

초·중·고등학생의 인공지능 윤리의식의 성차 분석

김귀식 · 신영준*

경인교육대학교

An Analysis of Gender Differences in Primary, Middle and High School Students' Artificial Intelligence Ethics Awareness

Gwisik Kim · Youngjoon Shin*

Gyeongin National University of Education

Abstract : The purpose of this study is to analyze the gender differences of elementary, junior high, and high school students in the artificial intelligence ethics awareness (hereinafter referred to as AIEA). This is a study to investigate whether there is a gender difference in the AIEA, and if so, when the gender difference will occur. This study was conducted with 198 elementary school students (98 female students, 100 male students), 265 middle school students (166 female students, 99 male students), and 114 high school students (58 female students and 56 male students) in I Metropolitan City. The results are as follows: First, a gender difference in the AIEA between all boys and girls was confirmed. Second, the gender difference in the AIEA tended to be solidified as the school age increased from elementary school to middle school and high school. Third, female students at all stages of elementary school, junior high school, and high school are not yet very reliable in artificial intelligence, and there is a greater concern about non-discrimination than boys. It turns out that they have a negative position on permission to enter the territory. Fourth, the interaction effects of school age and gender have been identified in 'stability and reliability,' and in 'permit and limit' categories. Taken together, these results show that an educational strategy that approaches the gender equality perspective of the educational program is necessary so that there will be no gender difference in the AIEA during artificial intelligence education activities.

keywords : intelligence ethics awareness, gender difference, test for intelligence ethics awareness, cross-sectional study

I. 서론

우리 사회에서 인공 지능(Artificial Intelligence, AI)의 활용은 광범위하게 퍼져나가고 있으며 그 영역은 앞으로도 계속 확장될 전망이다. 과학교육이나 수학교육에서도 인공지능은 이미 깊숙이 들어오고 있고 고등학교 교육과정에 '인공지능수학' 과목도 새롭게 마련되었다. 인공지능이라는 용어는 2016년 인공지능 알파고의 등장으로 많은 사람들의 뇌리에 깊숙이 박히게 되었다. 인공지능은 인간의 학습능력, 추론능력, 지각능력, 논증능력, 자연언어의 이해능력 등을 인공적으로 구현한 컴퓨터 프로그램 또는 이를 포함한 컴퓨터 시스템으로 인간을 포함한 동물이 갖고 있는 지능 즉, natural intelligence와는 다른 개념이다(Kim,

2020; Park, 2018). 1956년 미국 Dartmouth conference에서 처음 사용된 인공지능이라는 용어는 이제 너무 익숙한 용어가 되어 버렸다(Kline, 2010). OECD (2019)는 컴퓨터 과학을 기초로 구현된 인공지능을 인간의 통찰력에 가까운 이성적이고 지능적인 사고 행위를 하는 것으로 미래 사회와 교육에 있어서 잠재력이 높은 기술로 평가하고 있다.

인공지능은 수행 능력 측면에서 다음과 같이 세 가지로 나뉜다. 첫째, 약한(특수 혹은 좁은) 인공지능 단계이다. 약한 인공지능이란 특정분야에서 특화된 형태의 지능으로 인간의 한계를 보완 및 생산성 향상을 위해 활용되는데, 자의식이 없는 인공지능이 이에 속한다. 둘째, 강한(강 혹은 범용)인공 지능단계이다. 이 단계의 인공지능은 인간이 할 수 있는 지적 업무를

* 교신저자: 신영준 (yjshin@ginue.ac.kr)

** 2021년 3월 23일 접수, 2021년 4월 15일 수정원고 접수, 2022년 4월 26일 채택
<http://dx.doi.org/10.21796/jse.2021.45.1.105>

이해하거나 학습이 가능한 기계나 기술로, 약한 인공지능보다 훨씬 더 진보된 단계로 한계 없이 다양한 영역에서 그 능력을 발휘할 수 있는 기계를 의미한다. 영화 'her'의 사만다처럼 자유로운 사고가 가능한 자아를 지닌 인공지능이라고 할 수 있으며, AI를 친구처럼 교감하면서 의인화 할 수 있을 것이다. 그러나 고유성으로 본 자율성에 대한 의문과 인간 존엄성의 위협, 로봇의 권리 인정 여부와 같은 여러 가지 철학적, 윤리적 이슈를 낳을 수 있다(Yun, 2020). 셋째, 초(슈퍼)인공지능 단계이다. 이 단계의 인공지능은 인류의 모든 지성을 합친 것보다 더 뛰어난 지적 능력을 가진 인간 이상의 능력을 보유한 인공지능이며, 효율, 자기보존, 자원획득, 창의성 같은 원초적 욕구를 기반으로 끊임없이 자가 발전을 할 수 있다. 그 시점을 특이점(Singularity)이라고 한다(Kurzweil, 2005; Lee & Lee, 2019). 인류는 인공지능에 기반을 둔 나노 로봇의 도움으로 영생을 누릴 수도, 멸종 혹은 재앙이 될 수도 있다고 예측했으며(Barrat, 2013), 초인공지능은 신의 영역이라 할 수 있는 단계이다(Kim, 2016).

