회를 살아가고 있다. 급격한 기술의 발달로 사회, 문화, 경제, 교육은 물론 개인의 삶도 변화하고 있다. 그러나 인공지능 기술은 이전의 기술과 달리 사회에 미치는 영향력과 파급효과가 일반적인 상식 이상으로 크다고 할 수 있다. 인공지능 관련 기술이 주는 긍정적인 면은 인간에게 행복을 주지만 잘못된 의도로 만들어진 인공지능은 인간에게 심각한 해를 끼칠 수 있다는 사실에 주목할 필요성이 있다. 이러한 양면성 때문에 인공지능 기술은 인공지능 윤리와 함께 발전할 때 가장 편리하고 안전한 기술이 될 수 있다--

기존의 SW교육에 포함되어 있는 정보윤리 교육은 인공지능의 개념과 특성을 반영하지 못하는 한계가 있고 지식 전달 위주의 교육이 이루어지고 있다. 따라서 본 논문에서는 이러한 한계점을 극복하고, 실천력 중심의 인공지능 윤리교육을 위하여 메이커 교육의 단계마다 고려해야 하는 인공지능 윤리항목을 반영하고 적용하는 ‘인공지능 윤리교육 모델’을 개발하였다. 본 논문에서 제시한 프로그램 개발의 구체적인 과정은 다음과 같다.

첫째, 인공지능 기술의 핵심인 알고리즘과 데이터를 기반으로 한 인공지능 윤리항목을 도출하였다.

둘째, 인공지능 교육을 위한 프로세스로 ‘문제정의 및 주제선정’, ‘데이터 수집’, ‘데이터 준비’, ‘데이터 훈련’, ‘모델 튜닝’의 5단계를 정리하고 각 단계마다 중점적으로 적용하기 위해 타당하고 적합한 윤리 항목을 연결하여 제시하였다.

셋째, 단계별로 적용할 윤리항목과 관련하여 평가 문항을 만들고, 프로젝트 전후에 윤리 의식 진단을 위한 설문 문항을 개발하여 제시하였다.

넷째, 개발된 윤리항목과 인공지능 교육 프로세스를 기반으로 교수학습 지도안을 개발하고 전문가 집단을 구성하여 인공지능 윤리교육 프로그램에 대한 타당도 검사를 통해 그 타당성을 검증하였다.

본 연구의 내용과 결과를 바탕으로 후속 연구를 위한 제언은 다음과 같다.

첫째, 본 논문에서 개발된 프로그램에서는 책임·공공성, 다양성, 투명성, 개인정

보보호를 윤리항목으로 선정하였다. 하지만 여기에 언급되지 않은 더 많은 윤리항목이 있으며, 이를 교육에 적용하는 연구가 이어져야 할 것이다.

둘째, 본 논문에서 개발한 프로그램을 실제 학생들을 대상으로 해당 교육 현장에서 적용하여 한계를 찾아 보완하고 실천력 향상에 도움이 되는지에 대한 검증이 필요하다.

셋째, 교육 시수를 고려하였을 때 인공지능 교육만으로도 부족한 경우의 기본교육에서 대체하는 등의 다른 대안으로써의 윤리교육이 필요하다.

[참고 문헌]

- [1] 김명주(2017), 인공지능 윤리의 필요성과 국내외 동향, 한국통신학회지(정보와 통신), 34(10), 45-54.
- [2] 김윤정(2018). 인공지능 기술 발전에 따른 이슈 및 대응 방안 : 한국과학기술기획평가원, 2018-34(통권 제 252호).
- [3] 한국정보화진흥원(2019). 2018년 스마트폰 과의존 실태조사.
- [4] 교육부(2015). 2015 개정 교육과정.
- [5] 강인애, 윤희진, 황중원(2017). 메이커 교육= Maker education :4차 산업혁명 시대에 다시 만난 구성주의.
- [6] 서형준(2019). 4차 산업혁명시대 인공지능 정책의사결정에 대한 탐색적 논의. 정보화정책 제26권. 3-35.
- [7] 한상기(2016). 인공 지능 기술의 사회적 이슈와 윤리 문제. 한국멀티미디어학회 지, 20(3), 41-52.
- [8] 한국정보화진흥원(2019). 인공지능 윤리 가이드라인-일본과 EU 사례를 중심으로. 지능정보사회 법제도 이슈리포트.
- [9] IEEE(2019). EAD1e(ETHICALLY ALIGNED DESIGN,First Edition) 보고서.
- [10] 한국정보화진흥원(2018). 지능정보사회 윤리 가이드라인.
- [11] 김정민(2019). 인공지능 윤리 이슈와 교육 과정 동향. 월간SW중심사회_2019년_7월호.
- [12] 김효은(2020). 의사결정 자동화에 대한 대응으로서의 인공지능 윤리교육. 한국 윤리교육학회, 윤리교육연구, Vol. 55.
- [13] 김한성(2020). 인공지능의 이해와 사회적 영향력에 관한 교육 프로그램 개발 및 적용. 컴퓨터교육학회.
- [14] 강신혜(2018). 실천력 향상을 위한 스크래치 기반의 정보윤리 교육 프로그램. 석사학위논문. 제주대학교 교육대학원.
- [15] 교육부(2014). | 교육과정 기반 정보통신윤리 교육사례 및 활용가이드. 교육자료 TL-2014-1.

- [16] 이맹화(2008). 초등학교 저학년 학생들의 정보통신윤리 실천력 강화를 위한 知·情·意 통합 교수·학습 모형 개발 및 적용. 석사학위논문. 한국교원대학교 교육대학원.
- [17] 한국정보화진흥원(2016). 2016 정보문화포럼 정책연구보고서. 지능정보사회윤리분과, 인공지능에 관한 윤리.
- [18] 한국과학창의재단(2019). 초·중등 인공지능 교육 내용체계 탐색적 연구 이슈 보고서.
- [19] 한국과학창의재단(2019). 2019 글로벌 SW교육 컨퍼런스 이슈페이퍼. AI 시대, SW교육을 다시 생각하다.
- [20] ISTE 2019 (PA, Philadelphia, 2019년 6월 23일 -26일) 세션의 “인공지능에 대한 K-12 지침: 학생들이 알아야 할 사항” 이란 주제로 AI4K12 이니셔티브에 대한 토론 발표 자료. 번역 : 2019, 미래인재연구소.
- [21] 이승철, 김태영(2020). 초등학생을 위한 인공지능 교육 내용 및 방법 제안. 한국컴퓨터교육학회 학술발표대회논문집, 24(1), 177-180.
- [22] 김수환, 김성훈, 김현철(2019). 해외 인공지능 교육동향과 학습도구 분석. 한국컴퓨터교육학회 학술발표대회논문집, 23(2), 25-28.
- [23] 한국교육학술정보원(2019), 김한성, 전수진, 최승윤, 김성애. 모두를 위한 인공지능 윤리.
- [24] Blakeley H. Payne(MIT)(2019). An Ethics of Artificial Intelligence Curriculum for Middle School Students.
- [25] 문동규(2004). 정보기술에 대한 철학적 숙고 : 하이데어 사유를 중심으로. 범한철학 34, 2004.09, 177-201(25 pages)
- [26] 한국인공지능윤리협회(2019). 인공지능 윤리 헌장, 서문.
- [27] 조정원(2018). 2018 제주로on코딩 컴퓨팅사고력 정의 및 소프트웨어 교육체계 고도화 결과보고서.
- [28] 강신혜, 조정원(2018). 실천력 강화를 위한 EPL기반 정보윤리 교육 프로그램. 한국컴퓨터교육학회 동계학술 발표 논문지. 22(1). 25-28.
- [29] 우현정, 조혜정, 최울(2018). ICT를 활용한 교육의 동향과 전망. 정보화정책 제25권 제4호, 3-36.

