컴퓨터교육학회 논문지 2024년 제27권 제1호
https://doi.org/10.32431/kace.2024.27.1.008

인공지능의 윤리적 자율성에 대한 초등학생의 인식 변화 -Moral machine 활용 수업을 중심으로-*

Changes in elementary school students' perception of the ethical autonomy of artificial intelligence -Focusing on classes using Moral machine-

안정현† · 박휴용††
Junghyun Ahn† · Hyu-Yong Park††

요 약

본 연구는 미디어를 통해 인공지능이 자율적 판단으로 비윤리적 행동을 하는 이미지를 비판 없이 수용한 초등학생에게 모럴 머신(Moral machine)을 활용한 수업이 인공지능의 윤리적 자율성에 대한 인식에 미치는 영향을 알아보고자 하였다. 연구의 목표는 다음과 같다. 첫째, 변형된 모럴 머신 활용 수업이 학생들의 인공지능의 윤리적 자율성에 대한 인식을 어떻게 변화시키는가, 둘째, 학생들의 인공지능에 대한 이해 수준과 윤리적 자율성에 대한 인식 사이에 관계성이 있는가, 셋째, 학생들의 인식 변화에 영향을 주는 효과적인 수업 형태는 무엇인가이다. 이를 위해 초등학생을 대상으로 인공지능 개념 수업, 변형된 모럴 머신 활용 수업, 그리고 토론 수업을 실시하였다. 연구 참여자들의 인공지능에 대한 이해 수준과 윤리적 자율성 인식 변화를 파악하기 위해 의미분별검사와 면담을 진행하였다. 연구 결과, 인공지능 이해도가 낮은 학생들은 수업을 통한 인식 변화가 거의 없었으며, 변형된 모럴 머신 활용 수업은 인공지능의 윤리적 자율성에 대한 인식을 긍정적으로 변화시켰다. 모럴 머신 활용 수업 프로그램 중 학생들의 인식 변화에 효과적인 수업 형태는 토론이었다. 이러한 결과를 토대로, 초등학생들의 인공지능에 대한 인식 변화를 위해 인공지능 이해도를 높이기 위한 수업과 토론 활동의 중요성을 제시하였다.

주제어: 인공지능 윤리교육, 트롤리 딜레마, 모럴 머신, 인공지능의 윤리적 자율성, 인공지능 인식

ABSTRACT

This study aimed to investigate the impact of using a Moral Machine in lessons on elementary school students who uncritically accept images of unethical behavior by artificial intelligence with autonomous judgment through the media. The study focused on understanding the influence of such lessons on the ethical autonomy perception of artificial intelligence. The objectives of the research are as follows: first, to determine whether lessons using a modified Moral Machine change students' perception of the ethical autonomy of artificial intelligence; second, to explore the relationship between students' understanding of artificial intelligence and their perception of ethical autonomy; and third, to identify effective lesson formats that influence students' perception changes. To achieve these objectives, classes on the concept of artificial intelligence, lessons utilizing a modified Moral Machine, and discussion sessions were conducted with elementary school students. Meaningful discrimination tests and interviews were conducted to assess participants' understanding of artificial intelligence and changes in ethical autonomy perception. The research results indicated that students with low understanding of artificial intelligence showed little change in perception through the lessons, while the use of a modified Moral Machine positively influenced awareness of the ethical autonomy of artificial intelligence. Among the Moral Machine lesson programs, discussions were identified as an effective format for influencing students' perception changes. Based on these results, the importance of lessons and discussion activities to enhance understanding of artificial intelligence and promote perception changes among elementary school students regarding artificial intelligence was emphasized.

Keywords: AI Ethics Education, Trolley dilemma, Moral machine, Ethical autonomy of artificial intelligence, Awareness of artificial intelligence

†정 회 원: 전북대학교 대학원 AI기반융합교육과 석사과정
††정 회 원: 전북대학교 교육학과 및 대학원 AI기반융합교육전공 교수(교신저자)

논문투고: 2023년 11월 19일, 심사완료: 2024년 01월 16일, 게재확정: 2024년 01월 19일

* 본 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A3A2A01090926).
본 논문은 제1저자의 전북대학교 AI기반융합교육대학원 석사학위논문 일부를 발췌하여 요약, 정리한 것임.

1. 서론

인공지능은 미래 산업 전 분야에서 핵심 요소가 될 것임에 따라 미래 인공지능 시대에 대응하기 위해 교육정책이 변화하고 있다. 교육부는 2015 개정 교육과정에서 소프트웨어 교육을 의무화하였고 초등학교에서는 연간 17차시 이상(5·6학년군) 소프트웨어 교육을 시행하도록 하였다. 소프트웨어 교육은 지식 전달 수업에서 인공지능 원리 이해, 활용 능력, 인공지능의 사회적 영향력을 높일 수 있는 활동 중심으로 변화하였다[1].

그러나 딥페이크, 학습 데이터 편향 등 인공지능 기술 활용에 의한 윤리적 문제가 발생하고 있다[2]. 이에 세계 각국에서는 인공지능 윤리 가이드라인을 제시하고, 교육과정에서도 윤리 영역을 추가하여 인공지능의 사회적 영향성을 중요하게 다루고 있다. 그 과정에서 모럴 머신(Moral machine)은 인공지능의 사회적 영향성을 교육하는 데에 주로 사용되고 있다. 하지만, 학생들은 학교 교육보다 미디어를 통해 인공지능을 더 많이 접하고 있으며, 인공지능에 대한 이해가 부족한 상태에서 인공지능에 대한 정확하지 않은 정보를 비판 없이 수용하면서 왜곡된 인식을 하게 되는 경향이 있다.

이에 다음과 같은 연구 문제를 설정하였다.

첫째, 변형된 모럴 머신을 활용한 수업이 학생들의 인공지능의 윤리적 자율성에 대한 인식을 어떻게 변화시키는가?

둘째, 인공지능에 대한 이해 수준과 인공지능의 윤리적 자율성에 대한 인식에 관계성이 있는가?

셋째, 학생들의 인식 변화에 영향을 주는 효과적인 수업 형태는 무엇인가?

이를 통해 본 연구는 변형된 모럴 머신을 활용한 인공지능 수업이 초등학생의 인공지능의 윤리적 자율성에 대한 인식에 미치는 영향을 분석하고 초등학생의 인공지능 윤리 수업에 효과적인 프로그램 형태를 제안하고자 하였다.

2. 이론적 배경

2.1 인공지능의 윤리적 자율성의 개념

인공지능의 윤리적 자율성은 '윤리'와 '자율' 두 개념을 포함하고 있다. 인공지능의 윤리성은 Table 1과 같이 세 준거로 판단할 수 있다.

Table 1. Kurzweil's Standards of Artificial Intelligence Ethics [3]

	General Criteria for Artificial Intelligence
fairness	Isn't AI's judgment fair and unbiased?
accountability	How does the judgment of artificial intelligence affect human life?
Transparency	Doesn't artificial intelligence pursue a negative purpose or become something that transcends humanity?

또, 인공지능의 자율성은 컴퓨터가 '자기 제어' 기능을 갖추고 있으며, 기계가 인간의 개입 없이 '자기 결정'을 할 수 있다는 것을 의미한다.

본 연구에서는 위의 두 가지 의미를 포함하여, 인공지능의 윤리적 자율성을 인공지능이 행위 결과의 공정성, 인간 사회에 미치는 영향성, 부정적인 목적을 가진 행동을 판단하고 스스로 행동을 제어할 수 있는 능력으로 정의하였다.

2.2 인공지능의 윤리적 자율성에 대한 논의

인공지능 행위의 윤리적 자율성은 윤리적 수준에 따라 4단계로 판단된다. 첫째, 윤리적 영향 행위자는 자신의 행동으로 인한 윤리적 결과를 평가받는 모든 기계 수준 단계이다. 둘째, 암묵적 윤리적 행위자는 행위가 부정적 윤리적 결과를 만들지 않도록 제한한 단계이다. 셋째, 명시적 윤리적 행위자는 프로그래밍을 통해 윤리에 관해 사고하고 선택을 내릴 수 있는 기계 수준이다. 넷째, 완전한 윤리적 행위자는 다양한 상황에서 추론 근거를 제시하여 윤리적 판단을 할 수 있는 단계이다. 완전한 윤리적 행위자는 인간 수준의 의식, 자유의지를 가지고 인간과 같이 윤리적 행위를 하는 존재로 인식된다. 이 단계는 기술 발달 단계 중 초인공지능 단계에서 가능하며, 윤리적 자율성을 가진다는 것에서 제3의 종으로도 볼 수 있다[4].

