

인공지능 윤리 교육을 위한 윤리적 딜레마 시나리오 내용 개발 및 효과성 분석: Moral machine 대안의 탐구

Development and Effectiveness Analysis of Ethical Dilemma Scenario for Artificial Intelligence Ethics Education: Inquiry of Moral Machine Alternatives

홍은정[†] · 오경선^{††}

Eun-Jung Hong[†] · Kyung-Sun Oh^{††}

요 약

인공지능 기술의 사회적 영향 증대로 인해 인공지능 윤리 교육의 중요성이 강조되고 있다. 교육 현장에서 인공지능 윤리 교육을 위한 교육 실습 자료가 부족한 상황에서 Moral machine은 교육 도구로 활발하게 활용되고 있다. 그러나 Moral machine을 활용하는 과정에서 학습자에게 인공지능 기술에 대한 부정적인 인식을 형성할 수 있다. 이에 본 논문은 기존의 Moral machine을 활용한 인공지능 교육의 한계를 고려하여 새로운 윤리 교육의 시나리오를 제안하고 개발하였다. Moral machine을 기반으로 새로운 윤리 교육 시나리오를 초등학교 6학년생을 대상으로 실험하였으며, 긍정적인 효과를 보였다. 이러한 결과를 통해, 초등학교생을 대상으로 하는 인공지능 윤리 교육에서 새로운 윤리 교육 시나리오인 변형 Moral machine의 활용이 부정적 영향을 최소화하고 긍정적인 학습 경험을 제공할 방안으로 제안된다. 본 연구는 초등학교생의 교육 과정에서 인공지능 윤리 교육의 중요성을 강조하며, 미래 연구를 위한 기초를 마련하고자 한다. 더 많은 연구가 인공지능 윤리 교육 프로그램의 개발과 효과를 탐구하여 교육의 효율성을 향상하는 데 기여하길 바란다.

주제어: 인공지능 윤리 교육, 초등학교생, 인공지능에 대한 인식, 태도, 모델머신, 윤리적 딜레마

ABSTRACT

The importance of artificial intelligence ethics education is being emphasized due to the increasing social impact of artificial intelligence technology. Moral machines are actively used as educational tools in a situation where there is a lack of educational practice materials for artificial intelligence ethics education in the educational field. However, in the process of using Moral machines, learners can form negative perceptions of artificial intelligence technology. Therefore, this paper proposed and developed a new ethics education scenario in consideration of the limitations of artificial intelligence education using existing Moral machines. Based on the Moral machine, a new ethics education scenario was tested on 6th graders of elementary school, and it showed a positive effect. Through these results, the use of a modified Moral machine, a scenario of new ethics education in artificial intelligence ethics education for elementary school students, is proposed as a way to minimize negative effects and provide a positive learning experience. This study emphasizes the importance of artificial intelligence ethics education in the curriculum of elementary school students and intends to lay the foundation for future research. It is hoped that more research will contribute to improving the efficiency of education by exploring the development and effectiveness of artificial intelligence ethics education programs.

Keywords: Artificial intelligence ethics education, Elementary school student, Perception of AI, Attitude, Moral machine, Ethical dilemma

[†]중신회원: 건국대학교 교육대학원 석사

^{††}정 회 원: 건국대학교 상허교양대학 조교수(교신저자)

1. 서론

인공지능의 급속한 발전이 사회 전반에 영향을 미치면서 인공지능 기술에 대한 기대가 나날이 높아지고 있다. 현재 인공지능이 사회에 미치는 관심과 영향력이 커진 만큼 인공지능 윤리 문제 또한 인공지능 전문가들만의 문제가 아닌 모든 사회 구성원의 문제가 되었다[1]. 이에 따라 국내외의 여러 기관에서는 인공지능과 관련된 윤리적 이슈에 대응하기 위한 윤리 원칙들을 제시하고 있다. 유럽연합(EU)은 2019년 ‘신뢰할 수 있는 AI 윤리 가이드라인’을 발표하였고[2], 미국도 IEEE에서 2016년 ‘EAD(Ethically Aligned Design)’를 통해 인공지능 윤리 기준을 제시하였다[3].

우리나라에서는 과학기술정보통신부가 2020년 발표한 ‘사람이 중심이 되는 국가 인공지능 윤리 기준안’을 통해 ‘인간성을 위한 인공지능(AI for Humanity)’을 목표로 3대 기본원칙과 10대 핵심 요건을 제시했다[4]. 또한 국가인권위원회는 2022년 ‘인공지능 개발과 활용에 관한 인권 가이드라인’을 마련하여 인공지능 윤리 기준을 발표하였다[5]. 교육 분야에서도 인공지능 윤리 교육의 중요성이 강조되고 있다. 2022년 개정 교육 과정에서는 인공지능 윤리를 새로운 정보기술 사회에서 필수적으로 요구되는 디지털 역량으로 명시하고 있다[6]. 이처럼 인공지능 윤리 교육의 중요성이 강조되고 있음에도 불구하고, 현재까지도 인공지능 윤리의 체계적이고 효과적인 교육을 위한 연구는 미비한 실정이다[2]. 현재 직접적인 인공지능 교육 실습 자료가 부족한 상황에서 Moral machine은 윤리적 결정이나 책임성, 안전성 등을 학습할 때 인공지능 윤리 교육 자료로서 교육 현장에서 활발하게 활용되고 있다. 그러나 Moral machine은 처음부터 교육용으로 개발된 것이 아니며, 이를 교육적으로 활용하는 과정에서 발생할 수 있는 학생들의 감정적 반응과 인공지능 기술에 대한 부정적 인식에 대한 우려의 목소리가 커지고 있다. 김은경(2022)은 Moral machine을 활용한 인공지능 윤리 교육을 통해 학생들이 죽음과 희생을 반복적으로 선택함으로써 생명을 잃는 상황을 간접적으로 경험하게 되면, 학생들에게 인공지능 기술에 대한 부정적 인식을 형성할 수 있다는 연구 결과를 제시했다[7].