- [30] 이철현(2020), AI 시대 역량 관점의 실과 소프트웨어교육의 방향. 한국실과교육학회 학술대회논문집. 95-123.
- [31] 한국정보화진흥원(2016). 2016 정보문화포럼 정책연구보고서. 지능정보사회윤리분과, 지능정보사회에서의 인공지능의 출현과 새로운 윤리학의 모색.
- [32] 한국정보화진흥원(2016). 2016 정보문화포럼 정책연구보고서. 지능정보사회윤리분과, 소비자 관점에서 본 지능정보사회의 윤리적 이슈.
- [33] 가트너 특약(2019). <https://www.thescoop.co.kr/news/articleView.html?idxno=36550>, 윤리적인 AI 만드는 5가지 방법. 검색 : 2020. 2. 18.
- [34] 교육부(2017). 디지털 기술-세상을 바꾸는 디지털 기술.
- [35] 정보통신정책연구원(2018). 4차산업혁명시대 산업별 인공지능 윤리의 이슈 분석 및 정책적 대응방안 연구 보고서
- [36] 최유라, 이태욱(2018). 효과적인 정보윤리의식과 컴퓨팅사고력 함양을 위한 초등 소프트웨어 융합 교육 지도의 실제. 한국컴퓨터정보학회 학술발표논문집 26(2)
- [37] 유럽연합집행위원회(2019). AI윤리가이드라인, 신뢰할 수 있는 AI 실현을 위한 요건.
- [38] 아실로마컨퍼런스(2017). <https://futureoflife.org/ai-principles-korean>, 아실로마 AI원칙 23. 검색 : 2020. 2. 1.
- [39] 윤혜선(2018). 인공지능 기술을 윤리적으로 탐하다. KIST 저널(33), 48-52.
- [40] 생활을 변화시키는 인공지능(2018). 다쿠치 카즈히로, 모리시마 료코, 양성건 번역

[ABSTRACTS]

Development of Project-Based Artificial Intelligence Ethics Education Program

JinAh Bae

Graduate School of Education Jeju National University
(Major in Computer Education)

Supervised by Professor Jungwon Cho

We are living in the era of intelligent informatization. Technologies such as big data, block chain, and artificial intelligence are leading social changes, and artificial intelligence and robots, among them, are classified as the most influential technologies along with biotechnology. Also, artificial intelligence is a unique software with intelligent and autonomous, it has a significant development speed and ripple effects, and behind these technologies are adverse effects. Therefore, in order for artificial intelligence technology to be used safely and beneficially, ethics and ethics education are needed together in response. In line with this change, various software education is being conducted in the education community for training future convergence talents that had both computational thinking, and future software education will be expanded to include artificial intelligence. Information ethics education in existing software education has limitations in reflecting the characteristics or contents of artificial intelligence, and there is also a limit to the existing teaching methods that are briefly mentioned or omitted as the level of knowledge transfer at the beginning or end of the

problem-solving process. Therefore, the research purpose in this thesis is developing a new AI ethics education reflecting the characteristics of artificial intelligence, and as an important tool for problem-solving when artificial intelligence education was conducted, developing an "Artificial Intelligence Ethics Education Model" that applied ethics to know and practice at each stage of problem-solving.

The research contents of this thesis analyzed that AI education contents and research trends based on domestic and foreign AI ethics guidelines and existing ethics. The contents developed based on the analysis are as follows.

First, derived the core value of artificial intelligence ethics and the ethical items that could understand the algorithms and data that are the core of the technology.

Second, developed a five-step education process that reflected the computational thinking process, which is the goal of software education.

Third, developed the artificial intelligence ethics education program that applies the ethical items and the five-step artificial intelligence education process.

Last, go through expert verification and developed an artificial intelligence ethics education program that can be applied to middle school students.

The developed model is an ethics education program that reflects the characteristics and contents of artificial intelligence technology, and it made that agonized and applied ethical items to each problem-solving process through SW.

Through the 'Artificial Intelligence Ethics Education Model' developed in this thesis, it is hoped that it will be the basis for establishing itself as an ethical education that can be practiced.

[부록]

부록1. 학습 목표

1단계 : 문제정의 및 주제선정	
학습목표	<ul style="list-style-type: none"> 인공지능 알고리즘을 체험하고 이해할 수 있다. 인공지능 기술을 이용할 수 있는 아이디어를 찾을 수 있다. 인공지능 윤리의 필요성을 알 수 있다.
윤리항목	<ul style="list-style-type: none"> 인공지능 윤리 핵심 가치 이해하기 : 사회적 영향력, 이해관계자 배려 윤리 항목 : 공공성 및 책임성, 투명성
주요활동	<ul style="list-style-type: none"> 생활 속 인공지능 찾아보기 인공지능 기술을 이용하여 우리 주변의 문제를 해결할 수 있는 프로젝트 기획하기 인공지능 윤리가 배제된 상태에서 기획되고 만들어진 서비스가 사회에 미치는 영향에 대해 생각해보기 인공지능 윤리 의식 사전 진단
2단계 : 데이터 수집	
학습목표	<ul style="list-style-type: none"> 현대사회에서 데이터의 의미를 이해할 수 있다. 기획한 프로젝트에 맞는 데이터를 수집할 수 있다.
윤리항목	<ul style="list-style-type: none"> 윤리 항목 : 개인정보보호, 투명성
주요활동	<ul style="list-style-type: none"> 인공지능 기술을 이용한 아이디어를 구현할 방법 설계하기 데이터를 이해하고 분류하여 수집하기 내가 만든 그림이나 정보에 이름과 의미를 부여하고 저작권, 초상권 등이 중요한 이유를 생각해보기 빙고 활동을 통해 하나의 무의미한 데이터가 재가공 될 때 어떻게 의미가 바뀌는지 이해하고 생각해보기
3단계 : 데이터 수정	
학습목표	<ul style="list-style-type: none"> 데이터가 왜 중요한지 이해할 수 있다. 수집된 데이터를 훈련에 적합한 형태로 변형할 수 있다. 데이터에 따른 훈련 결과를 확인하고 수정할 수 있다.
윤리항목	<ul style="list-style-type: none"> 윤리 항목 : 다양성, 데이터 편향성 배제, 투명성
주요활동	<ul style="list-style-type: none"> 인공지능 알고리즘의 실패 사례를 알아보기 수집된 데이터의 분류 없이 훈련시키기 수집된 데이터를 카테고리별로 분류하고 편향성 제거 후 훈련시키기 사람의 편식과 인공지능의 편식이 사회에 미치는 영향에 대해 생각해보기