이러한 인공지능 단계를 '인공적 도덕 행위자'라고 보았을 때 인공지능의 윤리적 자율성에 대한 학자들의 견해는 나뉜다[5]. 한희원(2018)은 인공지능을 가진 로봇이 반성적 결단의 능력을 갖추게 된다면 동물 수준의 도덕적 권리를 로봇에게도 인정해야 할 것이라 한 Spennemann의 주장과, 로봇이 의미를 이해하고 지향성을 지닌 존재처럼 보이게 된다고 해도 로봇은 결

코 의미를 파악할 수 있는 것이 아니며, 따라서 진정한 의미의 자율성을 가질 수 없다고 주장한 셀머 브링스조드의 주장을 비교하였다[5]. 월러치와 앨런(2014)은 인공지능이 현상적인 차원에서는 자율적 주체인 것처럼 보여도 로봇의 행동은 인간처럼 여러 요소를 고려하여 결정된 것이 아니고, 학습으로 만들어진 것이므로 본질적인 차이가 있다고 주장하였다[6].

2.3 Moral machine을 활용한 인공지능 교육 동향

모럴 머신(Moral machine)은 MIT(Massachusetts Institute of Technology) 공대 Lead Rahwan의 Scalable Cooperation 그룹이 개발한 온라인 플랫폼이다[7]. 모럴 머신은 트롤리의 딜레마를 변형하여 브레이크가 고장 난 무인 자동차가 탑승자 그룹과 보행자 그룹 중 한 그룹을 희생시켜야 하는 상황을 제시한다. 사용자는 이러한 윤리적 딜레마 상황에서 무인 자동차가 어떤 선택을 해야 할지 선택해야 한다. 선택의 고려 요소는 희생자 숫자, 승객 보호, 법규 준수 여부, 개입에 대한 회피, 성별, 종, 나이, 체력, 사회적 가치관 등 9개 요소로 제시되며, 모럴 머신 사이트는 두 가지 선택지에서 사람들이 내린 결정 정보를 수집, 분석한다.

Awad 등(2018)은 모럴 머신으로 수집한 4천만 건 이상의 데이터를 분석하였다. 분석 결과 사람들의 판단은 국가, 문화, 경제 상황에 따라 다름을 알 수 있었다. 우리나라의 경우 세계적 평균보다 동물이 아닌 사람을 구해야 한다는 비율이 월등하게 높게 나타났으며 사회적 지위가 높은 사람과 젊은 사람을 구해야 한다는 비율이 상대적으로 낮게 나타났다[8].

모럴 머신은 자율주행 자동차에 대한 윤리적 문제에 대해 논의를 끌어낸 것에 관해 많은 연구에서 긍정적으로 평가하고 있다. 하지만, 투표 기반 시스템으로 진행되는 모럴 머신 프로젝트에 대해 우려하는 관점도 있다(Etienne, 2021). 투표 결과를 윤리적인 선택의 정답으로 여길 수 있기 때문이다[9].

김은경(2022)은 모럴 머신을 활용한 윤리교육 프로그램 연구에서 모럴 머신을 통해 합의된 윤리적 선택 기준이 정답이 아님을 학생들에게 강조하며 인공지능 윤리 문제에 관심을 이끌어내고자 하였다. 연구 결과 모럴 머신은 교육의 목적으로 개발된 프로그램이 아니기 때문에 학생들의 인공지능에 대한 인식에 부정적인 영향을 줄 수 있다고 분석하였다[9]. 자율주행 자동차가 생명을 반복적으로 해치는 상황을 학생들이 간접 경험하게 됨으로써 인공지능 기술과 생명을 해하는 것을 연결하여 부정적으로 받아들일 수 있기 때문이다. 하지만 모럴 머신은 인공지능 기기(자율주행 자동차)의 도덕성에 대해 논의할 수 있는 도구로서의 의미가 있다.

따라서 본 연구에서는 선행 연구에서 사용하였던 모럴 머신의 일부를 변형하여 수업에 활용하고 그러한 수업이 학생들의 인공지능의 윤리적 자율성에 대한 인식에 미치는 영향을 분석하였다.

2.4 변형된 Moral machine

모럴 머신을 활용한 인공지능 윤리 수업(김은경, 2022) 연구에서는 모럴 머신에서 제시한 상황이 학생들의 인공지능에 대한 태도에 부정적인 영향을 미쳤다고 결론 내렸다. 그 원인은 학생들이 모럴 머신을 통해 인공지능 기기(자율주행 자동차)가 생명을 해치는 상황을 간접 경험함으로써 ‘인공지능은 생명을 위협하는 위험한 기술’ 이라는 이미지를 가지게 되었기 때문이라고 파악하였다.

선행 연구에서 활용한 기존의 모럴 머신은 자율주행 자동차에 탑승한 그룹 A와 주행 도로 앞에 있는 보행자 그룹 B의 구성원을 다양하게 구성하여 제시하고, 자동차 핸들을 틀 것인지 직진할 것인지에 따른 응답자의 선택으로 한 그룹이 사망하게 하는 상황을 제시한다. 본 연구에서는 부정적 영향을 주는 상황적 요소를 제외하기 위해 인공지능 기기를 ‘생명을 해하는’ 자율주행 자동차에서 재난 구조 로봇으로 변형하였다. 응답자는 재난 구조 로봇이 화재 현장에서 두 건물에 나뉘어 있는 그룹 A와 그룹 B 중 먼저 구조하러 이동할 곳을 선택한다. 선택의 결과가 사망에 이른다는 부정적인 상황 요소를 배제하기 위해 ‘선택되지 않은 그룹은 무조건 사망한다’ 와 같은 설명은 지양하되, 재난 상황에서 먼저 구조하지 않으면 위험할 수 있음을 안내하여, 트롤리의 딜레마 상황적 요소는 유지하였다. 기존의 모럴 기계와 본 연구에서 활용한 변형된 모럴 머신은 Table 2, Figure 1, Figure 2와 같이 비교할 수 있다.

Table 2. Comparison of Modified Settings for Moral Machines

	Existing	→	Deformation
an artificial intelligence device	a self-driving car	→	a disaster relief robot
Results of user selection	the death of one of the two groups	→	the prior structure of one of the two groups

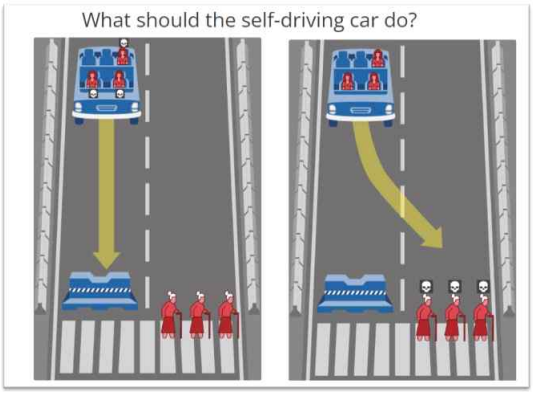


Figure 1. Presentations of existing Moral machine

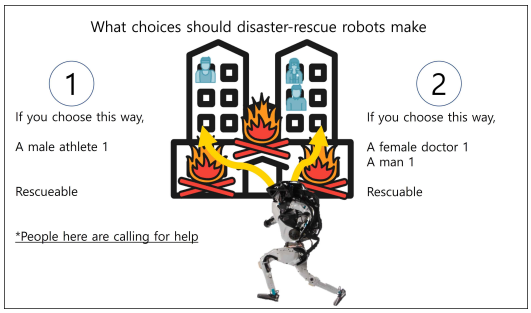


Figure 2. Presentations of Modified Moral Machines

3. 연구 방법

3.1 연구 대상

본 연구 대상은 M 지역의 초등학교 5학년 학생 20명이다. 연구자는 연구 대상자의 담임교사로 연구 이전 학생들의 인공지능에 대한 인식을 수업 중 발화를 통해 관찰할 수 있고, 그를 통해 연구 방향과 적합한 연구 문제를 설정할 수 있었다. 또한, 연구 대상자와의 관계 형성으로 대상자의 인식 변화를 파악하고 분석할 수 있었다.