이에 본 연구는 Moral machine을 활용한 교육에서 발생할 수 있는 인공지능 기술에 대한 부정적 영향을 고려하여 죽음과 희생을 배제하고 도움의 우선순위를 선택하는 새로운 윤리적 딜레마 시나리오를 개발하고,

초등학교 6학년 학생들을 대상으로 이를 활용한 교육을 함으로써 두 가지 교육 방법이 초등학교생들의 인공지능 태도와 인식에 미치는 영향을 비교 분석하고자 한다.

2. 연구 배경

2.1 인공지능 윤리 교육

전 세계적으로 증가하는 인공지능(AI)에 관한 관심은 윤리적 이슈에 대한 이해와 올바른 인식을 강조하고 있다. 이에 따라 올바르게 인식할 수 있는 능력을 함양할 수 있도록 인공지능 윤리에 대해 교육하는 것이 필요하다[8].

현재 국내외의 교육 분야에서 다양한 인공지능 윤리 교육을 연구하고 있다. 미국의 경우 하버드 대학교 ‘버크만 클라인 센터’(Berkman Klein Center)에서 2016~2019년 발표된 AI 윤리 원칙에 대한 다양한 국가 및 기구의 문서를 분석하여 ‘프라이버시’, ‘책임성’, ‘안전 및 보안’, ‘투명성 및 설명 가능성’, ‘공정성 및 차별금지’, ‘기술에 대한 인간의 통제’, ‘전문적 책임’, ‘인간 가치 증진’ 등 총 8개의 주요 주제로 정리하였다[9]. 또한 2019년에 미국 컴퓨터과학교사연합회(CSTA)는 AI4K12를 통해 초·중등 학생 대상의 인공지능 교육을 위한 ‘5가지 빅아이디어’를 ‘인식’, ‘표현 및 추론’, ‘학습’, ‘상호작용’, ‘사회적 영향’으로 구분하고 인공지능 윤리를 ‘사회적 영향’ 영역에 포함하였다[10].

우리나라에서는 교육부에서 2015년 개정된 교육 과정의 일부 보완을 통해, 고등학교에서 학생들이 선택할 수 있는 진로 과목으로 ‘인공지능 기초’ 교과가 신설되었[11], 이를 바탕으로 인공지능 윤리 교육을 하고 있다. 또한 2020년 ‘학교에서 만나는 인공지능’ 교재를 통해 ‘초·중등 인공지능 교육 내용 기준’을 제시하고, ‘인공지능의 사회적 영향’ 영역에 인공지능 윤리를 다루도록 하였다[12]. 2022년에는 ‘교육 분야 인공지능 윤리 원칙’을 발표하고 ‘사람의 성장을 지원하는 인공지능’을 대원칙으로 학습자의 주도성 강화, 교수자의 전문성 존중, 기술의 합목적성 제고를 위해 필요한 9대 세부 원칙을 제시하였다. 장연주 외(2022)는 초·중등 학생을 위한 인공지능 윤리 교육 주제를 ‘신뢰성’, ‘편향성’, ‘악용 가능성’, ‘책임성’ 네 가지로 선정하고 교육 프로그램을 개발 및

적용하여 효과성을 검증하였다. 그 결과, 학생들의 인공지능 윤리에 대한 흥미와 인공지능 윤리 교육의 필요성에 대한 인식 및 태도 모두 유의미한 변화를 나타냈다[13]. 송주영 외(2023)는 책임윤리를 기반으로 한 초등학생을 위한 인공지능 윤리 교육 프로그램을 개발하고 적용하여, 학생들이 인공지능 윤리에 대한 중요성을 바탕으로 인공지능 윤리 의식이 향상되는 유의미한 효과를 확인했다[14]. 이처럼, 교육 현장에서 학습자의 수준을 고려한 윤리 교육을 위해 다양한 연구가 진행되고 있다. 특히 인공지능의 빠른 발전으로 초등학생도 인공지능 기술과 상호작용하게 되면서 인공지능 윤리 교육의 중요성이 더욱 강조되고 있다.

2.2 Moral machine

Moral machine은 MIT(Massachusetts Institute of Technology)의 Lyad Rahwan 교수 연구실인 Scalable Cooperation 그룹이 개발한 온라인 플랫폼으로 자율주행 자동차의 도로 주행 상황에서 도덕적 선택 조건 상황을 제시하여 사람들이 두 가지 선택지에서 내린 결정 정보를 수집, 분석하는 사이트이다[15]. Moral machine은 현재 교육 현장에서 인공지능 윤리 교육 자료로 활발하게 활용되고 있으며, 윤리적 딜레마 상황에서 자율주행차가 결정을 내리는 과정을 통해 학습자들이 인공지능 윤리에 대해 생각해 보도록 하고, 13가지 딜레마 상황을 판단한 후에는 다른 사람의 응답과 자신의 응답을 비교해 볼 수 있도록 설계되었다.

Moral machine은 ‘트롤리 딜레마’ 라는 흥미로운 주제를 사용하여 누구나 쉽게 사용할 수 있는 직관적인 구성으로 학생들의 관심을 끌기에 적합한 도구로 평가된다. 그러나 반복적으로 인공지능 기기가 생명을 해하는 상황을 간접 경험하는 것이, 인공지능 기기나 기술에 대한 부정적인 인식을 유발할 수 있어 초등학생의 인공지능에 대한 태도에 부정적인 영향을 미칠 수 있다[16][17][18].

2.3 Moral machine 활용 교육

현재 교육 현장에서 인공지능 윤리 교육 시 자주 사용되는 실습 교육 도구로는 엔트리 인공지능, code.org, Moral machine 등이 있다. 그중 Moral machine은 인공지능 윤리의 책임성, 안전성, 윤리적 딜레마 교육 도구로 자주 사용되고 있다. 교육부에서 발행한 ‘인공지능

교육 길라잡이’에는 초등학생을 대상으로 한 수업 사례로 Moral machine을 활용한 인공지능의 도덕적 문제를 탐구하는 내용이 소개되어 있으며, 2021학년도 2학기부터 도입된 고등학교 인공지능 기초 교과서 7종은 모두 Moral machine이나 자율주행 자동차의 윤리적 딜레마 상황을 다루고 있다[16]. 또한 정보통신정책연구원(KISDI)과 과학기술정보통신부는 2020년 ‘인공지능(AI) 윤리기준’을 발표하고 핵심 가치 확산과 초·중·고등학생들의 인공지능 역량을 강화하기 위해 인공지능 윤리 교육 교재를 개발하였으며, 이 교재를 통해 Moral machine을 활용하여 책임성과 안전성에 관한 교육을 진행하고 있다.