4단계 : 모델 훈련	
학습목표	<ul style="list-style-type: none"> • 기획 의도대로 프로젝트를 완성할 수 있다. • 인공지능 플랫폼을 이용하여 훈련할 수 있다. • 훈련 결과를 분석하고 수정할 수 있다.
윤리항목	<ul style="list-style-type: none"> • 윤리 항목 : 개인정보보호, 투명성
주요활동	<ul style="list-style-type: none"> • 인공지능 알고리즘을 선택하여 모델 훈련하기 • 이미지 인식 알고리즘에 준비된 데이터로 훈련하기 • 완성된 이미지 분류기 테스트하기 • 테스트 결과에 따라 재훈련을 하기 • 이미지 인식 알고리즘 외에 다른 알고리즘 알아보기
5단계 : 모델 튜닝	
학습목표	<ul style="list-style-type: none"> • 완성된 이미지 분류기를 이용하여 인공지능 서비스를 만들 수 있다. • 다른 팀의 프로젝트를 보고 비판할 수 있다.
윤리항목	<ul style="list-style-type: none"> • 윤리 항목 : 투명성, 공공성 및 책임성
주요활동	<ul style="list-style-type: none"> • 완성된 분류기로 우리 주변의 문제를 해결하는 도구로 활용하기 • 스크래치나 앱 인벤터 등 EPL 프로그램과 연결하여 프로젝트 완성하기 • 자아 성찰과 비판적 사고의 필요성 인식하기 • 프로젝트 시연 및 발표하기 • 인공지능 윤리 의식 사후 진단

부록2. 교수·학습 지도안

< 1~2차시 지도안 : 1단계 문제정의 및 주제선정 >

학습 목표	<ul style="list-style-type: none"> ○ 인공지능 알고리즘을 체험하고 이해할 수 있다. ○ 인공지능 기술을 이용한 아이디어를 찾을 수 있다. ○ 인공지능 윤리의 필요성을 알 수 있다.
윤리 항목	<ul style="list-style-type: none"> ○ 인공지능 윤리 핵심 가치 : 사회적 영향력, 이해관계자 배려 ○ 윤리 항목 : 공공성&책임성, 투명성

단계	교수·학습 활동	자료 및 유의점
도입	<p>■ 우리 주변에서 인공지능 기술이나 시스템 알아보기</p> <ul style="list-style-type: none"> ○ EBS 프로그램 시청 : ‘인공지능 - 지능 만들기’ 편집 ○ ‘지능을 만든다.’라는 기술에 대한 의미에 대해 이야기해 보기 ○ 내가 알고 있는 인공지능 기술이나 시스템 나열해보기 ○ 우리도 만들 수 있을까? 누구나 만들 수 있어야 하나? ○ 인공지능 기술은 인간을 위해 꼭 필요한 기술인가? 	<p>※ 영상 시청 후 전체 발표 및 팀별 토론으로 진행</p> <p>인공지능에 대한 생각과 관심을 가지는 시간</p>
전개	<p>■ 인공지능 알고리즘 체험하기</p> <ul style="list-style-type: none"> ○ 머신러닝 기술 체험 <ul style="list-style-type: none"> - Quick Draw 플랫폼을 이용해서 내가 그리는 그림을 인공지능 시스템이 이해하는지 확인 - 그동안 학습된 학습데이터 확인 ○ 이미지 인식 분류기 사용하기 <ul style="list-style-type: none"> - Congnimates 플랫폼을 이용하여 사과와 수박 이미지 훈련시키고 빠르게 인식하는지 확인 <p>■ 인공지능 알고리즘 이용한 간단한 사례 안내</p> <ul style="list-style-type: none"> ○ 음성 인식 기술 ○ 추천 알고리즘 ○ 자율주행 자동차 <p>■ ‘우리 동네 Hero! 프로젝트’ 기획</p> <ul style="list-style-type: none"> ○ 체험한 인공지능 기술을 이용하여 우리 주변의 문제를 해결 할 수 있는 프로젝트 기획 ○ 제주도는 관광도시의 특성상 주변에 공중화장실이 많이 있다. 공중화장실 사용 시 안전에 도움을 주는 인공지능 시스템 구상해보기 ○ 선정 주제 : ‘우리 동네 Hero, 1탄! 안전한 공중화장실’ <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>목적 : 공중화장실에서의 안전한 사용과 편리함을 제공</p> <ul style="list-style-type: none"> - 범죄 예방 차원과 장애인·노약자에게 성별에 맞는 입장 유도 - 잘못된 출입구로 입장하려고 할 때 강제적으로 막는다거나 기분이 상하지 않도록 이미지나 사운드 제공 </div>	<p>※ https://quickdraw.withgoogle.com</p> <p>※ http://cognimates.me</p> <p>※ 사과와 수박 이미지 수집은 직접 그린 그림으로 훈련</p> <p>※ 여러 가지 알고리즘 소개로 인공지능기술이 다양한 분야에 활용되고 있는 부분 강조</p> <p>※ 프로젝트 기획시 유의점</p> <ul style="list-style-type: none"> - 초급자의 경우 같은 주제와 내용으로 진행할 수 있도록 유도 - 중급자 이상일 경우 팀별로 주제만 주고 세부 내용은 팀별로 기획할 수 있도록

록 유도

■ 기획된 프로젝트 검증

- 완성된 서비스라고 가정
- 이 서비스가 사회에 주는 영향력은 무엇일까?
- 이 서비스를 사용할 사람들은 누구일까?
 - 서비스 사용자 나열
 - 사용자별로 어떤 이로운 점이 있는지 작성
- 이 서비스로 인해 의도치 않은 불편을 겪는 사람은 없을까?
 - 피해를 보는 사람 나열
 - 불편한 점 작성
- 사용자를 포함한 이해관계자를 표로 만들어 정리

구분	이해관계자1	이해관계자2	이해관계자3	이해관계자4
사용이유				
장점				
단점				

※ 사회적 영향력, 이해관계자에 대한 의미를 이해시키고 작업

※ 언플러그드 활동 전지와 포스트잇을 이용해서 작성

■ 완성된 '인공지능 핵심 가치' 매트릭스 발표

- 팀별로 작성된 매트릭스 발표
- 우리 팀과 비교하여 수정·보안

■ 인공지능 윤리의 중요성 정리

- 기획부터 검증까지 진행했던 활동을 기반으로 인공지능 윤리 항목과 연계하여 설명
 - 사회적 영향력과 이해관계자의 배려가 인공지능 개발의 핵심가치임을 설명하고 프로젝트 전반에 걸쳐 고려해야 함

■ 실천 윤리 확인

중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	내가 기획한 프로젝트는 완성 후 사회에 어떠한 영향력을 끼칠 것인가에 대해 고려하고 기획하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	인공지능 기술이나 서비스를 개발하면서 서비스 사용자 관점에서 충분히 고려하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	기획한 프로젝트는 공공의 이익에 도움이 되는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
인공지능 기술이나 서비스를 기획하는 과정에서 안전이나 가격 등 여러 가지 고려해야 할 사항 중 자신이 생각하는 최우선 순위는 무엇인가요?		