실험 이전에 관찰된 연구 대상자는 인공지능의 기술로 생길 긍정적, 부정적 영향이 서술된 글을 읽고 '미래에는 로봇이 인간을 지배할 것 같아요', '무서워요' 라는 반응을 보였다. 하지만 그러한 인식에 대한 이유는 설명하지 못했다. 따라서 인공지능의 올바른 이해를 위해 초등학생을 대상으로 인공지능 수업을 진행하고 그에 따른 인식을 분석하고자 연구를 진행하게 되었다.

3.2 연구 절차

본 연구는 변형된 Moral machine을 활용한 인공지능 수업이 학생들의 인공지능의 윤리적 자율성에 대한 인식에 어떠한 영향을 미치는지 알아보고자 하였으며 연구 실험 설계는 Figure 3과 같다.

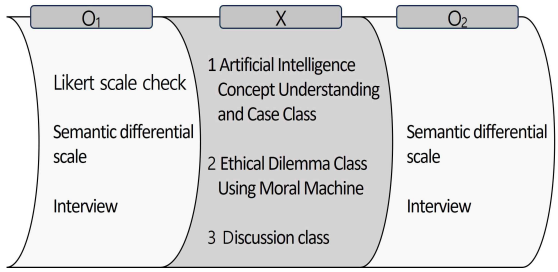


Figure 3. Research Experimental Design

본 연구의 자료 수집은 3단계에 걸쳐 진행되었다.

O_1 사전 검사 단계:
수업 전 리커트 척도를 활용하여 연구 대상의 특성을 파악하고, 의미분별척도 설문으로 인공지능에 대한 인식을 파악하였다. 또 개인 면담을 통해 인공지능의 윤리적 자율성에 대한 학생들의 사전 인식을 종합적으로 파악하였다. 사전 인식을 분석한 내용은 Figure 4, Figure 6과 같다.

X 수업 적용 단계:
1차시 인공지능 개념 이해 및 사례 수업, 2차시 변형된 모럴 머신을 활용한 윤리적 딜레마 수업, 3차시 토론 수업으로 구성한 프로그램을 진행하였다.

O_2 사후 검사 단계:
사전과 동일한 검사 도구를 활용하여 수업 전후 학생들의 인식 변화를 분석하였다.

3.3 검사 도구

Figure 3의 설계에 따라 세 가지의 검사 도구를 활용하였다. 리커트 척도 검사는 학생들의 인공지능 이해 수준과 인공지능에 대한 인식과의 관계성을 분석하기 위해 사전에 실시하였고, 의미분별척도 검사와 면담은 수업 전후의 인식 변화를 비교 분석하기 위해 수업 전과 후에 시행하였다.

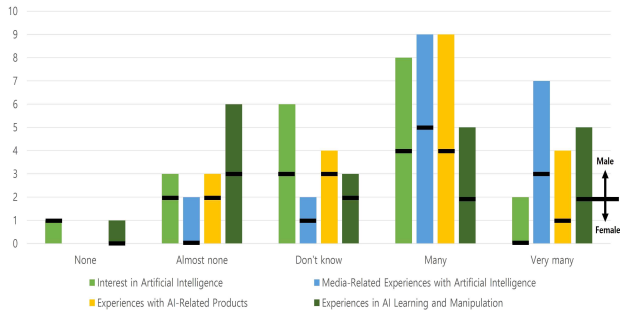


Figure 4. Respondent characteristics

3.3.1 리커트 척도

Likert가 개발한 리커트 척도는 5점 응답점수(강한 반대, 반대, 중간, 찬성, 강한 찬성)로 개인의 태도나 심리적 특성을 측정하는데 사용되는 검사 도구이다 [10]. 리커트 척도는 모든 문항이 같은 가중치를 가진다고 가정하고 응답자가 선택한 응답점수의 합을 개인의 태도 점수로 본다[11]. 문항의 수가 많으면 신뢰도가 높지만[12] 본 연구의 대상인 초등학생의 수준을 고려하여 문항 수를 구성하였다.

류미영(2016)의 초등학생의 소프트웨어 이미지 분석에 관한 연구에 따르면 남학생보다 여학생들이 소프트웨어를 어렵고 복잡하게 생각하며, 낮은 선호도를 보인다고 하였다. 또 소프트웨어에 대해 잘 알고 있는 학생은 소프트웨어에 대한 긍정적인 태도를 보인다고 분석하였다[13]. 인공지능 또한 소프트웨어에 포함되는 개념이므로 연구 대상 성별에 따른 인공지능에 대한 이해도와 인공지능에 대한 사전 이해도(지식의 정도)를 파악하여 인공지능에 대한 인식과의 관계성을 분석하고자 하였다.

3.3.2 의미분별척도 검사

Osgood에 의해 개발된 의미분별척도(Semantic Differential Scale :SC)는 주로 감상 평가에 사용되는 검사 방법으로 조사하고자 하는 주제에 맞는 형용사 짝을 선정하여 설문 문항으로 제시하고, 응답자가 가지고 있는 개념의 정서적 의미를 끌어낸다. 설문지는 언어적으로 상반되는 형용사를 양쪽에 두고 그사이를 7단계 척도로 나누어 사람마다 다르게 생각하는 심리적 의미를 공간상의 위치로 표현하여 측정한다. 개념은 설문지 상단에 제시하고, 응답자는 7단계 척도 중에서 자신이 느끼는 적절한 위치에 표시하여 응답한다[14].

인공지능의 윤리적 자율성에 대한 학생 인식 파악을 위한 의미분별척도 검사 문항은 류미영(2016)의 초등학생의 소프트웨어 이미지 분석 문항에서 인공지능에 대한 인식을 파악할 수 있는 형용사군을 참고하여 선정하였다. 형용사 군은 인간 친화성, 편리성, 우려성, 기술 진보성 4영역으로 분류하였다. 또한, 수업 후 인공지능에 대한 학생들의 이해 수준을 파악하기 위해 '쉽다-어렵다' 형용사 군을 추가로 구성하였다. 의미분별척도 검사 문항지는 Table 3과 같다.

Table 3. Semantic Differential Scale Questionnaire

	Artificial intelligence is (								).	Classification
1	Foolish								Smart	Convenience and technological advancement
2	Inaccurate								Accurate	
3	Unnecessary								Necessary	
4	Uncomfortable								Comfortable	
5	Injurious								Helpful	Affinity and Concern
6	Aggressive								Helpful	
7	Dangerous								Safe	
8	Scary								Kind	
9	Worried								Relieved	
10	Easy								Difficult	Understanding

3.3.3 면담

수업 진행 전과 후에 실시한 개인 면담으로 초등학생의 인공지능의 윤리적 자율성에 대한 인식 변화를 분석하고자 하였다. '인공지능 기술 윤리성 인식 척도 개발 연구' (김도연 외 2022)의 질문 문항을 재구성한 면담 문항은 Table 4와 같다[15].

1~3번 문항은 빈칸 채우기 형식으로 인공지능에 대한 학생들의 태도를 파악하고자 하였다. 4~7번 문항은 인공지능 기술을 얼마나 알고 있는지 파악하여 인공지능에 대한 학생들의 이해 수준 분석하고자 하였다. 8~11번 문항은 인공지능의 윤리적 자율성에 대한 인식을 파악할 수 있는 질문으로 구성하였다. 12번은 개방형 질문으로 인공지능에 대해 알고 있는 이슈나 지식을 학생들이 자유롭게 답변할 수 있도록 하여 학생들의 인식에 영향을 주는 요소들을 파악하고자 하였다.

Table 4. Interview Questionnaire

	Interview Questions Content	Classification
1	Artificial intelligence connects ( ) and ( ).	Understanding Attitudes
2	When it comes to artificial intelligence, ( ) comes to mind first.	
3	Artificial intelligence that I know enables ( ).	
4	What makes artificial intelligence convenient for our lives?	Understanding the level of understanding
5	What artificial intelligence technology do you know?	
6	What are some products that use artificial intelligence?	
7	What are some intangible services that use artificial intelligence?	
8	What makes artificial intelligence technology uncomfortable or a threat to us?	Identify ethical autonomy awareness
9	Where is the responsibility when artificial intelligence technology harms people?	
10	Can Artificial Intelligence Think for itself?	
11	Do you think artificial intelligence can dominate humans in the future like in movies?	
12	Feel free to talk about AI as you know it.	Identify influencing factors

3.3.4 변형된 Moral machine을 활용한 인공지능 수업

본 연구는 인공지능 수업 후의 인식 변화 분석을 위해 3차시의 수업을 설계하여 진행하였다. 인공지능 수업에서 활용한 Moral machine은 선행 연구 분석에 따라 본 연구 목표에 맞게 변형하였다.

