컴퓨터교육학회 논문지 2024년 제27권 제3호
https://doi.org/10.32431/kace.2024.27.3.014

인공지능 윤리감수성 검사 도구 개발 연구*

A Study on the Development of AI Ethics Sensitivity Scale

김은경† · 이영준††
Eungyeong Kim† · Youngjun Lee††

요 약

인공지능 윤리교육은 실제 윤리적 행동을 이끌어 낼 수 있는 교육이 되어야 하나 현재 인공지능 윤리교육 연구에서 윤리적 행동 변화에 초점을 둔 연구는 찾아보기 어렵다. 이는 윤리적 행동과 관련된 요소를 측정할 수 있는 검사도구가 부재하기 때문으로 생각된다. 이에 본 연구에서는 Rest의 도덕적 행동에 관련된 4구성 요소 이론을 기반으로 하여, 윤리적 행동을 위한 주요 요소인 도덕감수성을 인공지능 관련 상황에서 측정할 수 있는 인공지능 윤리감수성 검사 도구를 개발하였다. 인공지능 윤리감수성의 하위요인은 상황지각, 결과지각, 공감지각, 책임지각으로 설정하였으며, 인공지능 윤리 갈등상황을 나타내는 딜레마를 읽고, 인공지능 윤리 가치 문항과 다른 가치 문항에 대한 중요도를 7점 척도로 평정하는 형식으로 구성 하였다. 검사 도구에 대한 타당도는 전문가 대상 타당도 조사와 중학생 대상 예비검사를 통해 확보하였다. 본 연구에서 개발된 검사 도구는 인공지능 윤리교육 연구의 방향성을 윤리적 행동으로 확대하는데 기여할 것이다.

주제어: 인공지능 윤리, 인공지능 윤리감수성, 레스트, 도덕감수성, 인공지능 윤리감수성 검사도구

ABSTRACT

AI ethics education should be an education that can lead to actual ethical behavior, but it is difficult to find a study that focuses on changes in ethical behavior in current AI ethics education research. This is thought to be due to the lack of a test tool that can measure factors related to ethical behavior. Therefore, in this study, based on the four-component theory related to Rest's moral behavior, an AI ethics sensitivity test tool that can measure moral sensitivity, a major element for ethical behavior, in artificial intelligence-related situations was developed. The sub-factors of AI ethics sensitivity were set as situation perception, result perception, empathy perception, and responsibility perception, and the dilemma representing AI ethics conflict situations was read, and the importance of AI ethics value questions and other value items was evaluated on a 7-point scale. The validity of the test tool was secured through a validity survey for experts and a preliminary test for middle school students. The test tool developed in this study will contribute to expanding the direction of AI ethics education research into ethical behavior.

Keywords: AI Ethics, AI Ethical Sensitivity, Rest, Moral Sensitivity, AI Ethics Sensitivity Scale

1. 서론

인공지능이 사회에 미치는 영향력이 점점 커져감에 따라 인공지능 윤리 문제는 이제 일부 인공지능 전문가 집단의 문제가 아닌 사회 구성원 모두의 문제가 되었다. 세계 각국의 정부와 여러 기관은 인공지능 개발 및 활용 과정에서 지켜야 할 윤리 기준을 마련하고 있으며, 교육계 역시 이에 대응하기 위한 교육과정을 개발하고 다양한 인공지능 윤리교육을 실시하고 있다[1, 2]. 그러나 McNarama(2018) 연구에 따르면 윤리 기준을 제시하고 명시적으로 따르게 하는 것만으로는 학습자의 의사결정에 영향을 미치지 못하는 것으로 나

†정 회 원: 한국교원대학교 대학원 초등컴퓨터교육 박사과정
††종신회원: 한국교원대학교 컴퓨터교육과 교수(교신저자)

논문투고: 2024년 02월 29일, 심사완료: 2024년 04월 11일, 게재확정: 2024년 04월 17일

* 본 논문은 2024년 한국컴퓨터교육학회 동계학술대회에서 "인공지능 윤리 감수성 평가 도구 개발 필요성에 대한 고찰"의 제목으로 발표된 논문을 확장한 것임.

타났다[3]. 이는 인공지능 윤리 교육이 단순히 윤리 기준을 전달하는 것을 넘어, 윤리적 행동을 유도할 수 있는 방식으로 진행되어야 함을 시사한다.

인공지능분야의 석학인 제프리 힌튼 교수는 최근 한 방송국과의 인터뷰에서 인공지능의 위험성을 경고하며, 생화학무기 사용을 금지한 제네바협약과 같이 인공지능이 인간에게 위협이 되는 방향으로 사용되는 것을 제제 할 수 있는 협약이 필요하다고 주장하였다. 그러나 그것은 현실적으로 불가능할 것이라고 이야기하며, 그것이 가능해지려면 대중이 인공지능에 대한 경각심을 가져야 한다고 하였다[4].

인공지능 기술이나 정책을 개발하는 것은 소수지만 이들은 결국 대중의 여론과 선호에 영향을 받는다. 따라서 인공지능이 윤리적으로 개발되고 사용되게 하려면 대중이 인공지능을 윤리적 관점으로 바라보는 시각을 가져야 하고, 더 나아가 인공지능 윤리문제 상황에 직면했을 때 윤리적 행동을 할 수 있어야 한다. 따라서 인공지능 윤리교육 또한 이러한 관점에서 이루어져야 한다.

다행히도 교육계에서는 현재 윤리적 딜레마를 활용한 토론 수업과 같이 단순 윤리기준 전달이 아닌 다양한 형태의 인공지능 윤리교육이 진행되고 있다[5-8]. 그러나 이러한 교육이 윤리적 행동에 영향을 주는지에 대한 연구는 미흡하다. 현재 진행된 대다수의 인공지능 윤리교육의 효과성 연구에서는 인공지능 윤리의식이나 인공지능에 대한 인식만을 살펴보고 있다[9,10]. 이는 아직 인공지능 윤리 분야에서 윤리적 행동을 가져오는 요인과 관련된 검사 도구가 개발되어 있지 않기 때문으로 짐작된다. 따라서 인공지능 윤리교육이 윤리적 행동 변화를 가져오는지를 평가하기 위해서는 윤리적 행동과 관련된 요소를 측정할 수 있는 새로운 검사 도구의 개발이 필요하다.

윤리적 행동과 관련된 연구는 다양한 학문에서 계속 진행되어 왔으나, 인간이 어떠한 심리과정을 거쳐 윤리적 행동을 하게 되는지를 밝혀내는 것은 쉽지 않았다. 윤리적 행동의 심리과정을 밝혀보려는 노력 속에서 Rest(1983)는 윤리적 행동에 영향을 주는 심리요소로 4가지 구성요소를 주장하였다. 그중 제 1구성요소인 도덕감수성은 '상황에 대한 인식이나 해석'의 영역으로, 윤리적 행동을 위해 선행되어야 할 중요한 요소이다[11,12]. 이에 본 연구에서는 인공지능 윤리 문제 상황에서의 도덕감수성을 평가할 수 있는 도구를 개발하고자 하였다.

2. 이론적 배경

2.1 인공지능 윤리

이제영 외(2019)는 인공지능과 관련된 모든 이해관계자들이 지켜야 할 보편적인 사회 규범과 기술을 인공지능 윤리로 설명했으며, 허유선 외(2020)는 인공지능 기술의 개발, 연구, 적용을 포함한 모든 과정에서 인공지능에 관련된 규범적인 이슈를 다루는 것으로 정의했다[13,14].

이러한 인공지능 윤리는 기계의 책임을 강조하는 기계 중심적인 관점에서 출발했다. 아시모프의 로봇 3원칙이나 후쿠오카 세계 로봇 선언 같은 기준들이 이를 잘 나타낸다. 그러나 시간이 흐르면서 이러한 접근은 인공지능을 개발, 배포, 사용하는 인간의 책임을 강조하는 방향으로 진화했다. 2007년의 유럽로봇연구의 로봇 연구 원칙은 이러한 변화를 상징적으로 보여준다[15]. 2021년 유네스코(UNESCO)의 193개 회원국이 채택한 첫 번째 글로벌 인공지능 윤리 협약에서도 인공지능의 중심에 인간이 있다는 것을 확인할 수 있다[16].