정리

■ 바꾸어 생각해 보기

- 인공지능 윤리가 배제 된 상태에서 기획되고 만들어진 서비스가 사회에 미치는 영향력에 대해 생각해보기

	<p>■ 다음 차시 안내</p> <ul style="list-style-type: none"> ○ 데이터 수집 ○ 데이터 수집에도 윤리가 있다.
--	--

< 3차시 지도안 : 2단계 데이터 수집 >

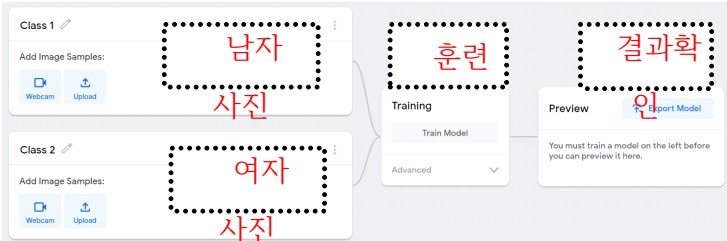
학습 목표	<ul style="list-style-type: none"> ○ 현대사회에서 데이터의 의미를 이해 할 수 있다. ○ ‘우리 동네 Hero 1탄, 안전한 공중화장실’ 프로젝트에 맞는 데이터를 수집할 수 있다.
윤리 항목	<ul style="list-style-type: none"> ○ 윤리 항목 : 개인정보보호, 투명성

단계	교수.학습 활동	자료 및 유의점						
도입	<p>■ 빅데이터를 활용한 사례 알아보기</p> <ul style="list-style-type: none"> ○ 제주 버스정보, 제주관광지 관련 사례 소개 ○ 만약 잘못된 데이터로 만들어진 정보가 제공된다면? 	<p>※ 영상 시청 후 전체 발표 및 팀별 토론으로 진행</p>						
전개	<p>■ 프로젝트 구현하기</p> <ul style="list-style-type: none"> ○ 상세 내용 <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 15%;">입력</th> <th style="width: 35%;">처리</th> <th style="width: 50%;">출력</th> </tr> </thead> <tbody> <tr> <td>화장실 출입구에서 안면 인식</td> <td>-성별 인식 : 유형 정의 -C형일 경우 : 추천 알고리즘 가동</td> <td>-A형. 올바른 입장 : 출력 없음 -B형. 입장 대기 시간이 길 경우 : 남녀 안내 -C형. 잘못 입장할A 경우 : 추천 알고리즘에 의해 이미지나 사운드 출력</td> </tr> </tbody> </table> <p>■ 데이터 수집</p> <ul style="list-style-type: none"> ○ 화장실 출입구에서 남자, 여자를 판별할 수 있는 인공지능 서비스를 만들기 위해서 훈련 데이터 필요 ○ 저작권 보호, 개인정보보호 등 정보검색 유의사항 안내 ○ 남자, 여자 사진 수집하기 <ul style="list-style-type: none"> - 공개된 데이터 제공 사이트에서 수집 - 인물과 관련한 그래픽, 이미지 등 다양한 자료 수집 ○ 기분을 좋게 하는 자료 수집하기 <ul style="list-style-type: none"> - 공개된 데이터 제공 사이트에서 수집 - 이미지, 사운드, 비디오 자료 수집하기 - 사운드, 비디오는 10초미만의 짧은 자료 수집하기 <p>■ 데이터 정리</p>	입력	처리	출력	화장실 출입구에서 안면 인식	-성별 인식 : 유형 정의 -C형일 경우 : 추천 알고리즘 가동	-A형. 올바른 입장 : 출력 없음 -B형. 입장 대기 시간이 길 경우 : 남녀 안내 -C형. 잘못 입장할A 경우 : 추천 알고리즘에 의해 이미지나 사운드 출력	<p>※ 저작권이나 개인정보보호에 관한 정보윤리교육은 선행되었다는 전제 아래 간단하게 정리</p> <p>※ 인증, 연령대 구분 없이 수집</p> <p>※ 추가로 제공할 데이터 준비 (k-ict 빅데이터센)</p>
입력	처리	출력						
화장실 출입구에서 안면 인식	-성별 인식 : 유형 정의 -C형일 경우 : 추천 알고리즘 가동	-A형. 올바른 입장 : 출력 없음 -B형. 입장 대기 시간이 길 경우 : 남녀 안내 -C형. 잘못 입장할A 경우 : 추천 알고리즘에 의해 이미지나 사운드 출력						

	○ 전체 데이터의 기준에 따라 크기 정리	터)																																																																			
	<table border="1"> <tr> <th rowspan="2">구분</th> <th colspan="3">10대 미만</th> <th colspan="3">20~30대</th> <th colspan="3">40~50대</th> <th colspan="3">60대 이상</th> <th rowspan="2">합</th> </tr> <tr> <th>황</th> <th>백</th> <th>흑</th> <th>황</th> <th>백</th> <th>흑</th> <th>황</th> <th>백</th> <th>흑</th> <th>황</th> <th>백</th> <th>흑</th> </tr> <tr> <td>남자</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr> <td>여자</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr> <td>합</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> </table>	구분	10대 미만			20~30대			40~50대			60대 이상			합	황	백	흑	황	백	흑	황	백	흑	황	백	흑	남자														여자														합													
구분	10대 미만			20~30대			40~50대			60대 이상			합																																																								
	황	백	흑	황	백	흑	황	백	흑	황	백	흑																																																									
남자																																																																					
여자																																																																					
합																																																																					
	<p>■ 실천 윤리 확인</p> <table border="1"> <thead> <tr> <th>중요도 낮음 ← → 높음</th> <th>인공지능 윤리 핵심 가치&윤리항목</th> <th>실천 정도 낮음 ← → 높음</th> </tr> </thead> <tbody> <tr> <td>1 2 3 4 5 □ □ □ □ □</td> <td>타인의 개인정보가 포함되지 않고, 초상권을 침해하지 않은 데이터를 수집 하였는가?</td> <td>1 2 3 4 5 □ □ □ □ □</td> </tr> <tr> <td>1 2 3 4 5 □ □ □ □ □</td> <td>저작권의 사용표기와 출처를 정확히 하였는가?</td> <td>1 2 3 4 5 □ □ □ □ □</td> </tr> <tr> <td>1 2 3 4 5 □ □ □ □ □</td> <td>원하는 모델을 생성하기 위해 저작권이 부여된 데이터를 저작권 허락 없이 사용하려는 욕망이 발생하였는가?</td> <td>1 2 3 4 5 □ □ □ □ □</td> </tr> <tr> <td colspan="3">수집된 데이터가 객관적이고 신뢰성이 있는지에 대한 검증방법으로 아래 예시에서 제안하고 싶은 것이 있는가? (예를 들어, 남녀를 구분하는 방법으로 키나 헤어스타일 같은 신체 특징 이외의 방법)</td> </tr> </tbody> </table>	중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음	1 2 3 4 5 □ □ □ □ □	타인의 개인정보가 포함되지 않고, 초상권을 침해하지 않은 데이터를 수집 하였는가?	1 2 3 4 5 □ □ □ □ □	1 2 3 4 5 □ □ □ □ □	저작권의 사용표기와 출처를 정확히 하였는가?	1 2 3 4 5 □ □ □ □ □	1 2 3 4 5 □ □ □ □ □	원하는 모델을 생성하기 위해 저작권이 부여된 데이터를 저작권 허락 없이 사용하려는 욕망이 발생하였는가?	1 2 3 4 5 □ □ □ □ □	수집된 데이터가 객관적이고 신뢰성이 있는지에 대한 검증방법으로 아래 예시에서 제안하고 싶은 것이 있는가? (예를 들어, 남녀를 구분하는 방법으로 키나 헤어스타일 같은 신체 특징 이외의 방법)																																																							
	중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음																																																																		
1 2 3 4 5 □ □ □ □ □	타인의 개인정보가 포함되지 않고, 초상권을 침해하지 않은 데이터를 수집 하였는가?	1 2 3 4 5 □ □ □ □ □																																																																			
1 2 3 4 5 □ □ □ □ □	저작권의 사용표기와 출처를 정확히 하였는가?	1 2 3 4 5 □ □ □ □ □																																																																			
1 2 3 4 5 □ □ □ □ □	원하는 모델을 생성하기 위해 저작권이 부여된 데이터를 저작권 허락 없이 사용하려는 욕망이 발생하였는가?	1 2 3 4 5 □ □ □ □ □																																																																			
수집된 데이터가 객관적이고 신뢰성이 있는지에 대한 검증방법으로 아래 예시에서 제안하고 싶은 것이 있는가? (예를 들어, 남녀를 구분하는 방법으로 키나 헤어스타일 같은 신체 특징 이외의 방법)																																																																					
정리	<p>■ 바꾸어 생각해 보기</p> <p>○ 내가 힘들여 만든 그림이나 음악 파일이 나도 모르게 사용되고 있다면 어떤 기분이 들까요?</p> <p>■ 다음 차시 안내</p> <p>○ 데이터의 변신 ○ 데이터 편향성</p>																																																																				