인공지능의 확장성은 계층과 노소와 남녀와 관계없이 펼쳐지지만 경우에 따라서는 각 그룹에 따라 다른 방향으로 본의 아닌 차별을 불러일으킬 수 있다. 이러한 차별은 인공지능에 대한 윤리의식에도 영향을 미칠 수 있다. 윤리란 사람들이 서로 간 지켜야 할 바른 행위나 도리 또는 인간이 살아가는 방식을 의미하는데, Ji (2020)는 인공지능 윤리를 인공지능 시대를 살아가기 위해 국가를 초월한 개인과 기업과 사회, 그리고 로봇 간의 서로 지켜야 할 행위나 살아가는 방식, 원칙 및 기술이라 정의하고 있다.

인공지능 기술을 도입하는 과정에서 예기치 않은 문제점들이 노출되기도 한다. 2021년 인공지능 챗봇 'AI 이루다'에서의 혐오와 차별에 관한 부적절한 대화라든가, 2016년 마이크로소프트사의 인공지능 챗봇 'Tay'의 혐오나 차별 문제가 대두되었다. 기술의 오남용, 데이터 편향성과 고용, 개인정보 보호, 그리고 허용과 한계, 로봇의 권리까지 거의 모든 분야에 걸쳐 인공지능 알고리즘이 인간의 규범(가치)적 판단과 행위에 지대한 영향을 미칠 것으로 예상되면서 인공지능 윤리 이슈가 제기되고 있다(Byun, 2020; Kim, 2017). 인공지능 윤리에서 다루는 범위는 인공지능 연구, 개발, 적용, 폐기 등과 관련한 인공지능 기술의 전체 단계에서 인공지능을 둘러싼 규범적 문제라 할 수 있다(Heo *et al.*, 2020). 인공지능 윤리에 대한 인식은 인공지능 관련 윤리 가이드라인을 제정하도록 추동하였으며, 우리나라도 과학기술정보통신부와 정보통신정책연구원에서 인공지능 시대 바람직한 인공지능 개발·활용 방향을 제시하기 위한 '사람이 중심이 되는

인공지능(AI) 윤리기준'을 제정하였다(Kim, 2020; MSIT, 2020).

인공지능의 다양한 적용 과정에서 사이버 공간에서의 일탈 행위들이 윤리적 문제에 있어서 성차를 일으키기도 한다. 기존의 왜곡된 성차별적인 권력관계가 사이버 공간에도 그대로 투영되기도 한다. 초등학생의 경우에도 여학생이 사이버 불링의 피해를 더 많이 받는다는 보고도 있다(Lee *et al.*, 2014). 실제 세계와 마찬가지로 사이버 세계에서도 일탈행위에 대하여 성별에 따른 위험 지각의 차이가 발생하게 된다. 여성이 사이버 일탈 행위의 심각성을 강하게 지각할 뿐만 아니라 두려움 또한 강하게 느끼는 것으로 보고되고 있다. 남성 대비 여성은 사이버 일탈행위에 대한 심각성 지각 수준과 두려움 수준이 컸음에도 사이버 일탈 행위 피해 가능성에 대해서는 남성보다 크게 지각하지는 않는 것으로 나타났다(Kim, 2010).

사이버 세계는 인공지능이 광범위하게 활용되는 공간이다. 일찍이 Sung (1999)이 제기했던 기존의 성차별 이데올로기가 고스란히 사이버 세계에서도 그 위세를 떨칠 가능성이 있다. 딥 페이크와 같은 다양한 인공지능 활용 기술로 더 강화된 차별과 혐오로 인해 인공지능에 대한 긍정적인 윤리 의식보다 부정적인 윤리 의식이 강화될 가능성도 열려 있으며, 이러한 과정은 남녀 사이에 차별적으로 일어날 가능성도 있다. 결국 이러한 상황은 인공지능 윤리의식에도 영향을 미칠 가능성이 매우 크다. 다양한 분야에서 성차(gender gap)가 존재해 왔으며(Kim & Kim, 2020; Lee & Park, 2015; Song *et al.*, 2015), 이러한 성차를 극복하려는 국가적·사회적 노력도 사실 기울여 왔다(Lee & Lee, 2011; Park *et al.*, 2007). 여러 가지 맥락을 보면 인공지능 분야에서도 성차가 존재할 가능성은 상존하고, 특히 인공지능 윤리 이슈에 대해서도 남녀 차이가 존재할 가능성은 매우 높다.