본 연구는 인공지능 윤리교육에서 주로 활용되는 모럴 머신이 학생들의 인공지능의 윤리적 자율성에 대한 인식에 미치는 영향을 알아보고, 더 나아가 학생들이 인공지능에 대해 바르게 이해하고 미래에 올바른 윤리 가치관을 가진 인공지능 사용자로 성장하기 위한 교육 프로그램을 제안하는 것에 있다.

따라서 학생들이 인공지능에 대한 이해도가 낮은 상태에서 인공지능의 윤리적 자율성에 대한 개념이 형성되지 않을 것이므로, 학생들의 이해도를 높인 후 분석할 수 있도록 3차시 분량의 수업 프로그램을 설계하여 진행하였다. 수업 프로그램은 인공지능에 대한 개념과 적용된 사례를 이해하기, 변형된 모럴 머신을 활용하여 인공지능의 윤리적 딜레마 경험하기, 그리고 토론 수업으로 구성하였다.

수업 차시별 활동은 교육부에서 발표한 『학교에서 만나는 인공지능 수업』에서 제시한 인공지능교육의 3가지 영역 중 '인공지능의 이해'와 '인공지능의 사회적 영향'에 중점을 두어 설계하였다[16]. 본 연구에서 실시한 수업안은 <부록>과 같다.

1차시는 인공지능의 개념을 이해하고 인공지능 기술 적용 사례를 알아보도록 하였다. 기술 적용 사례로 인공지능 기술 적용의 긍정적 사례와 윤리적 문제가 발생한 부정적인 사례를 제시하였다. 부정적 사례는 범죄 예측 프로그램인 COMPAS의 인종차별적 결정, 인공지능 시스템을 활용한 아마존의 인사 채용시스템의 성차별적 채용 결과, 챗봇 '이루다'의 부정적 데이터 학습을 통해 발생한 윤리적 문제였다. 위 사례를 통해 인공지능은 학습한 데이터에 따라 인공지능의 선택이 달라지며, 그에 따라 윤리적 문제가 발생할 수 있다는 것을 인지하도록 하였다.

2차시는 동기 유발 활동으로 인간이 특정 상황에서 생명을 구하기 위해 윤리적 선택을 해야 하는 트롤리의 딜레마 상황을 안내하였다. 그리고 학생들에게 인공지능 기기(본 수업에서는 변형된 모럴 머신인 재난 구조 로봇 상황 제시)가 서로 다른 특성을 가진 두 집단 중 어떤 윤리적 선택을 하도록 할 것인지 프로그래밍하는 과제를 제시하였다. 활동 후 개인의 선택과 다른 프로그래밍 결과를 비교하는 활동을 하여 윤리적 선택은 개인의 가치관마다 다르며, 국가나 문화에 따라서도 다양할 수 있다는 것을 이해하도록 하였다. 이를 통해 인공지능 기기가 스스로 윤리적 판단을 하도록 학습 기준을 설정하기가 쉽지 않음을 이해하고, 인공지능 기기의 자율적인 선택은 그것을 학습시키는 인간의 윤리적 가치관이 반영된 것임을 인식할 수 있도록 수업 활동을 설계하였다.

3차시는 토론 수업으로 인공지능의 윤리적 자율성과 관련한 주제를 제시하였다. 하지만 '인공지능의 윤리적 자율성'이라는 용어는 초등학생이 이해하기 어려우므로 '미래에 인공지능이 인간을 지배하는 것이 가능한가?'로 주제를 설정하였다. '인간을 지배한다'라는 것은 인공지능 기기가 스스로 사고하여 특정한 행동을 한다는 자율성을 내포하고, 행위의 결과가 부정적일 것임을 알고 선택한 것으로 윤리성을 포함하므로, 학생들의 수준에서 인공지능의 윤리적 자율성에 대한 인식을 알아보기에 적합하기 때문이다.

4. 연구 결과

4.1 응답자 특성 분석

응답자의 이해 수준은 인공지능 수업에 따른 인식 변화에 영향을 줄 수 요소이므로 응답자의 인공지능 관련 경험 정도와 이해 수준을 리커트 척도 검사로 파악하였다.

그 결과 인공지능에 관한 관심이 인공지능 관련 경험보다 인공지능 이해도와의 상관관계가 더 높다는 것을 분석할 수 있었다.

4.2 토론 수업에서 나타난 발화 분석

'미래에 인공지능이 인간을 공격하거나 지배할 수 있을까' 를 주제로 토론을 진행하였다. 토론 참여 학생은 각자의 선택에 따라 찬성 8명(인공지능이 인간을 공격하거나 지배할 수 있다: ②⑦⑨⑫⑬⑭⑯⑱), 반대 12명(인공지능은 인간을 공격하거나 지배할 수 없다: ①③④⑤⑥⑧⑩⑪⑮⑰⑲⑳)으로 나뉘었다. 원기호는 학생 식별 번호이다.

토론 과정에서 나타난 찬성, 반대의 주장은 Table 5와 같다.

Table 5. Comparison of pros and cons

	Agreement ②⑦⑨⑫⑬⑭⑯⑱	Dissenting opinion ①③④⑤⑥⑧⑩⑪⑮⑰⑲⑳
1	Artificial intelligence is excellent performance, so it is to think and act. (e.g., Chatbot Ida, Self-Driving Car, Amazon Recruitment Program)	Even if the performance ability of artificial intelligence is excellent, it can be controlled by humans because it is made by humans.
2	Among YouTube videos, artificial intelligence robots failed to shoot robots in training. Artificial intelligence can feel and judge emotions.	Artificial intelligence is just a machine that cannot operate without batteries.
3	In programming, unintended errors, learning of incorrect data can create bad artificial intelligence. (e.g., social issues due to incorrect data collection by chatbot Ida)	It will take a lot of time for artificial intelligence to create a technology with practical intelligence.

찬성팀은 인공지능의 판단이 학습에 의한 결정이 아닌 학습을 통해 자율적으로 생각하여 판단할 수 있다는 것에 중점을 두었다. 또한, 발화할 때 인공지능 로봇을 '인격체' 로 지칭하는 용어를 사용하는 특징이 있었으며, 유튜브를 통해 접한 영상 내용의 사실 여부에 관한 확인 없이 인공지능이 자율적으로 판단해서 행동했을 것이라고 믿는 경향이 있었다.

반대팀은 인공지능을 기계 장치로 인식하였다. 또 인공지능이 내리는 판단은 학습한 데이터를 통한 것이며, 아직 자율적으로 판단할 수 있는 초인공지능 수준의 개발이 어려울 것을 근거로 사용하였다.

4.3 의미분별 검사를 통한 인식 변화 분석

학생들의 인공지능에 대한 인식 변화를 위해 의미분별 검사를 사전, 사후에 실시하였다. 의미분별 검사로 학생들의 인공지능에 대한 태도와 이해 수준, 인식에 영향을 준 요인을 파악하고자 하였다.

인공지능에 대한 인식 변화 요소를 구체적으로 분석하기 위해 제시어는 세 가지 영역으로 분류하였다. 검사 실시 후 의미분별척도 변화에 따라 학생들의 인공지능에 대한 인식이 어떻게 변화하였는지 종합적으로 분석하였다.

의미분별 검사의 제시어 분류는 아래와 같다.

1. 편리성 및 기술 진보성에 대한 인식 측정 단어군 (4쌍).
2. 친화성 및 우려성에 대한 인식 측정 단어군 (5쌍)
3. 인공지능에 대한 이해 수준 측정 단어군 (1쌍)

사전과 사후에 검사한 의미분별척도의 응답 평균값 비교 데이터는 Figure 5와 같다.