한국에서는 2020년에 발표된 '사람 중심의 인공지능 윤리 기준'을 통해 이러한 인공지능 윤리의 변화를 명확히 하였다. 이 기준은 인간성을 존중하는 인공지능을 목표로 하는 3대 원칙과 10가지 핵심 요소를 도입하였고 이를 인공지능 윤리 문제에 관한 숙의의 시작점으로 삼겠다고 밝혔다[17]. 이에 따라 이후 국가 수준에서 발표하는 인공지능 윤리 관련 지침이나 정책뿐 아니라 인공지능 윤리교육 교재까지도 '사람이 중심이 되는「인공지능 윤리기준」'을 반영하여 제시되고 있다. 관련 자료 목록은 Table 1과 같다[18]. 이에 본 연구에서도 사람이 중심이 되는「인공지능 윤리기준」의 3대 원칙을 바탕으로 인공지능 윤리 딜레마 일화를 개발하였다.

Table 1. Content based on Human-centered AI ethical standards.

Year	Contents	Institution
2020.12	Human-centered 「AI ethical standards」	Joint ministries
2021.5	Reliable AI Realization Strategy for Human-centered AI [Plan]	Joint ministries
2021.5	AI Personal Information Protection Autonomous Checklist	Personal Information Protection Committee
2021.12	Development of AI Ethics Policy for	KISDI

Year	Contents	Institution
	Realizing Human-centered Artificial Intelligence	
2022.8	「Ethical Principles of AI in Education」 to support Human growth	Ministry of Education
2023.2	AI Ethics Education Textbook	KISDI

2.2 인공지능 윤리교육 연구 현황

인공지능 윤리에 관한 관심과 필요성이 증대됨에 따라 인공지능 윤리교육 관련 연구 또한 학계에서 활발하게 진행되고 있다. RISS에서 2020년에서 2023년 KCI에 등재된 논문 중 '인공지능 윤리교육' 이라는 키워드로 검색된 논문은 총 258건이었고 이 중 실제 인공지능 윤리교육 프로그램을 개발하고 그 효과성을 검증한 논문은 9건 있었다. 각 논문에서 효과성을 검증한 항목은 Table 2와 같다.

대부분의 연구에서는 인공지능 윤리의식이나 인공지능 윤리와 직접적으로 관계가 없는 것을 검증하고 있는 것을 확인해 볼 수 있었다. 이영호(2021), 박지민(2021)의 연구와 같이 인공지능 교육의 하위 요소로서 인공지능 윤리 교육을 다룬 연구의 경우에는 창의적 문제 해결력과 같이 다소 인공지능 윤리와 관련성이 떨어지는 항목의 효과성을 검증하고 있었다[10,19]. 정다이(2023)의 연구에서는 자체 개발한 검사 문항으로 도덕적 사고력과 AI 리터러시의 효과성을 검증하고 있었다. 이 검사 문항은 인지・정의・행동으로 영역을 나누어 검사하고 있다는 점에서 타 연구에서 인공지능 윤리 의식을 검사했던 것과 차별화하였으나 행동 영역에 관한 질문이 윤리적 행동에 대한 것이 아닌 인공지능 윤리 학습을 계속할 것인가에 대한 것이어서 윤리적 행동의 실천과는 관계가 멀다. 그리고 해당 검사 도구의 경우 자체 개발 검사 도구로 신뢰성과 타당성이 확보되지 않았다. 이에 대해 정다이(2023)는 인공지능 윤리 프로그램의 효과성을 입증하기 위해 보다 신뢰성 있는 측정 도구의 필요성을 지적하였다[20].

Table 2. A Study on the Effectiveness of AI Ethics Education

A Study	Effectiveness Validation Items
Development and effectiveness of AI ethics education program based on CAI model, (Jun,2023)	AI ethical awareness
The Influence of AI Ethics Education Using	Attitudes toward AI,
Moral Machine on Elementary School Students' Perception of Artificial Intelligence, (Kim, 2022)	Image toward AI
Development and Application of Ethics Education STEAM Projects using DeepFake Apps, (Hwang, 2021)	Consciousness of information and communication ethics
Validation of the Effectiveness of AI Ethics Education Program: focusing on application to elementary moral education, (Lee, 2023)	AI ethical awareness
Development and effectiveness analysis of AI STEAM education program, (Lee, 2021)	AI technology attitude, Creative problem solving abilities
Development and Application of AI Ethics Education Program for Elementary School Students based on Hans Jonas' Responsibility Ethics, (Song, 2023)	AI ethical awareness
Development and Implementation of an Elementary-school AI Ethics Education Program, (Jeong, 2023)	Moral thinking, AI literacy (Self-development)
Development and Application of Modular Artificial Intelligence Ethics Education Program for Elementary and Middle School Students, (Jang, 2022)	Perception of the need for AI ethics, Interest in AI ethics, Changing Perspectives of AI ethics
Effects of the Project-based AI Education Program on AI Ethical Consciousness and Creative Problem-Solving Skills using Flipped Learning, (Park, 2021)	AI ethical Awareness, Creative Problem Solving Abilities

2.3 도덕감수성

도덕성 발달이론을 처음 이야기한 Kohlberg는 인지적인 측면을 강조하여 도덕성 개념을 도덕적 사고(판단) 능력으로 정의하였다. 서양에서는 소크라테스가, 동양에서는 왕양명(王陽明)이 지행합일을 주창하였고, 그 이후로 도덕적 지식이나 사고와 도덕적 행동과의 관계에 대해서는 수많은 논쟁이 있었다. 그러나 지식만으로 행동을 설명하는데 한계가 있다는 것은 누구나 동의하는 주지의 사실이다[21].

Rest에 의하면 인간이 도덕적 행동을 하는 데에는 4가지 심리 과정을 걸친다고 한다. 그는 도덕적 행동을 구성하는 요소로 도덕감수성, 도덕판단력, 도덕동기화, 도덕품성화를 이야기하였다. 각 요소에 대한 정의는 Table 3과 같다.

Table 3. Rest's Four-Component Model of Morality

Four-Component	Concept
Moral	The ability to interpret situations and

Four-Component	Concept
sensitivity	understand how others might be affected
Moral judgment	The ability to judge which action is right and will lead to the desired outcome
Moral motivation	The ability to formulate actions to achieve the desired outcome
Moral character	The ability to follow through on intentions

도덕적 감수성은 Rest의 도덕적 행동과 관련된 4구성 요소 중 제1구성 요소로 특정 상황 속에 내포된 도덕적 이슈들을 지각하고, 상황을 해석하며, 자기 행동의 결과가 타인에 미치는 영향을 헤아릴 수 있는 능력을 말한다[11]. 그동안 심리학 분야에서 이루어진 많은 연구를 종합해보면, 사람들은 비교적 간단한 도덕적 사태를 지각・해석하는 데도 많은 어려움을 겪고 있고, 타인의 필요에 대한 감수성에서도 현저한 개인차가 발견되는데, 바로 이런 차이들은 도덕적 행동을 저해하는 요인으로 작용한다는 사실이 밝혀진 바 있다[11,22].

어떤 상황을 도덕적 사태로 지각・해석하지 못한다면 도덕적 행동은 일어날 수 없다[23]. 38명의 사람이 35분 동안 한 여성이 살해당하는 모습을 목격하고도 돕거나 신고하지 않았던 키티 제노비스(Kitty Genovese)사건은 이를 증명하는 대표적인 사건이라고 할 수 있겠다[21]. 또한 타인의 필요와 행동이 다른 사람에 미칠 영향성에 대한 심각성이 부족해도 도덕적 행동은 일어나지 않는다[12]. 이러한 도덕적 감수성은 현저한 개인차가 존재하며 연령 증가에 따라 발달하는 것으로 알려졌다[24].