3. 연구 절차 및 교육 프로그램 개발

3.1 연구 설계 및 절차

본 연구는 초등학생을 위한 변형 Moral machine이 학습자에게 미치는 영향을 알아보기 위한 것이다. 연구 설계는 실험 집단과 통제 집단으로 나누어 두 집단에 각각 변형 Moral machine과 Moral machine을 활용한 인공지능 윤리 교육을 진행하였고, 교육이 종료된 후 동일 검사 도구를 활용하여 검사를 시행하였다.

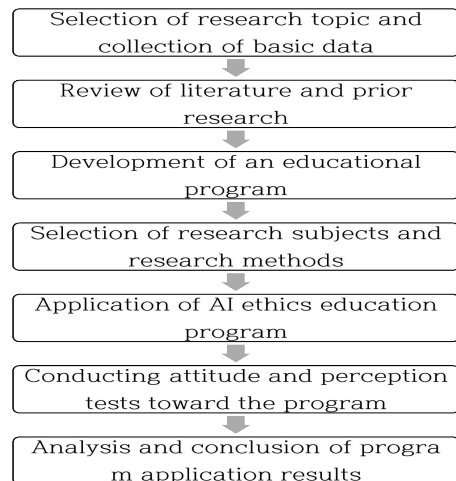


Figure 1. Research procedure

본 연구는 인공지능 윤리 교육에서 학습자들에게 인공지능에 대한 부정적인 영향을 최소화하기 위해, 죽음과 희생을 배제한 새로운 Moral machine 교육 프로그램

램을 개발하고자 하였다. 본 연구의 연구 절차는 Figure 1과 같다.

첫째, 국내의 인공지능 윤리 교육을 분석하고, 윤리적 딜레마에 대한 교육 프로그램에 대해 알아보고 연구 주제를 선정하였다.

둘째, 다양한 매체 분석을 통해 인공지능 윤리적 딜레마에 대한 문헌 고찰과 Moral machine이 학습자에게 미치는 영향에 관한 선행 연구를 분석하여 변형 Moral machine의 개발 내용을 도출하였다.

셋째, 선행 연구를 통해 선정된 주제에 따라 변형 Moral machine 시나리오를 설계 및 개발하였으며, 이후 개발된 변형 Moral machine은 인공지능 교육을 연구하고 있는 교수 및 대학원생 3인과 검토를 진행하였고, 이를 바탕으로 최종 프로그램을 산출하였다.

넷째, 인공지능 교육을 받은 경험은 있지만, 인공지능 윤리 교육과 관련된 경험이 없는 연구 대상자를 선정하고 개발된 프로그램에 대한 신뢰도와 두 집단 간의 차이를 알아보기 위해 SPSS 분석 도구를 선정하였다.

다섯째, 최종 프로그램의 효과성을 검증하기 위해 실험 집단과 통제 집단으로 나누어 개발된 인공지능 교육 프로그램을 적용하였다.

여섯째, 프로그램 적용 후 두 집단 간의 인공지능에 대한 태도와 인식의 차이를 알아보기 위해 김성원, 이영준(2020)이 개발한 ‘중학생의 인공지능에 대한 태도 검사 도구’ [19]를 김은경(2022)이 수정한 ‘인공지능에 대한 태도’ 검사 도구[7]와 류미영과 한선관(2017)의 ‘의미분별법을 이용한 초등학생의 인공지능에 대한 이미지’ 검사 도구를 수정하여 실시하였다 [20].

일곱째, 최종적으로 프로그램 적용 결과를 분석하고 결론 및 제언을 제시하였다.

3.2 교육 프로그램 개발 방향 및 설계

본 연구는 인공지능의 윤리적 딜레마 교육 프로그램의 필요성에 따라 부정적 요소를 최대한 배제하고 긍정적인 인공지능 윤리적 딜레마 교육을 위해 변형 Moral machine 시나리오를 개발하고자 하였다. 김은경(2022)은 반복적으로 인공지능 기기가 생명을 해하는 상황을 간접 경험하는 것이 인공지능 기기나 기술에 대한 부정적인 인식을 유발할 수 있어 초등학생의 인공지능에 대한 태도에 부정적인 영향을 미칠 수 있다

고 지적했다[7]. 안정현 외(2023)는 Moral machine을 활용한 인공지능 수업이 학생들의 인공지능 인식에 미치는 영향에 관한 연구가 부족하다고 언급하며 미디어보다 학교 교육을 통해 인공지능 윤리에 대한 이해를 높여야 한다고 교육 방향을 제안했다[15]. 이에 본 연구는 선행 연구의 결과와 문헌 고찰을 고려하여 변형 Moral machine 시나리오를 개발하였다.

첫째, 초등학생의 인공지능 윤리 교육에서 Moral machine을 활용한 교육이 초등학생에게 미칠 수 있는 부정적 영향을 고려하여, 희생과 죽음을 배제하고 도움을 주는 내용으로 구성된 변형 Moral machine 시나리오를 개발하여 학습자들의 윤리적 판단력과 의사결정 능력을 강화하고자 했다.

둘째, Moral machine을 변형 Moral machine과 동일한 학습 환경을 만들어 교육적 효과를 측정하기 위해 웹 기반의 Moral machine을 인쇄물로 대체하였다.