< 4-5차시 지도안 : 3단계 데이터 수정 >

학습 목표	<ul style="list-style-type: none"> ○ 수집한 데이터를 훈련에 적합한 형태로 변형할 수 있다. ○ 완성된 데이터로 이미지 인식 알고리즘으로 훈련시킬 수 있다. ○ 데이터에 따른 훈련 결과를 확인하고 수정할 수 있다.
핵심 내용	<ul style="list-style-type: none"> ○ 윤리 항목 : 데이터 편향성 배제, 다양성, 투명성

단계	교수.학습 활동	자료 및 유의점																																																																				
도입	<p>■ 인공지능 알고리즘 실패 사례 알아보기</p> <ul style="list-style-type: none"> ○ 구글의 포토서비스에서 흑인 여성을 고릴라로 인식 ○ MS사의 테이는 '히틀러가 옳으며 유대인이 싫다'라는 편견성 발언 ○ 두 사례 모두 훈련 데이터의 왜곡에서 오는 잘못된 서비스로 훈련데이터의 중요성 안내 	<p>※ 영상 시청 후 전체 발표 및 팀별 토론으로 진행</p>																																																																				
전개	<p>■ 1단계 : 수집된 자료의 수정 없이 훈련시키기</p> <ul style="list-style-type: none"> ○ 원본 자료를 그대로 훈련에 적용 <ul style="list-style-type: none"> - 사진의 크기, 수량과 관계없이 훈련 ○ 훈련 완료 후 문제점 이야기하기  <p>■ 2단계 : 수집된 자료의 정제 후 훈련시키기</p> <ul style="list-style-type: none"> ○ 얼굴을 중심으로 같은 사이즈로 조정 ○ 지난 시간에 성별, 연령, 인종에 따라 분류한 표를 확인하고 훈련 <ul style="list-style-type: none"> - 분류표 36칸의 사진의 개수가 상이해도 그래도 진행 ○ 훈련 완료 후 문제점 이야기하기 <ul style="list-style-type: none"> - 분류표의 숫자가 어떤 영향을 미치는지 확인하기 <p>■ 3단계 : 수집된 자료의 데이터 편향성 제거 후 훈련시키기</p> <ul style="list-style-type: none"> ○ 가장 많은 사진 개수를 기준으로 분류표 36칸에 동일한 숫자로 훈련 데이터 수정하기 <table border="1" data-bbox="424 1429 1145 1675"> <thead> <tr> <th rowspan="2">구분</th> <th colspan="3">10대 미만</th> <th colspan="3">20~30대</th> <th colspan="3">40~50대</th> <th colspan="3">60대 이상</th> <th rowspan="2">합</th> </tr> <tr> <th>황</th> <th>백</th> <th>흑</th> <th>황</th> <th>백</th> <th>흑</th> <th>황</th> <th>백</th> <th>흑</th> <th>황</th> <th>백</th> <th>흑</th> </tr> </thead> <tbody> <tr> <td>남자</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>120</td> </tr> <tr> <td>여자</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>10</td> <td>120</td> </tr> <tr> <td>합</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>20</td> <td>240</td> </tr> </tbody> </table> <ul style="list-style-type: none"> ○ 수정된 데이터로 훈련하기 <ul style="list-style-type: none"> - 데이터의 양에 따라 한 장씩 인식시키거나 사진 파일 전체를 업로드 할 수도 있다. ○ 1,2단계와 비교하여 차이점 이야기하기 ○ 데이터 편향성 이외에도 결과를 좌우하는 요소는 어떤 것이 있는지 이야기하기 	구분	10대 미만			20~30대			40~50대			60대 이상			합	황	백	흑	황	백	흑	황	백	흑	황	백	흑	남자	10	10	10	10	10	10	10	10	10	10	10	10	120	여자	10	10	10	10	10	10	10	10	10	10	10	10	120	합	20	20	20	20	20	20	20	20	20	20	20	20	240	<p>※ Teachable Machine을 이용해서 남녀 사진 분류하여 훈련시키고 본인의 얼굴이나 친구의 얼굴을 보고 성별을 구분하는지 확인</p> <p>※ pathmind.com/kr 사이트에서 정제된 데이터셋 제공</p>
구분	10대 미만			20~30대			40~50대			60대 이상			합																																																									
	황	백	흑	황	백	흑	황	백	흑	황	백	흑																																																										
남자	10	10	10	10	10	10	10	10	10	10	10	10	120																																																									
여자	10	10	10	10	10	10	10	10	10	10	10	10	120																																																									
합	20	20	20	20	20	20	20	20	20	20	20	20	240																																																									

■ 실천 윤리 확인		
중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	수집된 데이터 종류는 인종이나 지역, 성별 등으로 차별하거나 의도적으로 무시하지 않고 다양하게 구성되었는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	데이터 가공 시 민감한 데이터나 개인정보를 해칠 우려가 있는 데이터는 제외하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	차별적 결과를 도출할 수 있는 데이터 쏠림 현상이 발생하지 않을 만큼 고르게 분포시켰는가? (예. 여자 사진 50장, 남자 사진 100장과 같은 경우 여자 사진을 제대로 인식할 확률이 낮아 질 수 있음)	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	데이터를 수집하면서 개인적인 취향이나 선호도에 치우치지 않았는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	훈련된 결과에 대해 내가 원하는 결과가 나오지 않아도 결과를 조작 없이 정직하게 공개하고 훈련 데이터를 수정하여 사용하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
내가 예측하는 결과를 유도하기 위해 '편향'되거나 '왜곡'시키는 편집 행위가 없었나요?		
정리		
■ 바꾸어 생각해 보기 ◦ 사람의 편식과 인공지능의 편식이 사회에 미치는 영향에 대해 이야기 해보기		
■ 다음 차시 안내 ◦ 준비는 여기까지, 이제는 '우리 동네 Hero 1탄, 안전한 공중화장실' 프로젝트 만들기		

< 6-7차시 지도안 : 4단계 모델 훈련(알고리즘) >

학습 목표	<ul style="list-style-type: none"> ◦ 기획 의도대로 프로젝트를 완성할 수 있다. ◦ 인공지능 플랫폼을 활용하여 훈련할 수 있다. ◦ 테스트 결과를 분석하여 수정할 수 있다.
윤리 항목	<ul style="list-style-type: none"> ◦ 윤리 항목 : 알고리즘 투명성, 개인정보보호

단계	교수·학습 활동	자료 및 유의점
도입	■ 개발 도구 알아보기	

전개

- <https://machinelearningforkids.co.uk/#!/welcome>
- 스크래치3

■ 1단계 : 프로젝트 이름과 인식방법 선택

모든 학생을 위한 프로젝트입니까?