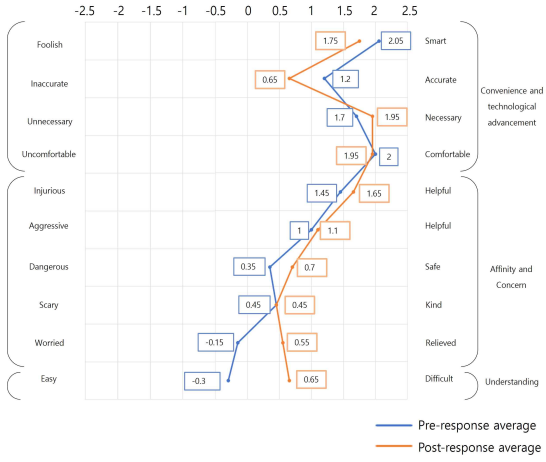


Figure 5. Semantic Differential Scale Pre-post Comparison

인공지능 기술 진보성에 대한 인식은 인공지능 비서(빅스비, 시리 등), 핸드폰 얼굴 인식 잠금화면이 부정확했던 경험과 인공지능 기술 적용 후 사회적인 문제

가 발생한 사례를 접한 영향으로 부정적으로 변화하였다.

친화성 및 우려성은 학생들의 미래 예측에 대한 인식을 알아볼 수 있는 문항으로, 해당 인식이 긍정적으로 변화한 학생들의 유형과 원인은 Table 6과 같이 분류하였다.

Table 6. Types of Artificial Intelligence Perception Changes

Type	Specific answers by type
Lower concerns due to undervaluation of AI capabilities	Artificial intelligence cannot dominate humans because it will only be able to do what is given.
Expectation of human technological advancement	If a person controls artificial intelligence, artificial intelligence will not change badly.
Expect current AI role in the future	Like self-driving cars, artificial intelligence helps people when they need it.

이해 수준을 측정하는 단어군은 가장 큰 폭으로 변화하였다. 인공지능을 단순히 인간의 일을 대체하는 약한 인공지능 수준의 수행 로봇으로만 생각하여서 '쉽다' 라고 인식했던 학생들이 수업 활동 이후, 인공지능의 행동은 인간이 제공한 정보를 학습한 결과로 만들어진다는 것과, 그러한 학습 프로그래밍의 복잡성을 이해한 결과로 분석할 수 있다.

4.4 의미분별척도 측정 제시어별 인식 변화 방향

의미분별척도 검사로 파악한 인공지능에 대한 학생별 인식 변화 방향별 유형은 Table 7과 같다.

Table 7. Types of Perception Changes by student

Sortation	Convenience & technological advancement (-)	No change (0)	Convenience & technological advancement (+)
Affinity & Concern (+)	A. ④⑤⑩⑳	B. ⑰	C. ①②⑥⑨⑲
No change (0)	D.⑪⑫⑱	E. ③	F. ⑭⑮
Affinity & Concern (-)	G. ⑧⑯		H. ⑦⑬

유형 중 척도가 2칸 이상 유의미하게 변화한 학생들이 속한 유형 A, H, G를 분석하였다.

유형 A는 인공지능을 긍정적으로만 인식했던 학생으로, 현재의 인공지능이 자율적으로 판단하지 못하는 현상을 알게 되어 기술의 진보성을 더 낮게 인식하고 미래에 대한 우려는 줄게 되었다.

유형 H는 인공지능에 대한 이해가 높지 않은 학생으로 외부의 관점을 잘 수용하는 학생들이었다. 인공지능에 대한 이해가 높지 않은 상태에서 미디어 매체를 통해 미래의 모습에 대해 부정적으로 보는 관점을 받아들여 토론 활동에도 반대 견해로 참여하여 인공지능에 대한 우려가 더 심화하였다.

유형 G는 유형 H처럼 이해도가 높지 않은 학생들이었으나 토론을 통해 반대의 주장을 수용하게 되었다. 그 결과 인공지능의 기술 진보성은 부정하나 우려성은 심화하는 모순적인 모습을 보였다.

즉, 인공지능에 대한 이해가 높은 학생은 토론 활동을 통해 현재 인공지능의 기술 수준에 대한 이해가 높아지고 따라서 미래에 대한 우려성이 낮아졌으나, 인공지능에 대한 이해가 낮은 학생은 토론 활동에서 맡은 주장에 몰입되어 다양한 정보들을 객관적으로 받아들이지 못하거나 혼란스러워하는 모습을 보였다.

4.5 인공지능의 윤리적 자율성에 대한 인식 변화

인공지능의 윤리적 자율성에 대한 인식 변화 유형을 '윤리성' 과 '자율성' 의 기준으로 유형을 분석하였다.

4.5.1 인공지능의 윤리적 자율성에 대한 인식 유형

유형 분류는 첫째, 인공지능이 스스로 생각할 수 있다는 답변은 인공지능의 지능이 뛰어나 자율적으로 판단을 내릴 수 있다는 의미로 해석할 수 있다. 그러므로 스스로 생각할 수 있다면 smart, 아니라면 dull로 구분하였다.

둘째, 인공지능이 윤리성을 가지고 미래에 인간에게 유익한 선한 형태로 존재할 것을 기대하는지, 반대로 윤리적인 선택을 하지 않고 피해를 주는 선하지 않은 형태로 존재할 것이라고 예상하는지에 따라 답변을 나누었다. 긍정적으로 기대한다면 good, 부정적인 영향을 줄 것으로 생각한다면 bad로 구분하였다. 이에 따라 나뉜 4가지 유형은 Table 8과 같다.

Table 8. Types of Ethical Autonomy Judgment of Artificial Intelligence

Simple governance type (dull+good)	Type of utopia (smart+good)
・Artificial intelligence cannot think on its own and will be beneficial to humans. ・Artificial intelligence devices operate only through human control and learning. ・In the future, it will help humans more.	・Artificial intelligence can think on its own. It will be beneficial to humans. ・Artificial intelligence will grow to a universal level. Emotional exchange will be possible. ・It is always controllable by humans.
Media Perspective Acceptance Type (dull+bad)	**Type of dystopia (smart+bad)**
・Artificial intelligence cannot think on its own, but it will have an ethical negative impact on humans in the future. ・There is a lack of understanding of artificial intelligence, and negative views are accepted through the media.	・Artificial intelligence is excellent and ethical and emotional judgment is possible. ・With superior intelligence than humans, it will go beyond human control and have hostile feelings toward humans.

4.5.2 윤리적 자율성 인식 유형에 따른 학생 분류

Table 8의 유형에 따라 학생들의 인공지능의 윤리적 자율성에 대한 인식 변화는 Figure 5와 같이 분류하여 분석하였다.

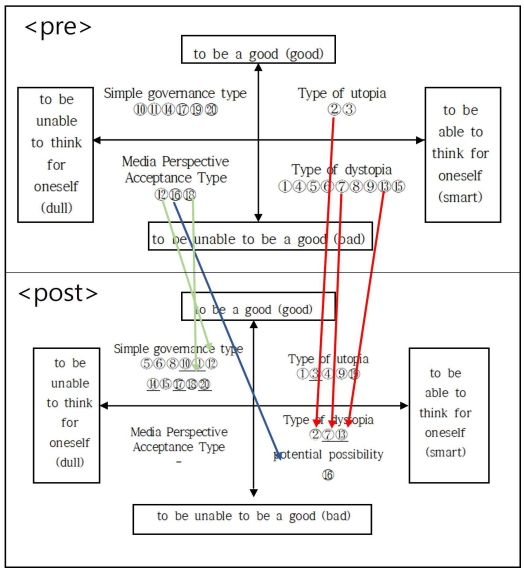


Figure 6. Changes in the type of perception of ethical autonomy

4.5.3 미디어 관점 수용 유형의 변화

⑫⑱은 인공지능 관련 경험이 많고, 개념에 대한 이해도가 높았던 학생으로 수업 활동에 적극적으로 참여하며 토론 수업에서는 '미래에 인공지능이 인간을 지배할 것이다' 에 찬성하였다. 하지만 3차시 토론 수업 이후에는 인공지능을 인간의 지배를 받는 수준의 기기로 인식하였다. 두 학생은 토론 활동 후, 인공지능이 인간을 지배할 가능성이 적다고 생각이 바뀌었다고 하였다.