2.4 도덕감수성 하위요소

Rest는 도덕감수성의 하위요소가 상황에 관련된 사람이 누구이고 어떠한 상황인지, 가능한 행동이 어떤 것인지 추측하고 그 행동이 다른 사람에게 어떠한 영향을 주는지를 파악하는 것과 자기 행동이 타인에게 어떠한 영향을 미치는지를 고려하기 위해서 역할 채택과 공감을 통해 타인의 관점을 고려하는 것이라고 주장하였다[21].

이러한 주장을 바탕으로 Volker는 상담가의 도덕감수성을 측정하는 검사를 개발하며 하위요소로 사실에 대한 지각(상황지각), 나의 결정이 타인에게 미치는 결과에 대한 지각(결과지각), 윤리적 딜레마 해결에 대한 자신의 책임에 대한 지각(책임지각)을 제시하였고 이들 사이에는 위계가 존재한다고 하였다[21].

상황지각은 도덕적 문제 상황을 이해하고, 문제 해결을 위해 도덕적 문제를 규명하는 것을 의미한다. 이는 의사결정에 따라 영향을 받을 측면을 인식하고, 결과에 영향을 줄 맥락뿐 아니라 가치나 원칙에 민감하게 반응하는 것을 말한다.

결과지각은 행위의 결과가 다른 사람에게 미칠 결과를 상상하고 이해하는 것으로 이를 위해서는 문제 상황을 정확히 지각해야 하고, 관련한 사람들의 모두 포함하여 생각해야 한다.

책임지각은 문제 상황을 해결하기 위한 행동에 대한 책임을 자신에게 두어야 한다는 것이다. Volker는 개인적 책임지각이 없으면, 문제 상황을 행동이 필요한 적극적 상황으로 간주하지 않는다고 하였다[25,26].

Volker의 도덕감수성의 세 가지 하위요소는 이후의 많은 도덕감수성 연구에 토대가 되었다. 본 연구에서도 이를 바탕으로 하위요소를 선정하여 검사 도구를 개발하였다.

2.5 도덕감수성 측정방법

도덕감수성을 측정하기 위한 최초의 검사는 치과의료 분야에서 개발되었다. Bebeau(1985)는 치과의료 상황에서의 딜레마에서 도덕감수성을 측정하는 의료윤리 민감성 검사를 개발하였다. 이 연구를 통해서 도덕감수성은 교육을 통해 향상될 수 있으며, 도덕적 행동을 위한 4가지 요소는 서로 영향을 줄 수 있지만 한 가지 요소가 높다고 다른 요소가 반드시 높은 것은 아니라는 것을 확인할 수 있었다[27].

McNeel 외(1992)는 대학생들을 대상으로 도덕감수성을 측정하는 연구를 진행하였다. 검사 방법은 복합적인 문제 상황이 담겨있는 드라마를 시청하고 개별 인터뷰를 통해 녹음하고 그것을 지침에 따라 채점하는 방식이었다[28].

이지혜(2005)는 의과대학생을 대상으로 하는 도덕감수성 척도를 개발하고 이를 측정하는 연구를 진행하였다. 검사는 3개의 딜레마를 제공하고 그 속에서 도덕적인 문제들을 직접 찾아내는 지필 검사 방식으로 진행하였다[24].

국가인권위원회가 주관한 인권감수성 측정 연구에서는 다수의 선행연구처럼 상황지각, 결과지각, 책임

지각을 하위요소로 상정하였다. 검사 대상은 초등학생부터 일반인까지였고 검사 형식은 인권 문제가 되는 갈등 상황을 제시하고, 문항에 대한 중요도를 5점 척도로 평정하는 형태였다[26].

선행연구를 종합적으로 분석해보면 도덕감수성을 측정하는 방식은 연구자가 제시한 윤리적 문제 상황에 대해 검사 대상자가 스스로 윤리적 문제를 찾아내는 방식(서술형)과 윤리적 이슈와 관련한 진술문에 대해 중요도를 평정하는 방식(객관형)으로 나뉜다. 도덕감수성의 특성상 세밀한 평정을 위해서는 서술형 평가가 더 유리할 것으로 생각되나 객관형 방식도 다수의 연구에서 그 타당도와 신뢰도를 인정받은바 본 연구에서는 다수의 학생을 대상으로 하는 교육 현장에서 사용하기에는 유용한 객관형 방식으로 인공지능 윤리감수성 척도를 개발하고자 한다.

2.6 도덕감수성과 인공지능 윤리감수성

‘도덕’과 ‘윤리’라는 용어는 서로 밀접하게 연관되어 있지만, 사용되는 맥락에서 약간의 차이를 보인다. ‘도덕’이라는 용어는 일상적인 대화맥락에서 사용되는데 반해 ‘윤리’라는 용어는 보다 학문적이거나 전문적인 맥락에서 사용된다. 특히, 철학, 법학, 의학 등 전문 분야에서 도덕적 원칙과 이론을 논의할 때 ‘윤리’라는 용어를 사용한다. 본 연구에서는 일상적인 상황이 아닌 ‘인공지능’이라는 특정 분야에 대한 상황에서의 감수성을 이야기하는 것이므로 ‘도덕감수성’이 아닌 ‘윤리감수성’이라는 용어를 사용하고자 한다. 다만, ‘도덕감수성’과 ‘윤리감수성’이라는 용어는 사용되는 맥락만 다를 뿐 그 의미를 동일하게 적용한다.

‘감수성’과 ‘민감성’에 대해서도 여러 선행연구에서 혼재하여 사용하고 있는데 이것은 Rest 연구의 ‘Moral sensitivity’를 번역하는 과정에 생긴 차이로 특정 자극에 대한 반응의 정도로 동일하게 해석할 수 있다.

이에 본 연구에서는 인공지능으로 인해 나타나는 문제 상황에서의 도덕감수성을 ‘인공지능 윤리감수성’으로 정의하였고, 타 연구에 대해서는 각 연구에서 사용한 용어를 그대로 사용하였다. 따라서 본 논문에는 ‘윤리감수성’, ‘도덕감수성’, ‘도덕민감성’의 용어가 혼재되어 있음을 밝힌다.

2.7 인공지능 윤리감수성 선행연구

윤리감수성을 평가하기 위한 연구는 꾸준히 이루어져왔다. 특히, 치과대학생의 윤리감수성을 측정한 Bebeau의 연구를 시작으로 Akira(2004), 홍성훈(1999), 이지혜(2005) 등 의료인을 대상으로 한 윤리감수성 연구가 활발하게 이루어져왔다[24,27,29,30]. 국내에서는 문용린(2002)의 인권감수성 측정 연구 이후 문미희(2004), 장지원(2015), 김한솔(2019), 김지만(2020), 조윤희(2024) 등 윤리감수성을 인권과 결부지은 연구가 활발하게 이루어지고 있다[26,32-36]. 이중 조윤희(2024)의 연구는 메타버스 기반 메이커교육을 통행 인권 감수성을 함양하려는 데서 기존의 연구와 차별을 보인다.

정보영역에서는 최용성(2014)이 디지털 기술과 환경이 도덕적 감수성에 미치는 영향에 대해 연구하여 사이버 공간이 도덕적 감수성을 약화시킨다는 연구결과를 제시하였고, 김항인(2004)은 초등학생의 정보윤리 실태조사를 바탕으로 정보윤리 감수성을 발달시킬 방안을 연구하였다[37,38]. 김항인에 따르면 도덕적 행동과 도덕적 민감성 간에는 정적 상관관계가 있는 것으로 나타났다. 이혜령(2023)은 초등학생의 인공지능 윤리 인식 척도 개발연구에서 인공지능 윤리감수성, 판단력, 동기화 측정할 수 있는 척도를 개발하였다[39]. 이 척도는 인공지능 분야에서 처음으로 윤리감수성을 평가했다는 점에서 의의가 있지만 이야기마다 평가하고 있는 요소가 고르지 않고, 한 가지 척도에서 다양한 요인을 측정하려다 보니 각 요인의 하위 요소까지 세부적으로 평가하고 있지 못하고 있다는 점에서 아쉬움이 있다.

이에 본 연구에서는 Rest의 도덕적 행동에 영향을 주는 4가지 구성요소 중 윤리감수성 영역에만 집중하여 인공지능 윤리감수성의 하위요소까지 평가할 수 있는 인공지능 윤리감수성 검사 도구를 개발하고자 한다.