셋째, 변형 Moral machine의 시나리오 대상 선택 기준은 Moral machine에서 노인, 어린아이, 강아지, 임산부 등 학생이 물리적으로 만날 수 있는 대상을 선택하였고, 교통 규칙 준수와 위반 등 윤리적으로 만날 수 있는 상황을 고려하여 결정하였다.

넷째, 개발된 변형 Moral machine에 대해 인공지능 교육을 연구하고 있는 교수 및 대학원생 3인과 검토를 진행하였다. 검토 결과, 딜레마에 대한 설명이 너무 길어져 가독성이 떨어진다는 의견이 많아 설명을 간략하게 수정하고 선택지에 삽화를 제시하여 직관적으로 이해할 수 있도록 수정한 후 이를 바탕으로 최종 프로그램을 산출하였다.

3.2.1 변형 Moral machine 시나리오 1

시나리오 1은 윤리적 딜레마 상황에서 무거운 짐을 들고 걷는 것에 어려움을 겪고 있는 노인과 길을 잃어 도움이 필요한 어린아이 중 누구에게 우선적으로 도움을 제공할지 선택하는 문항으로 구성하였다. 이 시나리오에서는 Moral machine의 연령별 우선순위에 대한 사회적 가치관을 반영하여 상황은 그대로 유지하면서 죽음과 희생을 배제하고, 도움을 주는 인공지능의 이미지를 강조하여 내용을 재구성하였다. Moral machine 시나리오와 변형 Moral machine 시나리오 1의 재구성 산출물은 Figure 2와 같다.

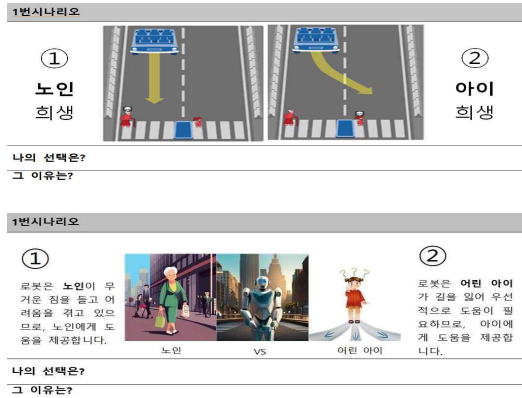


Figure 2. Modified moral machine scenario 1

3.2.2 변형 Moral machine 시나리오 2

시나리오 2는 길을 잃어 도움이 필요한 어린아이와 강아지 중 누구에게 우선적으로 도움을 제공할지 선택하는 문항으로 구성하였다.



Figure 3. Modified moral machine scenario 2

이 시나리오는 Moral machine의 인간과 애완동물 간의 우선순위에 대한 사회적 가치관을 반영하여 누구에게 우선적으로 도움을 주는 것이 옳은지 선택하는 시나리오로 설계되었다. 시나리오 2의 산출물은 Figure 3과 같다.

3.2.3 변형 Moral machine 시나리오 3

시나리오 3은 교통 법규 준수에 대한 가치관을 바탕으로 개발되었다. 이 시나리오에서는 상품을 전달하는 로봇이 다양한 상황을 고려하여 교통 규칙 위반과 교통 규칙 준수 사이에서 어떤 선택을 하는 것이 옳은지를 선택하는 문항으로 구성하였으며, 교통 상황에서의 윤리적 딜레마를 선택하도록 시나리오를 설계하였다. 시나리오 3의 산출물은 다음 Figure 4와 같다.

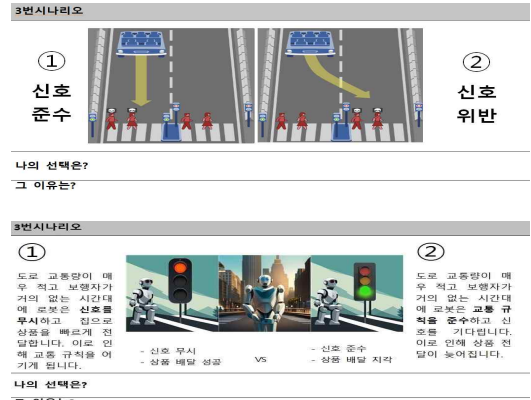


Figure 4. Modified moral machine scenario 3

3.2.4 변형 Moral machine 시나리오 4

시나리오 4는 힘들게 길을 걷고 있는 임산부와 노인 중 누구에게 우선적으로 도움을 제공할지 선택하는 윤리적 딜레마 상황을 시나리오로 개발하였다. 시나리오 4의 산출물은 다음 Figure 5와 같다.

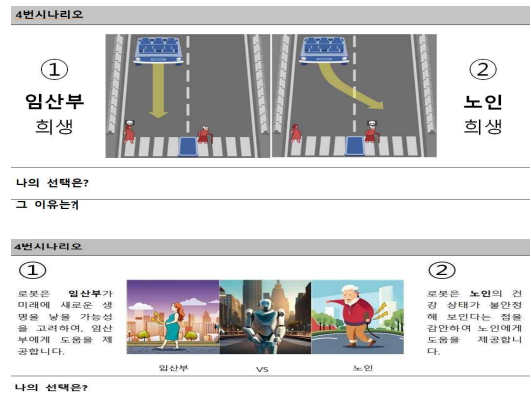


Figure 5. Modified moral machine scenario 4

본 교육 프로그램은 Moral machine을 활용한 인공지능 윤리 수업의 교육적 효과는 유지하면서 인공지능에 대한 인식과 태도에 부정적인 영향을 최소화하기 위해 도움을 주는 인공지능 로봇 시나리오로 개발하였다.

4. 연구 방법 및 교육 프로그램 적용

4.1 연구 가설

본 연구는 초등학교 6학년을 대상으로 한 인공지능 윤리 교육을 위해 Moral machine 프로그램의 대안으로써 변형 Moral machine 시나리오 내용을 개발하고, 이를 통해 변형 Moral machine이 초등학교 학생의 인공지능 윤리 교육에 미치는 영향을 알아보는 것이다. 김유상(2023)은 변형 Moral machine 활용 가능성에 관해 연구를 진행하였다. 그러나 연구 대상자가 8명으로 효과성을 검증하기에 부족했고, 포커스 그룹 인터뷰(Focus Group Interview, FGI) 연구 방법을 사용하여 객관적인 검증이 어려웠다[17]. 또한, 변형 Moral machine 선택지 중 하나는 생명을 구하는 것이었지만, 다른 선택지는 여전히 희생을 포함하고 있어 한계가 있었다. 따라서 이번 연구에서는 다음과 같은 두 가지 연구 가설을 설정하였다.