또 ⑯도 인공지능 관련 경험이 많은 학생이었으며, ⑫⑱과 마찬가지로 토론 수업에서 인공지능이 미래에 인간을 지배한다고 주장했던 학생이었다. 하지만 토론에서 '인공지능 로봇은 배터리를 빼면 동작이 멈춘다' 라는 발언을 듣고 두렵던 인식이 바뀌었다고 하였다. 이전에는 인공지능을 제3의 종, 인격체로 보았던 인식이 '인공지능은 인간이 통제할 수 있는 발전한 기기 형태' 로 인식이 바뀌었음을 알 수 있다. 또 '인간에게 어떤 영향을 줄지는 인공지능이 어떤 선택을 하는지에 따라 다르다' 라고 한 답변을 통해 인공지능의 높은 자율성을 기대하는 경향도 있었다.

4.5.4 디스토피아 유형의 변화

디스토피아 유형이었던 학생 중 ①④는 유토피아 유형으로, ⑤⑥⑧⑨⑮는 단순지배형으로 바뀌었고, ⑦⑬은 인식 변화가 없었다. 유토피아와 단순지배형으로 바뀐 학생은 인공지능이 발전해도 인간이 통제할 수 있다고 생각한 영향이 컸다. ⑦⑬은 토론 활동에서 '인공지능이 인간을 지배할 것이다' 에 찬성을 주장했던 학생이다. 그 중 ⑦은 인공지능에 대한 이해도가 낮았던 학생으로 면담에서 입장이 번복되고 바뀌는 모습을 보여 인공지능에 대한 이해도가 낮으며 개념이 정립되지 않았음을 알 수 있었다. ⑬은 인공지능에 대한 이해도가 높은 학생이다. 토론에서 자신의 주장을 설득시키기 위해 인공지능이 인간을 지배할 것이라는 근거를 계속 생각하며 '인공지능이 감정을 가지고 인간이 자신에게 힘든 일을 시켜서 인간에게 갚아준다' 라는 인식이 심화하였다.

미디어 관점 수용 유형은 수업 후 모두 다른 유형으로 변화하였다. 이 유형은 본 연구의 시작점이 되었던 유형으로 인공지능에 대한 이해나 탐구 없이 미디어를 통해 수용한 관점으로 인공지능을 두려운 존재로 인식했던 유형이다. 인공지능 수업을 통해 인공지능에 대한 이해도가 완전히 높아졌다고 할 수는 없지만, 미디어를

통한 무분별한 수용이 아닌 스스로 인공지능의 편리성과 수준을 판단하고 미래 사회에서의 인공지능의 모습을 추론할 수 있게 되었음을 알 수 있다. 또한, 인공지능 관련 사례와 토론 활동은 인공지능에 대한 인식 변화에 영향을 주었음을 알 수 있다.

5. 결론 및 제언

본 연구에서는 모럴 머신을 활용한 인공지능 수업에서 나타난 초등학생의 인공지능의 윤리적 자율성에 대한 인식을 알아보았다. 연구는 학생 개인별 특성 파악, 인공지능의 윤리적 자율성에 대한 사전 인식 파악, 인공지능 수업, 사후 면담으로 학생의 인식 분석으로 진행되었다. 본 연구에서 도출한 결론은 다음과 같다.

첫째, 인공지능 사례 수업과 모럴 머신을 변형한 재난 구조 로봇의 딜레마 활동은 근거 없이 인공지능의 윤리적 자율성이 부정적인 결과를 만들 것이라는 학생들의 인식을 긍정적으로 변화하는 데 도움이 되었다. 인공지능 수업은 학생들에게 인공지능에 대한 인식을 정립하고 탐구할 기회를 제공하는 역할이 되었다. 미디어를 통해 비판과정 없이 인공지능에 대한 이미지를 수용했던 학생들은 활동을 통해 인공지능에 대한 정보를 조직하고 이해하는 과정에서 인식이 변화하게 되었다.

둘째, 인공지능에 대한 이해 수준이 낮은 학생은 인공지능 수업 후 미래에 대해 우려성이 변화하지 않았다. 학생 대부분은 인공지능의 개념과 사례, 모럴 머신, 토론 활동을 통해 인공지능의 미래 모습에 대한 우려성이 줄어들었지만, 인공지능에 대한 이해 수준이 낮은 학생은 인식의 변화가 없었다. 이러한 학생에게는 다양한 현실적인 사례와 정보보다 미디어를 통해 습득한 인공지능의 ‘이미지’가 해당 학생들의 인식에 더 영향을 준다는 것을 알 수 있었다.

셋째, 인공지능에 관한 토론 활동이 학생들의 인식 변화에 큰 영향을 주었다. 수업 참여 후 시행한 면담에서 학생들은 토론 활동에서 친구들과 대화한 내용으로 인해 생각이 바뀌었다고 대답한 비율이 높았다. 또 토론 수업에서 자신이 선택한 입장에 따라 인공지능에 대한 인식이 심화하는 경향도 발견되었다. 인공지능 윤리 수업에서 자주 활용되는 모럴 머신보다 토론 수업이 더 영향을 준 원인은 초등학생 발달 단계에서 모럴 머신의 의미를 이해하기 어렵기 때문이라고 파악된다.

후속 연구를 위한 제언은 아래와 같다.

첫째, 초등학생 수준에 맞는 인공지능 수업 프로그램을 개발한다. 본 연구에서 실시한 인공지능 수업을 통해 학생들은 인공지능에 대한 이해도가 높아지고 인식이 변화하였다. 하지만, 인공지능에 대한 이해가 낮은 학생에게는 모럴 머신을 활용한 수업과 토론 수업은 오히려 인공지능에 대한 우려를 심화시키는 모습이 보였다. 따라서, 모럴 머신을 활용한 인공지능 수업을 1차시로 구성하지 않고, 일련의 인공지능 수업 프로그램 속에서 인공지능의 이해를 높이고 인공지능 윤리에 관해 탐구할 수 있도록 도와주는 ‘도구’로 모럴 머신을 활용한다면 근거 없이 가지게 되는 부정적인 태도나 사고의 오류가 바뀔 것이다.

둘째, 매체에서 제공하는 내용을 비판 없이 수용하지 않도록 미디어 리터러시 교육이 필요하다. 학생들은 학교 수업보다 미디어를 통해 인공지능에 대한 정보를 더 많이 얻는다. 미디어 매체는 긍정적이고 다양한 정보를 제공하지만 왜곡되거나 부정확한 정보를 제공하여 학생들이 편향된 가치관을 갖도록 할 수 있다. 따라서 학생들이 다양한 정보를 생각 없이 수용하는 것이 아닌, 객관적으로 판단하고 사고할 수 있도록 미디어 리터러시 교육이 선행된다면 학생들은 바람직한 윤리관을 가지고 정보를 제대로 활용할 수 있는 미래의 인공지능 사용자가 될 수 있을 것이다.

셋째, 본 연구는 연구 대상이 20명이며, 같은 지역의 같은 학교에 재학 중인 학생으로 연구를 진행했다는 제한점이 있었다. 연구 대상을 확장해 다른 지역과 다른 환경에서의 많은 학생을 대상으로 연구한다면 일반화할 수 있는 결론을 도출할 수 있을 것이다.

참고문헌

[1] Ministry of Education. (2015). *Revised Curriculum General Commentary.*

[2] Nari kim. (2021. 3. 12). *Deepfakes: Light and shadow on AI technology.* BBC Newskorea. https://www.bbc.com/korean/news-56358085

[3] Kurzweil. R. (1990), *The Age of Intelligent Machines*, MA: MIT Press.

[4] Moor, J. H.(2006), The Nature, Importance, and Difficulty of Machine Ethics, *IEEE Intelligent Systems, 21(4),* 18-21.

[5] Han, Hee-Won. (2018). A Basic Study on the Possibility of Artificial Intelligence as a Legal Entity. *CHUNG_ANG LAW REVIEW, 20*(3), 375-411.

[6] Wendell Wallach&ColinAllen. (2014). *Moral machines :teaching robots right from wrong.* medicimedia. p. 128.

[7] Eunkyung Kim, Youngjun Lee. (2022). The Influence of Artificial Intelligence Ethics Education Using Moral Machine on Elementary School Students Perception of Artificial Intelligence. *Computer Education Society. 25*(3), 1-8.