프로젝트 이름:
My, Hero!

인식방법:
이미지

만들기 취소

- 프로젝트 이름은 영어로 작성
- 인식방법은 남녀 성별을 구분하는 프로젝트이므로 '이미지'

My, Hero!

인식 이미지

- 프로젝트 이름을 선택하면 훈련/평가/만들기로 이동

"My, Hero!"

<p>훈련</p> <p>컴퓨터가 훈련할 수 있도록 다양한 데이터를 준비하세요.</p> <p>훈련</p>	<p>학습 & 평가</p> <p>데이터를 사용하여 컴퓨터를 학습시키세요. images</p> <p>학습 & 평가</p>	<p>만들기</p> <p>스크래치나 파이썬을 사용하여 여러분이 만든 마신러닝 모델로 게임이나 프로그램을 만들어보세요.</p> <p>만들기</p>
--	---	---

■ 2단계 훈련 : 데이터 준비

- 분류 레이블 설정 : woman, man 2개 준비
- 각 레이블에 사진 이미지 저장
 - 웹캠, 웹에서 링크, 직접 그리기 가능

인식 images as womam or man

프로젝트로 돌아가기

새로운 레이블 추가

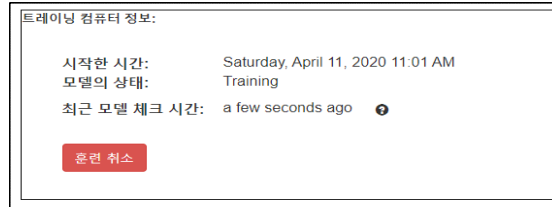
<p>womam</p> <p>사진을 드래그하여 가져오세요.</p> <p>업로드 설명 그리기</p>	<p>man</p> <p>사진을 드래그하여 가져오세요.</p> <p>업로드 설명 그리기</p>
<p>womam</p> <p>10</p>	<p>man</p> <p>10</p>

※

※ pathmind.com /kr 사이트에서 정제된 데이터셋 제공

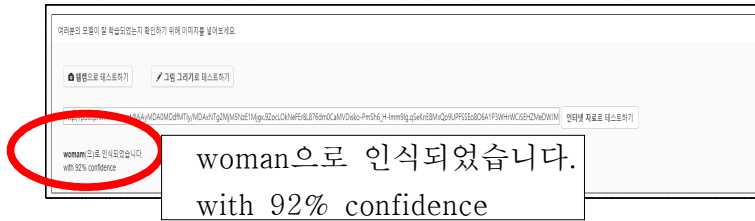
■ 3단계 훈련 시작

- IBM 클라우드가 제공하는 이미지 인식 API 준비
 - 자체 라이브러리를 통해 훈련 진행



■ 테스트 데이터로 분류기 테스트하기

- 테스트용 데이터 준비
 - 웹캠, 이미지, 직접 그린 그림 등
 - 기존에 훈련 데이터를 사용하면 안 됨



- 테스트 테이블을 만들어 정확도 확인하기

테스트 이미지	인식 레이블	정확도	실제 레이블
이미지1	woman	92%	woman
이미지2	woman	82%	woman
이미지3	;	;	;

- 테스트 결과 정리하기

1. woman 인식을 더 잘한다.	2. woman과 man 둘 다 인식을 잘한다.	3. man 인식을 더 잘한다.
---------------------	----------------------------	-------------------

- 팀별로 3가지 어느 쪽에 더 가까운지 발표한다.
- 만약 1이나 3의 결과가 나온다면 왜 그런지 이야기해 보자.

■ 테스트 결과를 반영하여 최종 데이터로 재훈련

- 위에서 분석된 테스트 결과를 반영하여 최종 데이터를 준비
- 테스트 결과 1,3번의 경우 문제점을 반영한 다양한 데이터 준비
- 최종 데이터는 기존에 훈련 데이터나 테스트 데이터로 사용했던 데이터와 다른 데이터를 추가로 사용해야 한다.

※ IBM Cloud 제공 이미지 인식 Visual recognition 텍스트 인식 Watson assistant

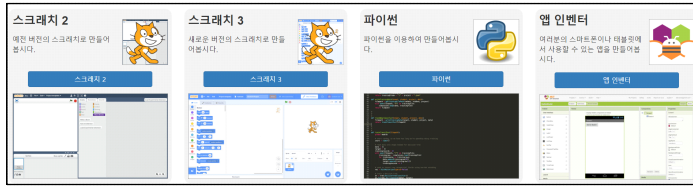
※ 훈련시간이 오래 걸리므로 시간 배분에 유의

	<p>○ 재훈련 후 위에 테스트를 반복 진행한다.</p> <p>■ 실천 윤리 확인</p> <table border="1"> <thead> <tr> <th>중요도 낮음 ← → 높음</th> <th>인공지능 윤리 핵심 가치&윤리항목</th> <th>실천 정도 낮음 ← → 높음</th> </tr> </thead> <tbody> <tr> <td>1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></td> <td>체계적인 정보 공유 등 정직한 프로세스에 맞추어 모델링 하였는가?</td> <td>1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></td> </tr> <tr> <td>1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></td> <td>오픈소스 알고리즘의 의도를 이해하고 용도에 맞게 사용하였는가?</td> <td>1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></td> </tr> <tr> <td colspan="3"> <p>내가 개발해보고 싶은 인공지능 관련 알고리즘이 있나요? (정답은 없습니다. 여러분의 생각을 적어주세요)</p> </td> </tr> </tbody> </table>	중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	체계적인 정보 공유 등 정직한 프로세스에 맞추어 모델링 하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	오픈소스 알고리즘의 의도를 이해하고 용도에 맞게 사용하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<p>내가 개발해보고 싶은 인공지능 관련 알고리즘이 있나요? (정답은 없습니다. 여러분의 생각을 적어주세요)</p>		
중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음											
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	체계적인 정보 공유 등 정직한 프로세스에 맞추어 모델링 하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>											
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	오픈소스 알고리즘의 의도를 이해하고 용도에 맞게 사용하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>											
<p>내가 개발해보고 싶은 인공지능 관련 알고리즘이 있나요? (정답은 없습니다. 여러분의 생각을 적어주세요)</p>													
정리	<p>■ 바꾸어 생각해 보기</p> <p>○ 내가 만든 라이브러리를 누군가 자기 것처럼 사용한다면 어떤 기분이 들까요? 그리고 그때는 어떻게 대처해야 할까요?</p> <p>■ 다음 차시 안내</p> <p>○ 완성된 이미지 분류기를 이용해서 최종 프로젝트 만들기 ○ 스크래치로 구현</p>												