가설 1. 변형 Moral machine이 죽음과 희생을 배제하면서 도움을 주는 긍정적인 상황을 제시함으로써 초등학교 학생들의 인공지능에 대한 태도 및 인식에 긍정적인 변화를 불러올 것이다.

가설 2. 인공지능 윤리 교육에 변형 Moral machine을 활용함으로써 초등학교 학생들은 인공지능에 대한 도덕적인 가치를 다양한 측면에서 고려하게 되며, 결과적으로 인공지능에 대한 정의에 유의미한 영향을 미칠 것이다.

4.2 연구 대상

본 연구는 경기도 소재의 초등학교에 재학 중인 6학년 학생 46명을 대상으로 하였다. 연구 대상자 전원은 인공지능 교육을 받은 경험이 있지만, 인공지능 윤리 교육과 관련된 경험은 없는 상태였다. 본 교육 프로그램에 참여하는 학생은 도덕적인 사고 능력을 갖춰야 하며, 추상적이고 논리적으로 문제를 해석할 수 있어야 한다. 아동의 도덕적 사고 발달에 관해 연구한 대표적 이론가인 피아제(Piaget)에 따르면, 아동의 도덕적 발달

은 타율적 도덕성(Heteronomous Morality)과 자율적 도덕성(Autonomous Morality) 두 주요 단계로 나뉜다. 일반적으로 10세 이상의 학생은 자율적 도덕성의 단계에 있다고 여겨지며, 이 시기의 학생은 도덕적인 판단 시 행동 결과뿐만 아니라 동기를 고려할 수 있으며, 규칙이 상대적이라는 개념을 이해한다[14]. 따라서 초등학교 학생을 대상으로 한 인공지능 윤리 교육 프로그램을 개발하고 적용하기 위해, 도덕적인 판단을 형성하면서도 개인의 동기를 고려하며 사회 질서를 유지할 수 있는 초등학교 6학년 학생들을 대상으로 연구를 진행하였다. 연구에 참여한 대상은 다음 Table 1과 같다.

Table 1. Research participant information

Division	number of classes	N	percentage
Experimental group	1	23	50
Control group	1	23	50
Total	2	46	100

4.3 분석 도구 및 방법

인공지능 윤리 교육을 위한 윤리적 딜레마 시나리오 내용을 개발하고 효과성을 검증하기 위해 김성원, 이영준(2020)이 개발한 ‘중학생의 인공지능에 대한 태도 검사 도구’ [19]를 김은경(2022)이 수정한 ‘인공지능에 대한 태도’ 검사 도구[7]와 류미영과 한선관(2017)의 ‘의미분별법을 이용한 초등학교 학생의 인공지능에 대한 이미지’ 검사 도구를 사용하였다[20]. 이는 Moral machine을 활용한 인공지능 윤리 교육이 어떠한 영향을 미치는지에 대한 사전, 사후를 비교하기 위한 도구로 본 연구에서는 Moral machine과 변형 Moral machine을 실험 집단과 통제 집단으로 나누어 수업 후 개발된 인공지능 윤리 교육 시나리오의 효과성 검증에 적용하였다.

본 검사 도구는 인공지능에 대한 태도와 관련된 문항 17개의 타당도 검사를 한 후 신뢰도를 저해하는 문항을 제거하고 15개 문항을 3개의 하위 요소인 인공지능과 감정적 교류, 인공지능과 생활, 인공지능의 사회적 영향으로 세분화하였다.

각 문항은 5점 리커트 척도를 사용하였고 각 항목에 대한 문항 신뢰도 검사와 문항 구성은 다음 Table 2와 같다. 일반적으로 Cronbach's α 가 0.6 이상이면 일정

수준의 신뢰성을 가지며 분석 활용에 적합하다고 판단한다. Cronbach's α 알파 값이 높을수록 해당 도구의 결과에 대해 매우 높은 신뢰성을 가진다. 본 연구에서의 Cronbach's α 는 인공지능에 대한 태도 문항 결과 하위요소별로 인공지능과 감정적 교류는 .732, 인공지능과 생활은 .652, 인공지능의 사회적 영향은 .821의 값으로 나타났기 때문에 모두 알파 계수의 값이 .6 이상으로 신뢰도가 검증되었다.

Table 2. Test tool of attitude toward AI

factor	Sub items	N	Cronbach's Alpha
Emotional exchange with AI	<ul style="list-style-type: none"> • Conversation with AI is comfortable. • AI with emotions is a friend. • AI with emotions is comfortable. 	3	.732
Life with AI	<ul style="list-style-type: none"> • AI works well. • AI is useful. • AI processes tasks quickly. • You can communicate fluently with AI like a friend. • I can understand what AI is saying. 	5	.652
Social influence of AI	<ul style="list-style-type: none"> • I am nervous about using AI. • The decisions made by AI are terrible. • I feel nervous when I stand in front of AI. • I think I will become more sensitive when I talk to AI. • If we rely on AI, bad things are likely to happen. • As AI develops, ethical issues are likely to arise. • Strong means to control AI are needed. 	7	.821
Total	15	15	

인공지능에 대한 이미지는 이미지를 인식으로 수정 후 관련된 문항 23개 중 이번 연구와 연관성이 있는 15개 문항의 신뢰도를 검증하였다. 15개 문항은 2개의 하위 요소 긍정적 인식, 부정적 인식으로 세분화하여 구성하였다. 각 문항은 5점 리커트 척도를 사용하였고 각 항목에 대한 문항 신뢰도 검사와 문항 구성은 다음 Table 3과 같다.