3. 연구 방법

본 연구의 목적은 평정 대상의 인공지능 윤리감수성을 측정할 수 있는 도구를 개발하고 타당성을 검증하는 것이다. 인공지능 윤리감수성 검사 도구는 인공지능 윤리 문제 상황을 설정하고, 관련한 문제에 대해 자신의 생각대로 평정한 답을 바탕으로 인공지능 윤

리감수성을 측정한다. 따라서 인공지능 윤리 문제 상황을 읽고 해석할 수 있는 사람이라면 누구나 본 검사 도구로 인공지능 윤리감수성 측정이 가능하다. 단, 초등학생의 경우 어휘력 부족으로 문제 상황을 제대로 이해하지 못할 수 있기에 본 검사 도구를 그대로 사용하기에 무리가 있다. 따라서 본 검사 도구의 평정 대상은 중등학생 이상으로 한다. 검사 도구 개발 절차는 다음과 같다.

첫째, 문헌 연구 및 선행연구 분석을 통해 인공지능 윤리 감수성의 개념과 하위요소를 선정한다.

둘째, 인공지능 윤리 딜레마 가치 요소를 선정하고, 인공지능 윤리 딜레마 일화를 개발한다.

셋째, 전문가 자문 및 타당도 검사를 통해 인공지능 윤리 딜레마의 타당도 및 적합성을 확인한다.

넷째, 인공지능 윤리 딜레마에 대한 인공지능 윤리 감수성 검사 문항을 작성한다.

다섯째, 전문가 자문 및 타당도 검사를 통해 인공지능 윤리감수성 문항의 타당도를 확인한다.

여섯째, 예비조사 및 전문가 타당도 검사를 중학교 수준에 적합한지 확인한다.

일곱째, 최종 문항을 확정한다.

4. 인공지능 윤리감수성 검사 도구 개발

4.1 인공지능 윤리감수성 정의 및 하위요소

문헌 고찰 및 분석을 통해 인공지능 윤리감수성의 개념을 정의하고 하위요소를 선정하였다.

먼저 인공지능 윤리감수성은 다음과 같이 정의하였다.

Rest의 도덕성 4구성 요소 이론을 인공지능 윤리 영역에도 적용해보면, 인공지능 윤리 문제 상황에서 윤리적 행동이 이루어지기 위해서 인간은 4가지 심리적 요소를 거쳐야 한다. 즉, 문제 상황을 인공지능 윤리 문제 상황으로 인식해야 하고(인공지능 윤리감수성), 어떤 행동이 옳은지 그른지 판단해야 하며(인공지능 윤리판단력), 다른 가치에 비해 인공지능 윤리의 가치를 우선시 하고(인공지능 윤리동기화), 옳다고 생각하는 행동을 끝까지 밀고 나갈 수 있는 실행력(인공지능 윤리품성화)을 지녀야 한다.

본 연구에서는 위 4가지 요소 중 제1요소인 인공지능 윤리감수성을 ‘인공지능 윤리 문제 상황에 당면했을 때 이를 인공지능 윤리와 관련된 문제 상황으로 지각하고 해석하며, 그 상황에서 특정 행동이 타인에게 어떠한 영향을 미칠지 예상하고, 그 상황을 해결하기 위한 책임이 자신에게 있다고 인식하는 것’으로 정의한다.

다음으로 인공지능 윤리감수성 하위요소는 다음과 같이 정의하였다.

문헌 분석을 통해 국내외 많은 연구에서 Volker(1984)가 제시한 상황지각, 결과지각, 책임지각을 바탕으로 윤리감수성 하위요소를 선정한 것을 확인할 수 있었다. 한국청소년정책연구원(2012)의 연구에서는 Volker의 도덕감수성 하위요소 중 상황지각과 결과지각만을 하위요소로 선정하여 연구를 진행하는 대신 상황지각과 결과지각을 더 세부적으로 나누었다[29]. 이 중 결과지각을 행위결과지각과 행위에 대한 타인의 감정으로 나누었다. 국가인권위원회(2002)의 연구에서도 결과지각을 타인의 정서 인식 능력과 결과 예측 능력으로 표현하였다[26]. 이혜령(2023)도 인공지능 윤리민감성의 하위 요소를 상황 인식과 결과 인식으로 나누고, 이중 결과 인식을 다시 행위 결과 지각과 행위에 대한 타인의 감정 지각으로 나누었다. 이처럼 다수의 연구에서 결과지각을 단순히 다른 사람에 미치는 결과를 예상하는 것으로만 정의하지 않고 타인의 감정에 대한 공감까지 포함하는 것을 확인할 수 있었다[39]. 이는 윤리감수성의 하위요소로 행위의 결과를 예측하는 것뿐만 아니라 타인의 감정에 공감하는 것 또한 중요하다가는 것을 방증하는 것이다.

이에 본 연구에서는 인공지능 윤리 감수성의 하위요인으로 Volker의 하위요인을 바탕으로 하되 공감지각을 추가하여 아래의 4가지로 설정하였다.

첫째, 상황지각은 상황에 대한 해석 능력으로 문제 상황을 인공지능 윤리 문제 상황으로 인식하는가에 대한 것이다.

둘째, 결과지각은 타인에게 미칠 결과에 대한 인지로 인공지능 윤리 문제가 타인에게 가져올 수 있는 잠재적 결과를 상상하고 이해하는 것이다.

셋째, 공감지각은 인공지능 윤리 문제가 타인의 감정에 미치는 영향을 인식하고 타인의 감정에 공감하는지에 대한 것이다.

넷째, 책임지각은 문제 상황에 대한 해결책임을 자신에게 돌리는 것으로 인공지능 윤리 문제를 자신의 문제로 인식하고 이를 실제 행동으로 실천하려는 의지를 말한다.

4.2 문항 형식 및 채점 방식

문항 형식 및 채점 방식은 국내 도덕 발달이론의 권위자인 문용린의(2002) 인권감수성 지표 개발 연구를 기본으로 하여 본 연구에 맞게 수정하였다.

문용린에 따르면 선다형 문항의 경우 사회적 바람직성 제거에 어려움으로 인해 감수성 측정에 적절하지 못하다. 이에 본 연구에서는 인공지능 윤리 가치와 관련된 진술문과 그에 대비되는 가치를 진술문에 중요도를 평정하도록 하는 리커트 형식으로 문항을 제작하였다.

채점은 인공지능 윤리 가치 문항에 대한 점수가 다른 가치 문항에 대한 점수보다 높은 경우에만 점수를 부여하고 합산하여 계산하는 방식으로 Figure 1과 같다.

If A > B A-B=Question score
else 0=Question score

A = Scores rated for AI ethical value question
B = Scores rated for other value question

Figure 1. AI Ethics Sensitivity Scoring Method

4.3 인공지능 윤리 딜레마 일화 개발

문헌 고찰 및 분석을 통해 인공지능 윤리에 대한 관점이 로봇에서 인간중심으로 옮겨온 것을 확인할 수 있었고, 특히 국내에서는 '사람이 중심이 되는 「인공지능 윤리기준」' 이후 발표된 인공지능 윤리 관련 정책이나 지침에 근간이 된 것을 확인할 수 있었다. 이에 본 연구에서는 인공지능 윤리의 최고 가치를 '인간성(Humanity)'에 두고 '사람이 중심이 되는 「인공지능 윤리기준」' 3대 기본원칙인 '인간존엄성', '사회 공공선', '기술 합목적성'에 기반을 두어 딜레마 일화를 제작하였다.

딜레마 일화는 인공지능 윤리 관련 논문, 기사, 영상 등을 참고하여 1차로 총 15개의 소재를 추출하였고, 2차에서는 기본 원칙 당 2개씩 총 6개의 소재를 선정 후 이를 딜레마 사례로 만들었다. 3차에서는 인공지능교육 분야를 연구하고 있는 5명의 교사와 논의하여 각 원칙별 1개씩 총 3개의 딜레마 사례를 선정하였다. 인간존엄성 원칙 딜레마에서는 '로봇 보안관 채용'의 경우 효율성과 안전요소가 함께 묶어 있어 삭제하였고, 사회공공선 원칙 딜레마에서는 지점폐쇄로 인한 불편함을 강조하기 위해 '고객센터' 사례를 삭제하였으며, 기술합목적성 원칙 딜레마에서는 '탄소를 뿜는 AI'의 경우 AI 개발 과정에서 발생하는 환경오염을 학생들이 이해할 수 있는 수준으로 설명하기에는 문항이 너무 길어져 삭제하였다.