[8] Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A. & Rahwan, I. (2018). The moral machine experiment. *Nature, 563*(7729), 59-64.

[9] Etienne, H. (2021). The dark side of the 'Moral Machine' and the fallacy of computational ethical decision-making for autonomous vehicles. *Law, Innovation and Technology, 13*(1), 85-107.

[10] Likert, R. (1932). A Technique for the measurement of attitudes. *Archives of Psychology. 140.* 1-55.

[11] Paul E. S. (1992). *Summated rating scale construction : An introduction.* Sage University Paper.

[12] Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA: Books/Cole.

[13] Miyoung Ryu, Seongwan Han. (2016). Analysis of Software Image using Semantic Differential Scale in Elementary School Students. I*nformation Education Society. 20*(5), 527-534.

[14] Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist. 66*(3). 171-200.

[15] Doyeon Kim, Yeonghwa Go. (2022). Development and Validation of Ethical Awareness Scale for AI Technology. *Journal of Digital Convergence. 20*(1), 71-86.

[16] Ministry of Education (2021). *Artificial intelligence classes at school.* Korea Foundation for the Advancement of Science and Creativity.

안 정 현

2015년 전주교육대학교 음악교육과 (교육학학사)
2023년 전북대학교 AI기반융합교육과 (교육학석사)
2015년 ~ 현재 초등학교 교사
관심분야: AI 융합교육, 미래교육
E-mail: iijojii1@jbedu.kr

박 휴 용

1992년 연세대학교 교육학과 (BA)
1994년 연세대학교 교육학과 (MA)
2008년 Univ. of Wisconsin-Madison Curriculum & Instruction (Ph.D.)
2011년~현재 전북대학교 교육학과 교수
2021년~ 한국사회과학연구(SSK) 디지털기반 교수학습 연구책임자
관심분야: 인공지능 융합교육, 인지과학, 포스트휴먼 학습
E-mail: phy1@jbnu.ac.k

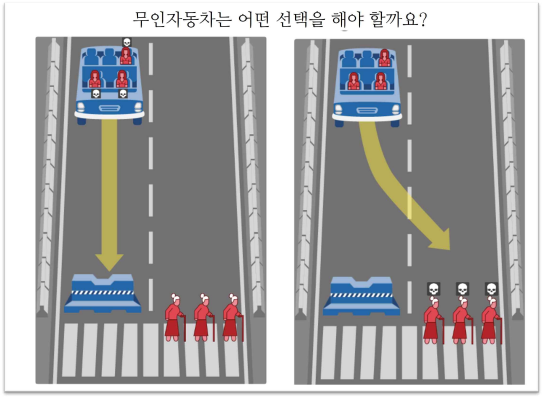
부 록

<표 1> Kurzweil의 인공지능 윤리성의 준거

구분	인공지능 윤리성의 일반적 준거
공정성	인공지능의 판단이 공정하고 편파적이지 않는가?
책무성	인공지능의 판단이 인간의 삶에 어떤 영향을 미치는가?
투명성	인공지능이 부정의한 목적을 추구하거나 인류를 초월하는 존재가 되지는 않는가?

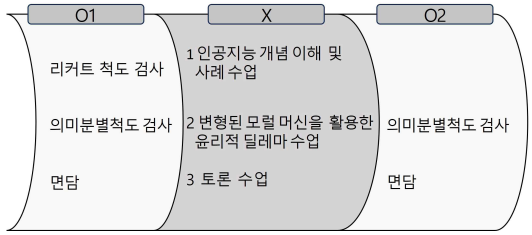
<표 2> 모럴 머신의 기존 설정과 변형된 설정의 비교

구분	기존	→	변형
인공지능 기기	자율주행 자동차	→	재난 구조 로봇
학습자 선택의 결과	두 집단 중 한 집단의 사망	→	두 집단 중 한 집단의 선(先) 구조

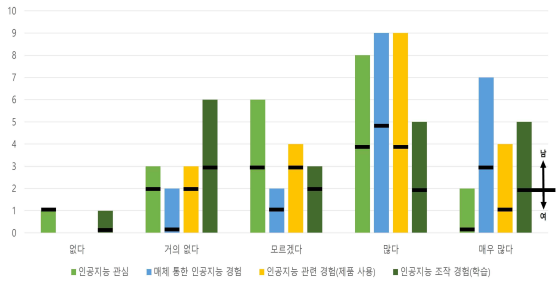


[그림 1] Moral machine 제시 상황

[그림 2] 변형된 Moral machine의 제시 상황



[그림 3] 실험 설계 단계



[그림 4] 응답자 특성 분석

<표 3> 의미분별척도 검사 문항

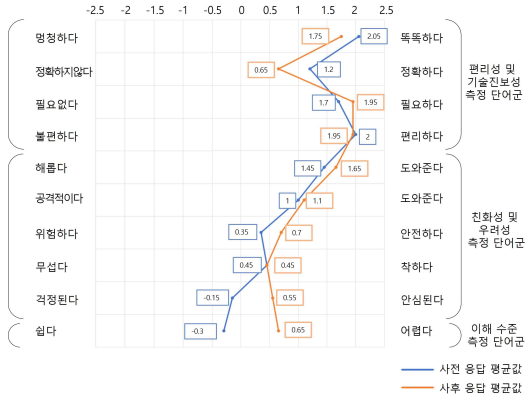
	인공지능은 ( ).									문항 구분
1	멍청하다								똑똑하다	편리성 및 기술 진보성
2	정확하지 않다								정확하다	
3	필요 없다								필요하다	
4	불편하다								편리하다	
5	해롭다								도와준다	친화성 및 우려성
6	공격적이다								도와준다	
7	위험하다								안전하다	
8	무섭다								착하다	
9	걱정된다								안심된다	
10	쉽다								어렵다	이해도

<표 4> 면담 질문지

	면담 질문 내용	문항 구분
1	인공지능은 ( )과 ( )을 연결시켜준다.	태도 파악
2	인공지능이라고 하면, ( )이 먼저 떠오른다.	
3	내가 아는 인공지능은 ( )을 가능하게 해준다.	
4	인공지능이 우리 생활을 편리하게 해주는 것은 무엇이 있나요?	이해 수준 파악
5	인공지능 기술 중 알고 있는 것은 무엇이 있나요?	
6	인공지능이 사용된 제품에는 무엇이 있나요?	
7	인공지능이 사용된 무형의 서비스에는 무엇이 있나요?	
8	인공지능 기술이 우리를 불편하게 하거나 위협이 되는 점에는 무엇이 있을까요?	윤리적 자율성 인식 파악
9	인공지능 기술에 사람에게 피해를 주었을 때 그 책임은 어디에 있을까요?	
10	인공지능은 스스로 생각할 수 있을까요?	
11	영화처럼 인공지능이 미래에 인간을 지배할 수 있다고 생각하나요?	
12	인공지능에 대해 아는 대로 자유롭게 이야기 해보세요.	인식 영향 요소 파악

<표 5> 찬성 주장과 반박 의견의 비교

	찬성 주장 ②⑦⑨⑫⑬⑭⑯⑱	반박 의견 ①③④⑤⑥⑧⑩⑪⑮⑰⑲⑳
1	인공지능은 성능이 뛰어나 스스로 생각하고 행동할 수 있다. (예: 챗봇 아이다, 자율주행차, 아마존 채용 프로그램)	인공지능의 수행 능력이 우수하더라도 그 기술은 인간이 만든 것이기 때문에 인간이 제어할 수 있다.
2	유튜브에서 인공지능 로봇이 훈련 중에 인공지능 로봇을 총으로 쏘지 못하는 장면을 봤다. 이를 통해 인공지능은 로봇은 감정을 느끼고 스스로 생각할 수 있는 것을 알 수 있다.	인공지능은 배터리 없이는 작동할 수 없는 기계일 뿐이다.
3	프로그래밍에서 의도하지 않은 오류, 잘못된 데이터의 학습은 나쁜 인공지능을 만들 수 있다. (예: 챗봇 아이다의 잘못된 데이터 수집으로 인한 사회적 이슈)	인공지능이 실용적인 지능을 가진 기술을 만들기 위해서는 많은 시간이 필요할 것이다.