인공지능에 대한 인식 문항의 Cronbach's α 역시 하위요소별로 긍정적인 인식은 .834, 부정적인 인식은 .635로 나타났기 때문에 모두 알파 계수의 값이 .6 이상

으로 신뢰도가 검증되었다.

Table 3. Test tool of perception toward AI

factor	Sub items	N	Cronbach's Alpha
Positive perception	<ul style="list-style-type: none"> • Smart • Friendly • Good • Kind • Creative • Human • Lovely • Varies • Thanks • Safe 	10	.834
Negative perception	<ul style="list-style-type: none"> • Scary • Alone • Attack • Worried • Depressed 	5	.635
Total	15	15	

4.4 교육 프로그램 적용 결과물

본 연구에서 경기도 소재의 초등학교 6학년 학생을 대상으로 1차시 인공지능 윤리 이론 수업 후 개발된 변형 Moral machine의 효과성을 분석하기 위해 두 집단으로 나누어 실험하였다. 총 2차시에 걸친 차시별 세부 내용은 다음 Table 4와 같다.


Table 4. Contents of the class

Session	Subject	Contents
1	Understanding AI ethics	<ul style="list-style-type: none"> - AI Ethics definition explained - AI Ethical issues - AI technology's impact on society and ethical considerations
2	AI Ethics practice	<ul style="list-style-type: none"> - AI ethical dilemma case - AI Ethics Scenario Practice

본 연구에서는 인공지능 윤리 교육을 총 2차시로 구성하였다. 1차시는 인공지능 윤리의 개념과 문제점, 인공지능 기술이 사회에 미치는 영향 및 윤리적 고려 사항에 대한 수업을 진행하여 인공지능 소양 교육에서 다뤄야 할 윤리 내용으로 구성하였다. 2차시는 인공지능 윤리 실습을 주제로 실제 인공지능 윤리적 딜레마

사례를 살펴보고 실험 집단과 통제 집단으로 나누어 개발된 변형 Moral machine과 기존 Moral machine을 활용한 수업을 하였다. 이를 통한 총 2차시 수업 산출물과 수업 장면은 Table 5과 같다.

Table 5. Students' answer to Modified moral machine and Class scenes

나의 선택은 ① 그 이유는 모든 것의 정해진 법칙을 잘 알고 있기 때문이다. 기술의 발달이 기술의 발달이	나의 선택은 ② 그 이유는 모든 것의 정해진 법칙을 잘 알고 있기 때문이다. 기술의 발달이 기술의 발달이
Students' answer to scenario no.1	
나의 선택은 ① 그 이유는... 기술의 발달이 기술의 발달이	나의 선택은 ② 그 이유는... 기술의 발달이 기술의 발달이
Students' answer to scenario no.2	
나의 선택은 ① 그 이유는... 기술의 발달이 기술의 발달이	나의 선택은 ② 그 이유는... 기술의 발달이 기술의 발달이
Students' answer to scenario no.3	
나의 선택은 ① 그 이유는... 기술의 발달이 기술의 발달이	나의 선택은 ② 그 이유는... 기술의 발달이 기술의 발달이
Students' answer to scenario no.4	
	
Class scenes	

5. 연구 결과

변형 Moral machine을 활용한 교육의 효과를 검증하기 위해 총 46명을 실험 집단(n=23)과 통제 집단(n=23)으로 나누어 인공지능에 대한 태도와 인식 변화를 측정하고 독립 표본 t-검정을 통해 분석하였다.

5.1 인공지능에 대한 태도 검사 결과

본 연구는 변형 Moral machine을 활용한 교육 방법이 초등학생들의 인공지능에 대한 태도에 미치는 영향을 알아보기 위해 하위 요소인 인공지능과 감정적 교류, 인공지능과 생활, 인공지능의 사회적 영향을 독립

표본 t-검정을 통해 검증하였다. 유의 수준은 p<.05로 분석하였다. 그 결과 Table 6과 같이 두 교육 방법 간의 유의미한 차이가 나타났다.

인공지능에 대한 태도 검사는 변형 Moral machine과 Moral machine을 비교한 결과 유의확률 p<.05 수준에서 두 하위 요소인 인공지능과 감정적 교류 .041, 인공지능과 생활 .037로 나타나 통계적으로 유의미한 결과를 확인하였다. 그러나, 인공지능의 사회적 영향은 평균에서 크게 차이를 보이지 않는 것으로 나타났으며, 유의확률 p값이 .369로 0.05 이상으로 나타나 인공지능의 사회적 영향에 차이가 없는 것으로 나타났다. 이러한 결과는 변형 Moral machine을 활용한 인공지능 수업이 인공지능에 대한 태도에 긍정적인 영향을 미친다는 것을 시사한다. 다만, 사회적인 영향에 대한 우려가 여전히 존재한다는 점이 확인되었다. 이는 미래에 인공지능 로봇과 함께 살아가야 하는 상황에서의 우려적인 시각이 여전히 존재함을 객관적으로 나타내고 있다.

Table 6. Factor analysis of attitude toward AI

Factor		N	M	SD	t	p
Division						
Emotional exchange with AI	Modified moral machine	23	3.23	1.00	2.102	0.041
	Moral machine	23	2.61	1.01		
Life with AI	Modified moral machine	23	2.77	0.81	2.149	0.037
	Moral machine	23	2.32	0.56		
Social influence of AI	Modified moral machine	23	2.82	0.83	-0.908	0.369
	Moral machine	23	3.04	0.82		

5.2 인공지능에 대한 인식 검사 결과

변형 Moral machine을 활용한 교육 방법이 초등학생들의 인공지능에 대한 인식에 미치는 영향을 알아보기 위해 하위 요소인 긍정적 인식, 부정적 인식을 독립 표본 t-검정을 통해 검증한 결과, Table 7과 같이 변형 Moral machine을 활용한 교육이 인공지능에 대한 인식 형성에 긍정적인 영향을 미치고 있음을 확인할 수 있었다.