선정된 딜레마 일화는 다음과 같다.

첫 번째 일화는 인공지능 채용시스템 이용과정에서의 효율성과 공정성 딜레마로 인간존엄성과 관련되어 있다.

두 번째 일화는 인공지능 기반 온라인 은행과 관련하여 비용절감과 노인접근성에 대한 딜레마로 사회공공선과 관련되어 있다.

세 번째 일화는 비윤리적 인공지능 개발과정에 대한 딜레마로 기술합목적성과 관련되어 있다.

이렇게 개발된 딜레마 일화는 전문가 타당도 검증 과정을 거쳐 수정 및 보완 후 다음과 같이 완성되었다. 인공지능 딜레마 일화 개발과정은 Table 4 와 같다.

Table 4. AI Ethics Dilemma Episode Development Process

Material extraction	
▪ Robotic sheriff recruitment ▪ AI recruitment and bias ▪ AI speaker social safety system ▪ Behind the digital transformation ▪ Shadow of AI filtering ▪ Excessive suppression of robot police officers ▪ Light and shadow of deep fake	▪ Cleaning robots and jobs for the elderly ▪ Drone delivery ▪ AI weapon ▪ AI wins contest ▪ Child-care robot ▪ Carbon-breathing AI ▪ AI customer center ▪ Ex-convicts and facial recognition CCTV ▪ Child-care robot

⇩

Material selection and Dilemma production	
Human dignity	▪ Robotic sheriff recruitment ▪ AI recruitment and bias
Social public good	▪ Behind the digital transformation ▪ AI customer center
Technical purposefulness	▪ Shadow of AI filtering ▪ Carbon-breathing AI

⇩

Final Dilemma Selection	
Human dignity	▪ AI recruitment and bias
Social public good	▪ Behind the digital transformation
Technical purposefulness	▪ Shadow of AI filtering
Final confirmation after expert validation	

전문가 타당도 검사 방법은 Lawshe(1975)의 내용 타당도 비율(Content Validity Ratio, CVR)을 활용하였으며, 총 10명의 전문가에게 각 딜레마 일화가 3대 원칙에 대한 딜레마로 적합한지 내용 타당도 검증을 받았다. 전문가 집단은 모두 인공지능 관련 분야를 전공하였을 뿐 아니라 모두 교육전문가로 인공지능교육 또는 인공지능 윤리교육 경험이 있다. 전문가에 대한 세부정보는 Table 5와 같다.

Table 5. Expert group

Degree	Years of Research Experience	Years of Teaching Experience	Occupation
Ph.D. in Computer Science	6	12	Professor
Ph.D. in Big Data Engineering	7	10	Teacher
Ph.D. in Computer Science	24	4	Professor
Ph.D. in Computer Education	6	12	Lecturer
ABD Ph.D. in Computer Education	8	6	Teacher
ABD Ph.D. in Computer Education	9	13	Teacher
M.Sc. in Computer Education	5	16	Teacher
M.Sc. in Educational Psychology	8	15	Teacher
M.Sc. in AI and Education Integration	3	22	Teacher
M.Sc. in Computer Education	4	10	Teacher

각 딜레마 일화의 CVR 값을 토대로 분석한 결과 각 딜레마가 '인간존엄성', '사회공공선', '기술합목적성' 의 딜레마로 적합한 것으로 확인되었다. 이에 따라 최종 인공지능 윤리 딜레마 일화를 확정하였다. 타당도 검사 결과는 Table 6과 같다.

Table 6. AI Ethics Dilemma Episode CVR

Dilemma	M	SD	CVR
AI recruitment and bias	4.4	0.52	1.0
Behind the digital transformation	4.2	1.23	0.9
Shadow of AI filtering	4.5	0.53	1.0

4.4 인공지능 윤리감수성 문항 개발

앞서 개발한 인공지능 윤리 딜레마 일화별로 인공지능 윤리감수성의 하위요소를 검사할 수 있는 5점 척도의 예비 문항을 만들었다. 문항은 딜레마 당 4문항이며 각 문항은 다시 인공지능 윤리 가치와 다른 가치에 대해 묻는 2개 하위 문항으로 구성하여 총 24개의 문항을 만들었다.

각 딜레마별 첫 번째 문항은 상황지각 문항으로 딜레마 상황을 윤리적 문제 상황으로 평가하고 있는지를 확인하는 문항으로 구성하였다.

두 번째 문항은 결과지각에 대한 문항으로 인공지능 윤리문제 상황이 타인에게 미칠 결과를 인지하고, 그 결과를 얼마나 중요하게 생각하고 있는지 묻고 있다.

세 번째 문항은 공감지각의 문항으로 문제 상황에 관련된 사람들의 감정에 얼마나 공감을 하고 있는지를 묻는 문항이다.

네 번째 문항은 책임지각 문제로 인공지능 윤리 문제를 자신의 문제로 인식하고 이를 실제 행동으로 실천하려는 의지가 있는지에 대한 문항이다.

이렇게 개발된 예비 문항에 대해서는 30여 명의 성인을 대상으로 세 차례에 걸쳐 파일럿 테스트를 시행하였다. 1차 테스트에서 결과지각 문항을 개발자의 의도와 다르게 해석하는 비율이 높은 것을 확인하였다. 이에 따라 결과지각 문항을 '위 사건과 관련하여 다음과 같은 결과가 있을 수 있습니다. 각각의 결과가 어느 정도 중요하다고 생각하십니까?' 와 같이 포괄적인 문항에서 '기존의 AI채용시스템을 그대로 이용할 경우 다음과 같은 결과가 예상됩니다. 당신은 각각의 결과가 얼마나 중요하다고 생각하십니까?' 와 같이 딜레마 문항이 반영된 형태로 수정하였다.

그리고 인공지능 윤리 가치와 다른 가치 평정 점수가 모두 5점에 몰려있는 문항을 선별하여

'이익을 내야하는 B은행의 대표 입장이 이해된다.' 를

'비용절감을 위해 모든 지점을 폐쇄하려는 B은행 대표 입장이 이해된다.'

와 같은 형태로 어조를 변경하였다.

문항을 수정 후 2차 테스트를 진행하여 수정된 문항에 대한 이해와 변별이 높아진 것을 확인 할 수 있었다. 그러나 앞선 수정에도 불구하고 인공지능 윤리 가치와 다른 가치 점수의 차이가 없는 응답자가 다수 존재하여 전문가 자문을 통해 5점 척도를 7점 척도로 변경하였다. 수정 후 3차 테스트 결과 변별력이 높아진 것을 확인할 수 있었다.

3차 테스트 후 인공지능 윤리감수성이 유난히 낮게 측정된 2명의 피험자를 면담해보았는데 실제로 윤리

적 가치보다 경제성이나 효율성의 가치를 더 중요하게 생각하고 있었으며, 이 때문에 같은 문제 상황을 보고도 다르게 해석하는 것을 확인할 수 있었다.

세 차례의 파일럿 테스트를 거친 후 수정된 문항들의 내용 타당도를 확인하기 위해 인공지능 윤리 딜레마 일화 타당도 검사에서와 마찬가지로 내용 타당도 비율(CVR) 방법을 사용했다. 타당도 검사에 참여한 집단은 인공지능 윤리 딜레마 일화 타당도 조사에 참여한 전문가와 동일하다. 타당도 조사에서는 인공지능 윤리 가치와 다른 가치를 반영하는 문항들을 한 쌍으로 묶어 평가했다. 각 문항의 CVR 값을 토대로 분석한 결과(Table 7 참조), 모든 문항이 인공지능 윤리감수성의 하위요소를 적절히 평가할 수 있음을 확인했다.