< 8차시 지도안 : 모델 튜닝(디버깅) >

학습 목표	<p>○ 완성된 이미지 분류기를 활용하여 인공지능 서비스를 만들 수 있다.</p> <p>○ 다른 팀의 프로젝트를 보고 비판할 수 있다.</p>
윤리 항목	<p>○ 윤리 항목 : 투명성, 책임성·공공성</p>

단계	교수.학습 활동	자료 및 유의점
도입	<p>■ 자아 성찰과 비판적 사고의 필요성 인식</p> <p>○ 성찰이나 비판적 사고에 대한 부정적 인식을 깨고 발전을 위해서는 꼭 필요한 절차라는 것을 알게 한다.</p>	
	<p>■ 완성한 이미지 분류기 사용하기</p> <p>○ 스크래치, 파이선, 앱 인벤터로 확장 가능 - 여기서는 '스크래치'으로 만들기</p>	<p>※ 확장프로그램은 참여 학생 수준에 따라 결정</p>



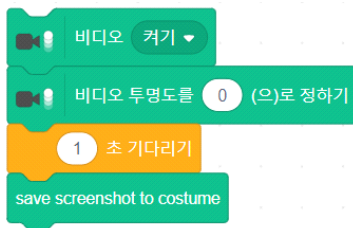
- 스크래치3으로 연결한 후 내가 만든 이미지 분류기 추가
 - 이미지 분류기 프로젝트명 : My, Hero!
 - 훈련데이터 관련 블록 생성



■ 이미지 분류기를 이용해서 ‘안전한 공중화장실’ 만들기

- 비디오를 통해 얼굴 인식하고 저장

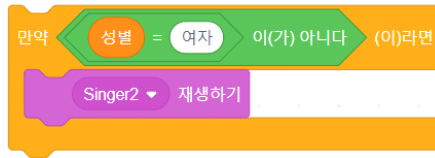
입력	처리	출력
화장실 출입구 : 이미지분류기 작동	-입력된 남/여 값에 따라 유형 정의 -C형일 경우 : 추천 알고리즘 적용 -추천알고리즘은 스크래치로 프로그래밍	-A형. 올바른 입장 : 출력 없음 -B형. 입장 대기 시간이 길 경우 : 남녀 안내 -C형. 잘못 입장할A 경우 : 추천 알고리즘에 의해 이미지나 사운드 출력



- 저장된 이미지를 이미지 분류기에 확인
- 이미지 분류기 기준에 따라 남자 또는 여자로 저장



○출입구와 성별이 다를 때 음악 재생하기



■ 팀 작업 후 최종 프로젝트 시연 및 발표

- 발표 팀의 시연 및 성찰
- 발표와 경청을 하면서 비판적 사고 연습하기

■ 실천 윤리 확인

중요도 낮음 ← → 높음	인공지능 윤리 핵심 가치&윤리항목	실천 정도 낮음 ← → 높음
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	팀 내에서 사용자, 개발자, 공급자 등 역할을 나누고 완성된 서비스에서 예상되는 부작용 및 해결 방안에 관해서 토론하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	완성된 서비스에 대해서 의도치 않게 발생하는 역기능에 대해 책임감을 가지고 공개하고 수정할 의사가 있다고 생각하는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	본 서비스 관련자들과 해당 서비스의 기능, 역할, 부작용 등에 대해서 책임을 가지고 의견을 수렴할 통로를 마련하였는가?	1 2 3 4 5 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
'우리 동네 Hero 2탄!, 0000000'을 만든다면 어떤 프로젝트를 기획하고 싶은가? (사회에 미칠 영향력과, 누구를 위한 프로젝트인가를 중심으로 작성)		

정리

■ 인공지능 기술과 인공지능 윤리의 관계 정리

- 기술의 발전은 편리함과 함께 부작용도 함께 존재
- 부작용을 최소화하는 방법은 윤리의식이 함께 발전해야 한다.

부록3. 문항별 전문가 설문 내용 및 결과

No.	문항번호	질문 내용	CVR	수렴 도	합의 도	안정 도
1	Q1.1_1	윤리항목_다양성(편향성 배제)이 적절한가?	1.00	0.00	1.00	0.08
2	Q1.1_2	윤리항목_투명성이 적절한가?	1.00	0.50	0.80	0.11
3	Q1.1_3	윤리항목_책임·공공성이 적절한가?	1.00	0.50	0.80	0.11
4	Q1.1_4	윤리항목_개인정보보호가 적절한가?	1.00	0.00	1.00	0.06
5	Q1.2_1	윤리항목 정의_다양성(편향성 배제)이 적절한가?	1.00	0.38	0.85	0.10
6	Q1.2_2	윤리항목 정의_투명성이 적절한가?	1.00	0.50	0.78	0.11
7	Q1.2_3	윤리항목 정의_책임·공공성이 적절한가?	1.00	0.50	0.78	0.11
8	Q1.2_4	윤리항목 정의_개인정보보호가 적절한가?	1.00	0.00	1.00	0.06
9	Q2_1	프로세스별 인공지능 윤리 적용 내용_1단계. 문제정의 & 주제선정 (프로젝트 주제 결정)	1.00	0.00	1.00	0.08
10	Q2_2	프로세스별 인공지능 윤리 적용 내용_2단계. 데이터 수집 (데이터 요건 정의 및 데이터 수집)	1.00	0.38	0.85	0.10
11	Q2_3	프로세스별 인공지능 윤리 적용 내용_3단계. 데이터 준비 (데이터를 훈련에 적합한 형태로 변환)	1.00	0.00	1.00	0.08
12	Q2_4	프로세스별 인공지능 윤리 적용 내용_4단계. 모델 훈련, 알고리즘 (주제에 맞는 최적의 알고리즘 선택)	1.00	0.00	1.00	0.06
13	Q2_5	프로세스별 인공지능 윤리 적용 내용_5단계. 모델 튜닝, 디버깅 (인공지능 서비스 적용, 결과 분석 및 공유)	1.00	0.00	1.00	0.06
프로세스별 실천윤리 확인 문항(프로세스 중심으로)						
14	Q3.1_1	문제정의 및 주제선정_내가 기획한 프로젝트는 완성 후 사회에 어떠한 영향력을 끼칠 것인가에 대해 고려하고 기획하였는가?	1.00	0.00	1.00	0.06
15	Q3.1_2	문제정의 및 주제선정_인공지능 기술이나 서비스를 개발하면서 이와 관련된 사람들을 충분히 고려하였는가?	0.80	0.50	0.78	0.15
16	Q3.1_3	문제정의 및 주제선정_기획한 프로젝트는 공공의 이익에 도움이 되는가?	1.00	0.38	0.85	0.10
17	Q3.1_4	문제정의 및 주제선정_인공지능 기술이나 서비스를 기획하는 과정에서 안전이나 가격 등 여러 가지 고려해야 할 사항 중 자신이 생각하는 최우선 순위는?	0.60	0.50	0.80	0.18
18	Q3.2_1	데이터 수집_저작권의 사용표기를 정확히 하였는가?	0.80	0.38	0.85	0.14
19	Q3.2_2	데이터 수집_타인의 개인정보가 포함되지 않은 데이터와 초상권을 침해하지 않은 데이터를 수집 하였는가?	1.00	0.50	1.00	0.08
20	Q3.2_3	데이터 수집_타인의 개인정보가 포함되지 않은 데이터와 초상권을 침해하고 싶은 유혹은 없었나요?	0.60	0.50	0.80	0.18
21	Q3.2_4	데이터 수집_수집된 데이터가 객관적이고 신뢰성이 있는지에 대한 검증방법으로 아래 예시에서 제안하고 싶은 것이 있는가? (예를 들어, 남녀를 구분하는 방법으로 키나 헤어스타일 같은 신체 특징 이외의 방법)	0.60	0.38	0.85	0.23
22	Q3.3_1	데이터 수정_수집된 데이터 종류는 인종이나 지역 등으로 차별하거나 의도적으로 무시하지 않고 다양하게 구성되었는가?	0.80	0.38	0.85	0.14
23	Q3.3_2	데이터 수정_데이터 가공 시 민감한 데이터나 개인정보를 해칠 우려가 있는 데이터는 제외하였는가?	1.00	0.50	0.80	9.11
24	Q3.3_3	데이터 수정_데이터 쓸림 현상이 발생하지 않을 만큼 고르게 분시켰는가? (예, 여자 사진 50장, 남자 사진 100장과 같은)	1.00	0.50	0.80	0.11