Table 7. Factor analysis of perception toward AI

Factor		N	M	SD	t	p
Division						
Positive perception	Modified moral machine	23	2.98	0.78	2.067	0.045
	Moral machine	23	2.53	0.69		
Negative perception	Modified moral machine	23	2.71	0.82	-2.257	0.029
	Moral machine	23	3.22	0.68		

인공지능에 대한 인식 검사는 변형 Moral machine과 Moral machine을 비교한 결과 유의확률 $p < .05$ 수준에서 두 하위 요소인 긍정적 인식 .045, 부정적 인식 .029로 나타나 인공지능에 대한 인식 검사 결과 모두 통계적으로 유의미한 결과를 확인하였다. 긍정적 인식과 부정적 인식에 대한 통계적으로 유의미한 결과는 변형 Moral machine이 인공지능에 대한 인식을 형성하는 데에 효과적인 교육 도구로 활용될 수 있음을 알 수 있다.

5.3 인공지능에 대한 정의 검사 결과

본 연구에서 실험 집단, 통제 집단은 인공지능 윤리 교육 후 검사지에 “인공지능은 ○○이다.” 라는 질문을 개방형 설문으로 작성하였고, 작성 내용은 워드 클라우드를 이용하여 두 집단 간의 차이를 비교 분석하였다. 결과적으로 두 집단 간의 유의미한 차이를 확인하였다. 인공지능 윤리 교육 후 인공지능에 대한 학생들의 응답은 다음 Figure 6과 같다.



Figure 6. Students' answer to Modified moral machine vs Moral machine

인공지능에 대한 정의 응답은 두 집단의 설문 결과

총 51건이 수집되었다, 이 중 2건 이상 의견을 제시한 학생의 응답과 응답을 하지 않은 경우도 포함되었다. 구체적으로 살펴보면, 변형 Moral machine은 ‘필요하다.’ 라는 의견을 응답한 경우가 5건, ‘편리하다.’ 를 응답한 경우가 4건, ‘착하다.’, ‘친구다.’, ‘똑똑하다.’, ‘무섭다.’ 를 응답한 경우가 각 2건, 그 외 ‘신기하다.’, ‘최첨단이다.’, ‘못생겼다.’, ‘인간적이다.’, ‘위험하다.’, ‘미래다.’, ‘낯설다.’, ‘불안하다.’ 등의 응답이 각 1건, 응답하지 않은 경우가 1건으로 나타났다.

Moral machine의 응답을 ‘무섭다.’ 라고 응답한 경우가 7건으로 가장 높았으며, ‘편리하다.’, ‘필요하다.’ 라고 응답한 경우는 각 3건, ‘똑똑하다’ 라는 응답 2건으로 나타났다. 그 외 ‘앞서간다.’, ‘재미있다.’, ‘위험하다.’, ‘위험적이다.’, ‘정복할 기계’, ‘발전한다.’, ‘멋지다.’, ‘철 덩어리이다.’, ‘사람이 만든 기계이다.’, ‘순종적이다.’ 등의 응답이 1건으로 다양하게 나타났다.

변형 Moral machine을 학습한 학생들이 제시한 내용 중에서 일부 발췌한 응답은 다음과 같다.

- “인공지능은 꼭 필요한 존재이다.”
- “인공지능은 편리한 기계이다.”
- “인공지능은 똑똑한 기계이다.”
- “인공지능은 편리하지만, 위험한 존재이다.”
- “인공지능은 친구 같은 존재이다.”
- “인공지능은 최첨단 기계이다.”
- “인공지능은 착한 로봇이다.”

6. 결론 및 제언

인공지능 기술의 발전은 사회 전반에 영향을 미치고 있으며, 이에 따라 교육 현장에서 인공지능 윤리 교육의 중요성이 증대되고 있음을 확인했다. 현재 직접적인 인공지능 윤리 교육 실습 자료가 부족한 상황에서 Moral machine은 인공지능 윤리 교육에 활발하게 활용되고 있다. 그러나 Moral machine은 교육 목적으로 개발되지 않았기 때문에, 학습자들이 죽음과 희생을 반복적으로 경험하면서 부정적 태도를 형성할 우려가 있다.

이에 본 연구에서는 이러한 한계를 극복하고 긍정적 학습 경험을 제시하기 위해, 죽음과 희생을 배제하고 도움의 우선순위를 선택하는 새로운 변형 Moral machine 교육 프로그램을 개발하였다. 또한 본 연구에서는 초등학생을 대상으로 Moral machine을 활용한 인공

지능 윤리 교육과 변형 Moral machine을 활용한 인공지능 윤리 교육이 인공지능에 대한 태도와 인식에 어떠한 영향을 미치는지를 비교 분석하였다. 연구를 통해 도출한 결과는 다음과 같다.

첫째, 비교 집단 간의 교육 방법에 따라 차이가 있었다. Moral machine을 활용한 교육은 인공지능 기술에 대한 부정적 영향을 주는 반면, 변형 Moral machine을 활용한 교육은 긍정적인 결과를 확인했다. 이는 변형 Moral machine 교육 프로그램이 죽음과 희생을 배제하고 도움을 주는 긍정적인 상황을 제시함으로써, 학생들의 긍정적 태도 및 인식 형성에 기여한 것으로 판단된다. 다만, 모든 영역에서 유의미한 향상이 관찰된 것은 아니다. 인공지능에 대한 태도 검사에서 하위 영역인 인공지능 사회적 영향은 유의확률 0.05 이상으로, 통계적으로 유의미한 향상이 있다고 보기 어렵다. 따라서 해당 영역을 보완할 수 있는 추가적인 연구가 필요하다고 판단된다.

둘째, 변형 Moral Machine을 활용한 수업에서 학습자들은 인공지능에 대한 도덕적인 가치를 긍정적으로 고려했다. 결과적으로, 학습자들은 인공지능을 필요하고 편리한 존재로 정의하는 긍정적 태도를 보였다.

이러한 연구를 바탕으로 후속 연구를 위한 제언은 다음과 같다.