Table 7. AI Ethics Sensitivity Question CVR

	AI recruitment and bias			Behind the digital transformation			Shadow of AI filtering		
	M	SD	CVR	M	SD	CVR	M	SD	CVR
Situational awareness	4.5	0.53	1	4.4	0.52	1	4.5	0.53	1
Result recognition	4.4	0.52	1	4.5	0.53	1	4.5	0.53	1
Empathy recognition	4.4	0.52	1	4.6	0.52	1	4.5	0.53	1
Responsibility recognition	4.5	0.53	1	4.4	0.52	1	4.4	0.52	1

4.5 예비검사를 통한 최종 문항 확정

4.4에서 개발한 인공지능 윤리감수성 문항이 중학생이 이해할 수 있는 수준인지 확인해보기 위해 예비검사를 실시하였다. 예비검사는 중학교 3학년 학생 21명 대상으로 실시하였으며, 예비검사 실시 후 각 딜레마별 문항을 이해하는 것이 어렵지 않은지 설문을 하였다. 설문결과 모든 딜레마 영역에 대해 90%이상의 학생이 이해하기 어렵지 않다고 응답하였다. 학생들의 설문결과는 Table 8과 같다.

Table 8. Question Level Appropriateness CVR

Dilemma	Not difficult at all	Not difficult	Difficult	Very difficult
AI recruitment and bias	62%	33%	5%	0
Behind the digital transformation	48%	43%	9%	0
Shadow of AI filtering	57%	33%	10%	0

최종적으로 평균 교육경력 10.2년의 중등교사 5인을 대상으로 본 검사도구가 중학생이 이해하기에 어렵지 않은지 전문가 타당도 검사를 실시하였다. 설문결과는 Table 9와 같다. 학생 설문결과와 전문가 타당도 검사결과를 통해 본 검사도구가 중학생의 수준에 적합한 것으로 확인되어 최종적으로 문항을 확정하였다.

Table 9. Question Level Appropriateness CVR

Dilemma	M	SD	CVR
AI recruitment and bias	4.4	0.52	1.0
Behind the digital transformation	4.2	1.23	0.9
Shadow of AI filtering	4.5	0.53	1.0

5. 연구 결과

본 연구는 Rest의 도덕적 행동의 네 가지 구성 요소 이론을 기반으로 하여, 인공지능 윤리 문제 상황에서 도덕적 행동을 유발하는 주요 요인 중 하나인 인공지능 윤리감수성을 측정하는 도구를 개발하는 것을 목적으로 하였다. 인공지능 윤리감수성의 하위요인은 상황지각, 결과지각, 공감지각, 책임지각으로 설정하였다.

본 도구는 인공지능 윤리 갈등상황을 나타내는 일화를 읽고, 인공지능 윤리 가치 문항과 다른 가치 문항에 대한 중요도를 7점 척도로 평정하는 형식으로 되어있다. 인공지능 윤리 갈등 일화는 총 3개로 각 사람이 중심이 되는「인공지능 윤리기준」’ 의 3대 원칙인 ‘인간존엄성’, ‘사회공공선’, ‘기술합목적성’ 의 가치와 다른 가치 간의 딜레마 형태이다. 문항은 각 일화와 관련하여 하위요소를 평가할 수 있는 문항으로 이루어져 있다. 문항은 일화별 8문항으로 총 24문항이다. 최종 개발된 문항 Table 10, 11, 12의 내용으로 구성되어 있다.

Table 10. AI recruitment and bias Summary

Dilemma	
Company D implemented an AI system for employee selection, which resulted in higher-performing employees than those chosen by human recruiters. However, the AI showed bias towards candidates from a specific region and gender, likely because it learned from the resumes of successful existing employees. Despite recognizing this bias, the CEO is contemplating continuing with the unaltered AI system due to its effectiveness in identifying high performers.	
Situational Perception Question	Is the inclusion of region of origin and gender in the employee selection process through an AI system recognized as a problem?
Result Perception Questions	What are the expected outcomes of continuing to use the AI system without modifications?
Empathy Perception Question	What is your level of empathy towards the perspectives of individuals involved in employee selection using an AI system?
Empathy Perception Question	If you were the CEO of Company D, would you modify the AI system?

Table 11. Behind the digital transformation Summary

Dilemma	
Kim's grandfather was surprised to find a bank branch missing when he went to get a loan, due to Bank B closing several branches as it shifted to AI-based mobile banking services. This strategy, aimed at reducing costs and eventually closing all branches, led to fewer customers using physical locations. Despite the bank's plans to improve deposit rates with the savings, Kim's grandfather found the mobile app challenging to use and decided against taking out the loan.	
Situational Perception Question	Do you view the lack of consideration for the accessibility of socially vulnerable groups in the development of AI technology as a problem?
Result Perception Questions	What are the expected outcomes of closing all bank branches?
Empathy Perception Question	What is your level of empathy for the situation between marginalized groups affected by AI technology and the providers of AI technology?
Empathy Perception Question	If you were the CEO of Bank B, would you maintain bank branches considering the marginalized groups affected by AI technology?

Table 12. Shadow of AI filtering Summary

Dilemma	
Company C, which offers generative AI services globally, relies on labor from underdeveloped countries to filter inappropriate content from vast data sets, due to limited job opportunities in these regions. Despite the high demand for these jobs, workers are exposed to large volumes of violent and disturbing content daily, leading to increasing reports of mental harm. Nonetheless, Company C's representative has chosen not to reduce the workload to avoid higher development costs.	
Situational Perception Question	To what extent do you recognize issues with generative AI services that have been developed unethically?
Result Perception Questions	What are the expected consequences of developing AI unethically?
Empathy Perception Question	What is your level of empathy for individuals who have been harmed by an unethical technology development process?
Empathy Perception Question	If you were the CEO of Company C, would you develop AI technology ethically, even if it incurs additional costs?

6. 결론

본 연구는 Rest의 도덕적 행동에 미치는 4가지 요소 이론에 근거하여 인공지능 관련 상황에서 윤리적 감수성을 평가할 수 있는 검사 도구를 개발하였다.

인공지능 윤리교육의 궁극적인 목표는 학생이 인공지능 윤리문제 상황에 직면했을 때 적절한 윤리적 행동을 취할 수 있도록 하는 데 있다. 그러나 현재 인공지능 윤리 분야에서는 윤리적 행동에 영향을 미치는 요인에 관한 연구가 미미하고, 이를 평가할 수 있는 도구 또한 찾아보기 힘들다. 이에 대다수의 인공지능 윤리교육 연구가 윤리의식 변화에 초점을 맞추고 있다. 이런 배경에서 개발된 인공지능 윤리감수성 검사 도구는 윤리적 행동 촉진이라는 인공지능 윤리교육의 목적을 명확히 하고, 교육 영역을 확장하는 데 중요한 역할을 할 것이라는데 의의가 있다. 또한, 이 도구는 교육 분야를 넘어 개발자, 사용자, 정책 결정자 등 다양한 이해관계자들이 윤리적 고려 사항을 이해하고 적용하는 데 도움을 줄 것이다.

본 검사 도구는 Rest의 이론과 다수의 윤리감수성

연구에 기반을 두어 개발되었지만, 실제 학생에게 투입해보고 인공지능 윤리감수성과 윤리적 행동의 상관관계에 대해서는 확인해보지는 못하였다는데 그 한계가 있다. 따라서 향후 윤리적 행동에 영향을 주는 타 요인과 인공지능 윤리감수성 사이의 관계를 분석하여 본 검사 도구의 타당도를 높이는 연구를 진행할 예정이다.

또한 추후에는 인공지능 윤리감수성 신장 교육 프로그램 및 이를 일반화 할 수 있는 교육 모형을 개발하고, 이에 대한 효과성을 본 연구에서 개발한 인공지능 윤리감수성 검사 도구를 이용하여 측정해보고자 한다. 이러한 연구는 학생들의 인공지능 윤리감수성을 높여, 인공지능 기술의 윤리적 사용과 발전을 촉진하는 실질적인 변화를 이끌어내는데 기여할 것으로 기대된다.