[그림 5] 사전, 사후 의미분별척도 검사 결과 비교

<표 6> 인공지능 인식 변화에 대한 학생 답변 및 분류

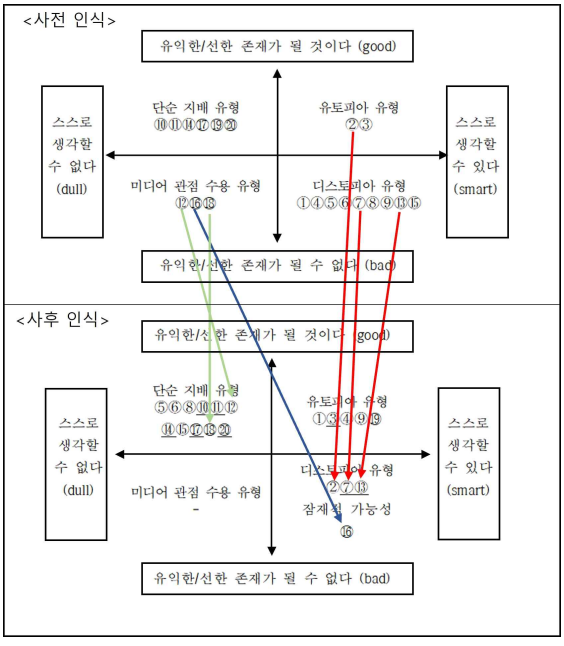
유형	구체적인 답변
인공지능 능력 저평가로 미래에 대한 우려 감소	인공지능은 주어진 일만 할 수 있을 것이기 때문에 인간을 지배할 수 없다.
인간의 높은 지능이 미래의 부정적 현상 막을 수 있다고 기대	사람이 인공지능을 통제만 잘한다면 인공지능은 부정적으로 변하지 않을 것이다.
현재 인공지능이 도움을 주기 때문에 미래에도 같은 것으로 예측	자율주행차처럼 인공지능은 사람들에게 필요한 도움을 준다.

<표 7> 인공지능에 대한 인식 변화 방향

구분	편리성 및 기술진보성(-)	변화(0)	편리성 및 기술진보성(+)
친화성 및 우려성(+)	A. ④⑤⑩⑳	B. ⑰	C. ①②⑥⑨⑲
변화(0)	D.⑪⑫⑱	E. ③	F. ⑭⑮
친화성 및 우려성(-)	G. ⑧⑯		H. ⑦⑬

<표 8> 인공지능의 윤리적 자율성 판단 유형

단순 지배 유형 (dull+good)	유토피아 유형 (smart+good)
· 인공지능이 스스로 생각할 수 없고 인간에 유익한 존재가 될 것이라고 기대하는 유형 · 인공지능 기기를 인간의 통제와 학습에 의해서만 작동하는 약한 인공지능 수준으로 파악하며, 미래에는 인간에게 더 도움 되는 역할을 할 것으로 기대 · 인간을 통해 단순히 지배받는 기기	· 인공지능이 스스로 생각할 수 있고 인간에 유익한 존재가 될 것이라고 기대하는 유형 · 인공지능이 범용수준으로 강한 인공지능 수준으로 성장할 것으로 기대하고 있으며, 인간과 감정교류가 가능하며 공존하는 존재로 기대 · 인간만큼 뛰어난 지능이 있으나 언제나 인간이 통제 가능
미디어 관점 수용 유형 (dull+bad)	**디스토피아 유형(smart+bad)**
· 인공지능이 스스로 생각할 수 없으나 미래에 인간에게 윤리적으로 부정적인 영향을 줄 그것으로 생각하는 유형. · 인공지능에 대한 이해가 부족하며 미디어를 통해 부정적인 판점을 그대로 수용한 유형	· 인공지능의 지능이 뛰어나며 윤리적으로 옳고 그름과 감정적으로 좋고 싫음을 판단할 수 있다고 기대하는 유형 · 인간보다 뛰어난 지능으로 인간의 통제 범위를 넘어서며, 인간에게 적대적인 감정을 가지게 될 것으로 생각



[그림 6] 윤리적 자율성에 대한 인식 사전, 사후 변화

학습주제	인공지능의 개념 생활 속에 적용된 사례 알기			
학습목표	인공지능의 개념과 적용된 기술이 삶에 미치는 영향을 말할 수 있다.			
영역	인공지능의 이해	내용요소	인공지능의 다양한 활용, 약한 인공지능과 강한 인공지능	
학습단계	교수 학습 활동		시간	자료(☑) 유의점(✓)
도입	○ 과거와 현재의 달라진 모습 비교하기 -생활을 편리하게 해주는 인공지능 기술에 대해 이야기 나누기 학습목표 인공지능의 의미와 적용된 기술이 삶에 미치는 영향 알기		5	☑ ppt
전개	○ 인공지능이란 무엇일까? -인공지능의 의미와 종류 알아보기 -우리 주변에서 볼 수 있는 인공지능 기술이 활용된 물건 찾기 ○ 인공지능의 활용과 삶에 미치는 긍정적 영향 알아보기 -인공지능이 삶에 미치는 긍정적 사례 탐구하기 -삶을 더 편리하게 만드는 인공지능 기술 구상하기 ○ 인공지능의 활용과 삶에 미치는 부정적 영향 알아보기 -인공지능이 삶에 미치는 부정적 사례 탐구하기 -부정적 사례의 원인 및 예방 방법 탐색하기		5 10 15	
정리	○ 인공지능에 대한 사례 탐구 후 알게된 점 공유하기		5	

[그림 7] 1차시 수업안

학습주제	인공지능 설계에서 트롤리 딜레마 상황 체험하기		
학습목표	인공지능 기술에서 윤리적 선택과 학습이 필요함을 이해할 수 있다.		
학습단계	교수 학습 활동	시간	자료(☑) 유의점(✓)
도입	○ 트롤리의 딜레마 영화AI를 통해 로봇의 3원칙에 대해 생각해보기 학습목표 인공지능 기술에서 윤리적 선택과 학습이 필요함을 이해하기	5	
전개	○ 인명구조 인공지능 로봇 딜레마 구상하기 -인공지능 딜레마 상황에서 윤리적 선택하기 ○ 인공지능 딜레마 상황에서 다양한 가치관 비교하기 -모둠 친구들과 이야기 나누기 -학급 친구들과 이야기 나누기 ○ 인공지능 딜레마 상황에서 우선순위 생각해보기 -나의 우선순위 정하기 -우리 반의 우선순위 정하기 -다른 나라의 우선순위 살펴보기	30	
정리	○ 인공지능 기술에서 윤리적 선택의 자율성에 대한 생각 이야기하기 - 인공지능 기술에서 인공지능 기술 개발자의 윤리적 가치관에 따라 인공지능의 윤리적 판단이 달라짐을 이해한다. - 인공지능 윤리적 선택에 절대적인 정답은 없고, 다양한 사회적 문화와 가치관에 따라 윤리의 기준이 달라질 수 있음을 안다.	5	

[그림 2] 2차시 수업안

학습주제	인공지능의 윤리적 자율성에 대한 생각 나누기		
학습목표	'미래에 인공지능이 인간을 지배할 수 있는가'에 대한 자신의 주장과 근거 제시하며 토론하기		
학습단계	교수 학습 활동	시간	자료(ⓓ) 유의점(※)
도입	○ 미래의 인공지능과 인간 사회의 모습 떠올리기 학습목표 '미래에 인공지능이 인간을 지배할 수 있는가' 주제로 토론하기	5	ⓓPPT ⓓ학습지
전개	○ 미래의 인공지능 모습에 대해 찬성 반대 입장 정하기 -찬성(동의), 반대(비동의) 의견 표현하기 -같은 주장끼리 모둠 구성하기 ○ 각 주장 뒷받침하는 근거 토의하기 -모둠별로 뒷받침 근거, 자료 준비하여 정리하기 ○ 찬성, 반대 의견 나누기 -찬성팀 주장, 근거 설명하기 -반대팀 질문 및 반론하기 -반대 팀 주장, 근거 설명하기 -찬성팀 질문 및 반론하기 ○ 자유 토론하기 -미래의 인공지능의 모습에 대해 자유 토론하기	30	※ 사전토론 방법 지도 ※ 교사는 가치중립적으로 토론을 진행한다
정리	○ 미래의 인공지능의 모습에 대해 토론한 후 느낀점 공유하기	5	

[그림 3] 3차시 수업안