첫째, 인공지능 윤리 교육에서 교육 방법의 선택이 중요하며, 특히 초등학생을 대상으로 하는 인공지능 윤리 교육은 학생들의 감정과 심리에 민감하게 영향을 미칠 수 있음을 시사했다. 향후 연구에서는 다양한 교육 방법의 실험과 평가가 필요하며, 학생들의 긍정적 태도 형성을 위한 최적의 교육 전략을 찾는 것이 중요하다.

둘째, 본 연구는 초등학생을 대상으로 한 인공지능 윤리 교육의 중요성을 강조하며, 변형 Moral machine을 활용한 교육을 통해 부정적 영향을 최소화하고 긍정적인 인공지능에 대한 태도와 인식을 형성할 수 있음을 확인했다. 변형 Moral machine의 추가적인 내용 개발과 다양한 연령 그룹 및 상황에 대한 연구가 필요하다.

본 연구에 개발된 인공지능 윤리 교육 프로그램이 더 다양한 교육 프로그램을 개발하는 데 기초 자료로 활용되길 기대한다.

참고문헌

- [1] Kim, E., Lee, Y. (2023). Development of AI Ethics Dilemma to Enhance AI Ethical Competence. *The Journal of Korean Association of Computer Education*, 26(5), 31-42. DOI : 10.32431/kace.2023.26.5.003
- [2] Jun, S. (2023). Development and effectiveness of Artificial Intelligence ethics education program based on CAI model. *The Journal of Korean Association of Computer Education*, 26(1) : 23-31. DOI : 10.32431/kace.2023.26.1.003
- [3] The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. (2016). *Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1*. IEEE. http://standards.ieee.org/develop/in-dconn/ec/autonomous_systems.html
- [4] Kim, J., Lee, C. (2021). Development of STEAM Program for Artificial Intelligence Ethic Education for Elementary School Student. *The Journal of Korean Association of Artificial Intelligence Education*, 21(1), 21- 28. DOI : 10.52618/aied.2021.2.1.2
- [5] National Human Rights Commission. (2022). *Human Rights Guidelines on the development and use of artificial intelligence*. <https://www.humanrights.go.kr/base/board/read?boardManagementNo=24&boardNo=7608423&menuLevel=3&menuNo=91>
- [6] Ministry of Education. (2022). *2022 National Curriculum Draft - Elementary School*.
- [7] Kim, E. (2022). *The Effect of Artificial Intelligence Ethics Education Using Moral Machine on Elementary School Students' Attitudes and Images toward Artificial Intelligence*. Master's thesis, Graduate School of Korea National University of Education.
- [8] Jun, S. (2022). A case study of activity-oriented AI ethics education for liberal arts education. *Korean Association of Artificial Intelligence Education Transactions*, 3(3), 30-39. DOI : 10.52618/aied.2022.3.1.4
- [9] Berkman Klein Center for Internet & Society (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*
- [10] AI4ALL. (2020 February 9). <https://ai-4-all.org/>
- [11] Ministry of Education (2020). *Information Department 2015 Revised Curriculum-Artificial Intelligence Basics*, National Curriculum Information Center.
- [12] Ministry of Education & Korea Foundation for

Science and Creativity (2020). *AI class that meets at school (content system by school level)*

- [13] Jang, Y., Choi, S., Cho, H. & Kim, H. (2022) Development and Application of Modular Artificial Intelligence Ethics Education Program for Elementary and Middle School Students. *The Journal of Korean Association of Artificial Intelligence Education*, 23(5), 1-14. DOI : 10.32431/kace.2022.25.5.001
- [14] Song, J., Jeon, Y. (2023). Development and Application of Artificial Intelligence Ethics Education Program for Elementary School Students based on Hans Jonas' Responsibility Ethics. *The Journal of Korean Association of Artificial Intelligence Education*, 26(2), 29- 39. DOI : 10.32431/kace.2023.26.2.004
- [15] Ahn, J., Park, H. (2023). Changes in elementary school students' perception of the ethical autonomy of artificial intelligence -Focusing on classes using Moral machine. *The Journal of Korean Association of Computer Education, Academic Presentation Conference Papers*, 27(1), 153-156. DOI : 10.32431/kace.2024.27.1.008
- [16] Kim, E., & Lee, Y. (2022). The Influence of Artificial Intelligence Ethics Education Using Moral Machine on Elementary School Students' Perception of Artificial Intelligence. *The Journal of Korean Association of Computer Education*, 23(3), 1-8. DOI : 10.32431/kace.2022.25.3.001
- [17] Kim, Y. (2023). *A Study on the Validity of Moral Machine as a Teaching Tool for AI Ethics Education*. Graduate Program in Educational Methods Graduate School of Education Gwangju National University of Education.
- [18] Kim, E., Kim, S., Lee, Y. (2021). Moral Machine Utilization Plan in Artificial Intelligence Ethics Education for Elementary School Students. *The Journal of Korean Association of Computer Education*, 25(2(A) : 35-36. DOI : 10.32431/kace.2022.25.3.001
- [19] Kim, S., & Lee, Y (2020). Development of Test Tool of Attitude toward Artificial Intelligence for Middle School Students. *The Journal of Korean Association of Computer Education*, 23(3), 17-30. DOI : 10.32431/kace.2020.23.3.003
- [20] Ryu, M., & Han, S (2017). Image Analysis of Artificial Intelligence Recognized by Elementary School Students. *The Journal of Korean Association of Computer Education*, 21(5), DOI : 527-535. 10.14352/jkaie.2017.21.5.527



홍 은 정

2024년 건국대학교
정보컴퓨터교육전공(석사)

관심분야: SW교육, 인공지능교육, 인공지능 윤리교육, 정보교육
E-Mail: eunjung8141@gmail.com



오 경 선

2016년 성균관대학교 컴퓨터교육
전공(박사)

2017년 ~ 2019년 단국대학교강의 전담조교수
2019년 ~현재 건국대학교 상허교양대학 조교수
관심분야: SW교육, 프로그래밍교육, 컴퓨팅적사고, 인공지능교육,
데이터과학
E-Mail: skyal@konkuk.ac.kr