참고문헌

[1] Byun, S. (2020). A Study on the Necessity of AI Ethics Education. *The Journal of Korea elementary education. 31*(3), 153-164. DOI : 10.20972/kjee.31.3.202009.153

[2] Kim, E & LEE. Y. (2023). Development of AI Ethics Dilemma Questions for AI Ethics Education. *The Journal of Korean Association of Computer Education, 26*(5), 31-42. DOI : 10.32431/kace.2023.26.5.003

[3] McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development?. *In Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*. 729-733. DOI : 10.1145/3236024.3264833

[4] CBS Mornings. (2023, March 1). *Interview: "Godfather of artificial intelligence" talks impact and potential of AI*[Video]. Youtube. https://www.youtube.com/watch?v=qpoRO378qRY&t=1222s

[5] Kim, E & LEE. Y. (2022). The influence of artificical intelligence ethics education using Moral machine on elementary school students ' perception of artificical intelligence. *The Journal of Korean Association of Computer Education, 25*(3), 1-8. DOI : 10.32431/kace.2022.25.3.001

[6] 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

[7] 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 Computer Education, 26*(2), 29-39. DOI : 10.32431/kace.2023.26.2.004

[8] 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 Computer Education, 25*(5), 1-14. DOI : 10.32431/kace.2022.25.5.001

[9] Lee, J. & Byun, S. (2023). Validation of the Effectiveness of AI Ethics Education Program:focusing on application to elementary moral education. *Korea Elementary Moral Education Society,* (84), 285-308. DOI : 10.17282/ethics.2023..84.285

[10] Park, J. & Chung, H. (2021). Effects of the Project-based AI Education Program on AI Ethical Consciousness and Creative Problem-Solving Skills using Flipped Learning. *Journal of Research in Curriculum & Instruction. 25*(5). 359-368. DOI : 0.24231/rici.2021.25.5.359

[11] Rest, J. R. (2008). *Moral Development: Advance in Research and Theory.* (Moon, Y, Trans.). Hakjisa. (Original work published 1986)

[12] William M. K & Jacob L. G. (2004). *Moral Development.* (Moon, Y, Trans.). Hakjisa. (Original work published 1995)

[13] Lee, J., Kim, D., & Yang, H. (2019). *A Prospective Analysis of Artificial Intelligence(AI) Technology and Innovation Policies-Focusing on System Improvement Plans for Safe and Ethical AI R&D and Utilization. Policy research*, 1-226. DOI :

[14] Heo, Y., Lee, Y., & Sim, J. (2020). Artificial Intelligence Ethics and RoboEthics, Differences and Continuity -Toward AI ethics as everyone's ethics-. *Pilosophy · Ieology · Clture*, (34), 41-72. DOI : 10.33639/ptc.2020.34.003

[15] NIA. (2019). Artificial Intelligence Ethics Guidelines-Focused on Japanese and EU cases. *Intelligent information society legal system issue report.*

[16] UN News . (2021). https://news.un.org/en/story/2021/1

1/1106612.

[17] Ministry of Science and ICT. (2020). National AI ethicalstandards. Human-centered 「AI ethical standards」

[18] Bae, J., Lee, J., Hong, M. & Cho, Jung. (2022). *The Journal of Korean Association of Computer Education, 25(6), 103-118.* DOI : 10.32431/kace.2022.25.6.008

[19] Lee, Y. (2021). *Journal of The Korean Association of Information Education, 25*(1), 71-79. DOI : 10.14352/jkaie.2021.25.1.71

[20] Jung, D. & Park, H. (2023). *Korean Journal of Elementary Education, 34*(3), 39-54. DOI : 10.20972/Kjee.34.3.202309.39

[21] Park, G., Hong, S., Seo, K., Han, H. & Kim, Y. (2011). *A Preliminary Study on making Tool for the Moral Sensitivity Test for Korean Youth.* National Youth Policy Institute.

[22] Schwartz, S. H. (1977). *Normative influences on altruism. In Advances in experimental social psychology, 10,* 221-279. DOI : 10.1016/S0065-2601(08)60358-5

[23] Sauberman, D. C. (1978). *Irrational attributions of responsibility: Who, what, when, where and why.*

[24] Lee, J. (2005). *Study of moral sensitivity scale development and the trait of moral sensitivity.* Seoul National University master's thesis.

[25] Volker, J. M. (1984). *Counseling experience, moral judgment, awareness of consequences and moral sensitivity in counseling practice*. University of Minnesota.

[26] Moon, Y. (2002). *Developing indicators of psychological scale for human rights sensitivity in Korea.* National Human Rights Commission Of The Republic Of Korea.

[27] Bebeau, M. J., Rest, J. R., & Yamoor, C. M. (1985). Measuring dental students' ethical sensitivity. *Journal of Dental Education, 49*(4), 225-235.

[28] McNeel, S. P. (1994). College teaching and student moral development. *Moral development in the professions: Psychology and applied ethics, 27*, 49.

[29] Lim, Y., Son, G., Seo, K., Shin, T. & Chung, K. (2012). *A study on the development of the moral sensitivity test for adolescents II.* National Youth Policy Institute.

[30] Akira, et. al(2004), *The development of a brief and objective method for evaluating moral senitivity and reasoning in medical student*, BMC Medical Ethics, 5(1) http://www.ethicsweb.ca/guide/moral-decision.html

[31] Hong, S. (1999). *A Study on the Development of Medical Education Program.* (Doctoral dissertation). Seoul National University, Seoul, Korea.

[32] Moon, M. (2004). *Developing a Human Right Education Program for Preservice Teachers.* (Doctoral dissertation). Seoul National University, Seoul, Korea.

[33] Jang, J., & Lee, Y. (2015). The effects of non-disabled elementary school student's emotional intelligence and human rights sensitivity on acceptance attitude of disability. *J Rights Child Disabil, 6*(1), 1-26.

[34] Kim H., & Han, Y. (2019). Human Rights Sensitivity as Mediating Factor in Relationship between Emotional Regulation Style and Attitude to Violences in Their 20s. *Journal of Human Rights & Law-related Education, 12*(1), 163-189.

[35] Kim, J., Hong, K., Lee, C., & Kim, H. (2020). A study on the sensitivity of human rights and the advocacy activities of Korean occupational therapists. *The Journal of Korean society of community based occupational therapy, 10*(2), 11-24. DOI : 10.18598/kcbot.2020.10.2.02

[36] Cho, Y. & Park, H. (2024). Development and Application of a Metaverse-based Maker Education Program to Cultivate Human Rights Sensitivity. *The Korean Society for Artificial Intelligence, 3*(1), 27-68.

[37] Choi, Y. (2014).The Study on the Moral Psychology of Cyber Space and Information Ethics Education's Teaching Strategies of J. Rest's Four Moral Components Model. *The Korean Association of Ethics, 1*(94), 277-325. DOI : 10.15801/je.1.94.201403.277

[38] Kim, H. (2004). A study on development of the information ethics sensitivity. *The Korean Society for the Study of Moral and Ethics Education, 19*, 1-24.

[39] Lee, H. Development of Artificial Intelligence Ethics Awareness Scale for Elementary School Students. *The Korean Society for Artificial Intelligence, 2*(1), 98-127.

김 은 경

2009년 청주교육대학교
초등교육과(교육학학사)
2022년 한국교원대학교 컴퓨터교육학과
(교육학석사)
2022년~현재 한국교원대학교 초등컴퓨터교육 박사과정
관심분야: 인공지능교육, 인공지능 윤리교육
E-Mail: kektb86@gmail.com

이 영 준

1988년 고려대학교 전산과학과(이학사)
1994년 미국 미네소타대학교 전산학과
(Ph.D.)
2003년~현재 한국교원대학교 컴퓨터교육학과 교수
관심분야: 지능형시스템, 학습과학, 정보교육, 인공지능교육
E-Mail: yjlee@knue.ac.kr

부 록

인공지능 윤리가치 문항(N-2)과 다른 가치 문항(N-1) 점수를 비교하여,

N-2 > N-1 보다 경우

N문항 점수 = (N-2) - (N-1)

N-2 ≦ N-1 보다 경우

N문항 점수 = 0

N-1은 다른 가치 문항, N-2는 인공지능 윤리가치 문항임.