25	Q3.3_4	데이터 수정_데이터를 수집하면서 개인적인 취향이나 선호도에 치우치지 않았는가?	0.80	0.38	0.85	0.14
26	Q3.3_5	데이터 수정_훈련된 결과에 대해 내가 원하는 결과가 나오지 않아도 결과를 조작 없이 정직하게 공개하고 훈련 데이터를 수정하여 사용 하였는가?	0.80	0.50	0.80	0.15
27	Q3.3_6	데이터 수정_내가 예측하는 결과를 유도하기 위해 '편향'되거나 '왜곡'시키는 편집 행위가 없었나요?	1.00	0.38	0.85	0.10
28	Q3.4_1	모델 훈련(알고리즘)_체계적 프로세스에 맞추어 모델링 하였는가?	0.60	0.50	0.78	0.18
29	Q3.4_2	모델 훈련(알고리즘)_오픈소스 알고리즘의 의도를 이해하고 용도에 맞게 사용하였는가?	1.00	0.38	0.85	0.10
30	Q3.4_3	모델 훈련(알고리즘)_내가 개발해보고 싶은 인공지능 관련 알고리즘이 있나요? (정답은 없습니다, 여러분의 생각을 적어주세요.)	1.00	0.00	1.00	0.08
31	Q3.5_1	모델 튜닝(디버깅)_팀 내에서 사용자, 개발자, 공급자 등 역할을 나누고 완성된 서비스에서 예상되는 부작용 및 해결 방안에 관해서 토론하였는가?	1.00	0.00	1.00	0.08
32	Q3.5_2	모델 튜닝(디버깅)_완성된 서비스에 대해서 의도치 않게 발생하는 역기능에 대해 책임감을 가지고 공개하고 수정할 의무가 있다고 생각하는가?	0.80	0.38	0.85	0.14
33	Q3.5_3	모델 튜닝(디버깅)_본 서비스 관련자들과 해당 서비스의 기능, 역할, 부작용 등에 대해서 차별 없이 의견을 수렴할 통로를 마련하였는가?	1.00	0.50	0.80	0.11
34	Q3.5_4	모델 튜닝(디버깅)_'우리 동네 Hero 2탄, 000000'을 만든다면 어떤 프로젝트를 기획하고 싶은가? (사회에 미칠 영향력과 누구를 위한 프로젝트인가를 중심으로 작성)	1.00	0.38	0.85	0.10
사전 · 사후 진단 문항(윤리 항목 중심으로)						
35	Q4.1_1	인공지능 윤리 핵심 가치_인공지능 기술이나 서비스는 기본적으로 인간 삶의 질을 향상 시키는데 기여해야 한다고 생각하는가?	0.80	0.38	0.85	0.14
36	Q4.1_2	인공지능 윤리 핵심 가치_인공지능 기술이나 서비스는 긍정적 혹은 부정적이든 사회 변화에 영향을 끼칠 수 있다고 생각하는가?	1.00	0.50	0.80	0.11
37	Q4.1_3	인공지능 윤리 핵심 가치_인공지능 기술이나 서비스는 기본적으로 인간의 자유와 존엄성을 존중해야 하는가?	1.00	0.50	0.80	0.11
38	Q4.2_1	다양성 & 편향성 배제_개발의 전 과정에서 성, 인종, 종교, 지역, 민족 등의 차이로 인한 근본적인 차별을 배제해야 하는가?	1.00	0.50	0.80	0.11
39	Q4.2_2	다양성 & 편향성 배제_인공지능 기술이나 서비스를 이용함에 있어 사회적 약자 및 취약 계층 등을 고려한 별도의 서비스가 필요한가?	1.00	0.50	0.78	0.11
40	Q4.2_3	다양성 & 편향성 배제_민감 데이터 또는 개인 데이터의 이용을 최소화하거나 사용하지 않고 훈련할 수 있는 방법을 고려했는가?	0.60	0.50	0.78	0.18
41	Q4.3_1	알고리즘 투명성_오픈 알고리즘에 개발자(나)만 알고 있는 숨은 기능을 넣어도 괜찮은가?	0.80	0.50	0.80	0.28
42	Q4.3_2	알고리즘 투명성_최종적인 의사결정은 인공지능 시스템이 아니라 인간이 할 수 있어야 하는가?	0.80	0.38	0.85	0.14
43	Q4.3_3	알고리즘 투명성_훈련에 사용된 데이터의 양은 결과 도출에 필요한 만큼 적절한가?	1.00	0.50	0.80	0.11
44	Q4.3_4	알고리즘 투명성_알고리즘의 성능을 위해 의도적인 수정을 하였는가?	0.80	0.50	0.80	0.28
45	Q4.4_1	책임성 & 공공성_최초 개발부터 이용에 이르기까지 전 과정과 결과에 대한 책임을 공유해야 할 책임이 있는가?	0.80	0.38	0.85	0.14
46	Q4.4_2	책임성 & 공공성_위급상황 발생 시 해결을 위해서 필요 데이터를 공유해야 할	0.60	0.50	0.80	0.18

		책임이 있는가?				
47	Q4.4_3	책임성 & 공공성_인공지능 기술이나 서비스에 사용된 데이터를 공유해야 할 책 임이 있는가?	1.00	0.38	0.85	0.10
48	Q4.5_1	개인정보보호_다수의 사람들에게 이로운 서비스를 위해서 소수의 저작권, 초상권, 개인정보는 희생되어도 괜찮은가?	0.80	0.38	0.85	0.27
49	Q4.5_2	개인정보보호_나와 관련한 데이터의 사용에 대해 알고 있어야 하고 원치 않을 경 우 사용을 거부할 수 있어야 하는가?	1.00	0.00	1.00	0.08
50	Q4.5_3	개인정보보호_암호화, 익명화 등 개인정보보호 강화를 위한 조치는 필요한가?	1.00	0.00	1.00	0.08