[그림 1] 인공지능 윤리감수성 검사도구 채점 방식

<표 10> AI채용시스템과 편향

D 회사는 지난해 직원을 뽑을 때 , 사람이 아닌 AI 가 이력서를 분석하여 직원을 선발하는 시스템을 도입하였다. 1년 후 AI 시스템을 통해 선발된 직원과 기존에 사람에 의해 선발된 직원들의 실적을 비교해 보았더니 AI 시스템을 통해 선발된 직원의 실적이 더 우수한 것으로 나타났다. 그러나 이 AI 시스템을 통해 선발된 직원들을 자세히 살펴보니, 대다수가 같은 지역 출신이며 한 성별에 집중되어 있었다. 이러한 현상은 AI 시스템이 기존 직원들 중 우수한 실적을 낸 직원들의 이력서를 학습하여 직원을 선발했기 때문으로 보인다. D 회사의 대표는 AI 시스템이 직원 선발과정에서 출신지역과 성별을 반영하는 것을 인지하였지만, AI시스템으로 선발된 직원이, 사람에 의해 선발된 직원에 비해 실적이 좋았기 때문에 기존의 AI시스템을 수정하지 않고 계속 사용할지 고민하고 있다.	
1. 위 상황과 관련하여 다음과 같은 주장을 할 수 있습니다. 각각의 주장에 어느 정도 동의하십니까?	
1-1	직원을 선발할 때에는 공정하게 뽑는 것보다 우수한 직원을 뽑는 것이 더 중요하다.
1-2	AI가 직원을 선발하는 과정에서 출신지역이나 성별을 반영하게 해서는 안 된다.
2. AI시스템을 수정하지 않고 그대로 이용할 경우 다음과 같은 결과가 예상됩니다. AI시스템을 수정할지 말지 결정할 때, 각각의 결과를 얼마나 중요하게 고려해야 한다고 생각하십니까?	
2-1	우수한 성과의 직원이 많아져 회사의 실적이 좋아지는 것
2-2	출신지역이나 성별로 인해 채용과정에서 차별받는 사람들이 생기는 것
3. 위 상황과 관련하여 다음과 같은 감정이 들 수 있습니다. 당신은 각각의 감정에 어는 정도 공감하십니까?	
3-1	AI시스템이 출신지역과 성별을 반영하는 것을 알면서도, 우수한 직원을 뽑기 위해, AI시스템을 수정하지 않고 이용하려는 D회사 대표의 입장이 공감된다.
3-2	출신지역이나 성별로 인해 공정한 경쟁을 할 수 없게 된 사람들이 안타깝다.
4. 위 상황과 관련하여 다음과 같은 행동을 할 수 있습니다. 당신이 D회사 대표라면 어떻게 하시겠습니까?	
4-1	현재의 AI시스템을 수정하지 않고 직원 선발 과정에 그대로 사용한다.
4-2	직원 선발을 선발할 때, 출신지역과 성별이 반영되지 않도록 AI시스템을 조정한다.

<표 11> AI서비스 전환의 이면

김씨 할아버지는 대출을 받기 위해 방문하던 은행 지점에 갔다가, 그 지점이 없어진 것을 보고 당황했다. 이는 B은행이 계좌 개설, 예금, 대출 등 다양한 은행 업무를 AI 기반의 모바일앱에서 가능하게 하면서 여러 지점을 폐쇄했기 때문이다. 모바일 앱에서 은행 업무가 가능하게 되자, 은행 지점을 이용하는 고객들이 크게 줄어들었고 B은행은 비용 절감을 위해 이용객이 적은 지점을 폐쇄한 것이다. B은행의 대표는 앞으로 남은 지점을 모두 폐쇄하고, 절감된 비용 중 일부를 신규 고객 유치를 위해 예금 금리 인상에 사용할 예정이다. 그러나 김씨 할아버지는 AI모바일 앱 사용방법이 어려워 결국 대출을 포기하였다.	
1. 위 상황과 관련하여 다음과 같은 주장을 할 수 있습니다. 각각의 주장에 어느 정도 동의하십니까?	
1-1	은행 운영비용을 줄이기 위해서라면 모든 지점을 폐쇄해도 된다.
1-2	AI기술은 사회적 약자와 디지털 취약 계층의 접근성을 보장하도록 개발 및 활용되어야 한다.
2. 모든 은행 지점을 폐쇄하고 AI기반 모바일앱으로 대체할 경우 다음과 같은 결과가 예상됩니다. 모든 지점을 폐쇄할지 말지 결정할때, 각각의 결과를 얼마나 중요하게 고려해야 한다고 생각하십니까?	
2-1	B은행의 운영 비용이 줄어드는 것
2-2	노인과 장애인, 어린이와 같은 디지털 취약계층이 은행 업무 처리에 어려움을 겪는 것
3. 위 상황과 관련하여 다음과 같은 감정이 들 수 있습니다. 당신은 각각의 감정에 어는 정도 공감하십니까?	
3-1	은행 운영 비용을 줄이기 위해 모든 지점을 폐쇄하려는 B은행 대표 입장이 공감된다.
3-2	기술 변화에 소외감과 좌절감을 느낄 김씨 할아버지가 안타깝다.
4. 위 상황과 관련하여 다음과 같은 행동을 할 수 있습니다. 당신이 B은행 대표라면 어떻게 하시겠습니까?	
4-1	비용을 줄이기 위해 모든 지점을 폐쇄하겠다.
4-2	비용이 들더라도 디지털 취약 계층의 접근성 보장을 위해 지점을 운영하겠다.

<표 12> AI필터링의 그림자

C사는 전 세계에 생성형 AI서비스를 제공하는 회사로, 대량의 데이터를 분석하여 유용한 정보를 만들어낸다. 이 과정에서 부적절한 내용을 걸러내기 위해 많은 노동력이 필요한데, C사는 저개발 국가의 사람들에게 적은 돈을 주고 이 작업을 시키고 있다. 저개발 국가에는 일자리가 많이 없기 때문에 하루 종일 많은 양의 폭력적이고 충격적인 이미지와 영상을 봐야 하는 이 작업에도 지원자가 넘쳐난다. 그러나 하루에 처리해야 하는 폭력적이고 충격적인 콘텐츠의 양이 많기 때문에, 작업을 수행하는 노동자들 사이에서 정신적 피해를 호소하는 사람들이 증가하고 있다. 그러나 노동자들의 작업량을 줄이면 개발비용이 증가하게 되기 때문에 C사의 대표는 작업량을 그대로 유지하기로 했다.	
1. 위 상황과 관련하여 다음과 같은 주장을 할 수 있습니다. 각각의 주장에 어느 정도 동의하십니까?	
1-1	노동자들이 스스로 작업에 참여하고 있기 때문에 문제가 되지 않는다.
1-2	AI 서비스 개발과정은 윤리적이어야 한다.
2. 노동자들의 작업량을 지금처럼 유지하면 다음과 같은 결과가 예상됩니다. 노동자들의 작업량을 결정할 때, 각각의 결과를 얼마나 중요하게 고려해야 한다고 생각하십니까?	
2-1	저렴한 가격에 AI서비스를 제공할 수 있어 C사가 다른 업체에 비해 경쟁력을 가지게 되는 것
2-2	정신건강에 문제를 호소하는 노동자들이 발생하는 것
3. 위 상황과 관련하여 다음과 같은 감정이 들 수 있습니다. 각각의 감정에 어는 정도 공감하십니까?	
3-1	AI서비스 개발비용을 줄이기 위해 작업량을 줄이지 않은 C사 대표의 입장에 공감된다.
3-2	적은 돈을 받고도 정신적 피해가 예상되는 일을 해야 하는 노동자들이 안타깝다.
4. 위 상황과 관련하여 다음과 같은 행동을 할 수 있습니다. 당신이 C회사 대표라면 어떻게 하시겠습니까?	
4-1	개발비용을 줄이기 위해 노동자들의 작업량을 현재대로 유지한다.
4-2	개발비용이 증가하더라도 노동자들의 건강을 위해 노동자들의 작업량을 줄이겠다.