인공지능 윤리에 대한 담론이 정리되어 가는 중에 학교 현장에는 2021년 2학기부터 고등학교 진로 선택 과목으로 '인공지능 기초'와 '인공지능 수학' 과목이 도입되며, 초중고교 2022년 개정 교육과정에는 '인공지능 교육'이 도입되어 프로그래밍, 인공지능 기초원리, 인공지능 활용, 인공지능 윤리 등이 포함될 예정이다(MOE, 2020). 인공지능 윤리교육이 진행되기 전에 실제로 학생들이 지니고 있는 인공지능 윤리의식이 어떠한가를 미리 진단해볼 필요가 있다. 또한, 인공지능 윤리의식에 대한 남학생과 여학생 사이의 차이가 존재하는지에 대해 확인해볼 필요도 있다. 따라서 본 연구에서는 초등학생과 중학생, 그리고 고등학생을 대상으로 남학생과 여학생사이의 인공지능 윤리의식에 있어서 성차가 존재하는 지를 알아보았다.

II. 연구 방법

1. 연구 대상

본 연구는 가급적 지역적 편차를 최소화하기 위하여 사회문화적 배경이 유사한 동일 지역인 I광역시에 위치한 초등학교, 중학교, 고등학교 총 세 학교급 학생을 연구 대상으로 횡단적 연구를 시행하였다. 초등학교(5~6학년)은 4개 남녀공학학교에서 표집한 남학생 100명과 여학생 98명으로 총 198명, 중학생(1~2학년)은 5개 남녀공학학교에서 표집한 남학생 99명, 여학생 166명으로 총 265명, 고등학생(1학년)은 남학교 1개교 학생 56명, 여학교 1개교 학생 58명으로 총 114명이다. 인공지능 윤리의식 조사는 2020년 10월 ~ 2021년 1월 초 사이의 13주간 진행하였으며, 연구 집단의 학생 구성은 Table 1과 같다.

2. 검사 도구 및 자료 처리

이 연구에서 사용한 검사 도구는 Kim & Shin (2021)이 개발한 인공지능 윤리의식 검사도구(TAIEA: Test for Artificial Intelligence Ethics Awareness)를 사용하였다(부록). 검사도구의 하위 범주는 책임성, 안정성 및 신뢰성, 차별 금지, 투명성 및 설명 가능성, 사람중심 서비스, 고용, 허용과 한계, 로봇의 권리 등 8개이다. 책임성 범주는 인공지능 정보 기술 및 서비스의 책임성이 누구에게 있는가에 대한 범주이다. 안정성 및 신뢰성 범주는 사용자, 자원 또는 환경에 피해가 되지 않도록 작업을 수행하고 의도하지 않은 결과와 시스템 작동 오류를 최소화하고 오작동에 대한 대비책을 마련하고 운영해야 하는 것에 대한 범주이다(Oh, 2020). 차별금지 범주는 인공지능에 의한 소수자 차별, 성적 차별에 대한 범주이다. 투명성 및 설명 가능성 범주는 인공지능 기술 개발, 서비스설계, 제품기획 등 의사결정 과정에서 이용자, 소비자, 시민 등의 의견을 반영하도록 노력해야 하며, 이용단계에서 예상되는 위험과 관련한 정보를 공개, 공유해

야 한다는 내용을 담은 범주이다(Bang, 2021; Byun, 2019; Lee, 2019). 사람중심 서비스 범주는 인공지능 시스템은 사람에게 해를 가하지 않고 인간의 타고난 기본적 요구를 보호하는 방향으로 개발되어야 한다는 내용의 범주이다. 고용 범주는 인공지능에 의한 노동 대체 현상에 관한 범주이다. 허용과 한계 범주는 인공지능 기술을 어디까지 허용하고 한계 지을지에 대한 범주이다. 로봇의 권리 범주는 로봇에게도 인간에 준하는 권리를 보장할 것인가에 대한 범주이다. 검사 도구는 각 하위 범주 당 3문항씩 모두 24문항으로 구성되어 있으며(Table 2), Likert 5점 평정척도로 측정하였다. 검사 도구의 범주별 신뢰도 분포는 .616 ~ .810을 보이고 있으며 전체 신뢰도 Cronbach's $\alpha = .812$ 이다(Kim & Shin, 2021). 초등학교, 중학교, 고등학교 남학생과 여학생의 인공지능 윤리의식에 대한 성별 차이가 있는지 알아보기 위하여 학교급별로 독립표본 *t*-검정을 실시했다. 학교급과 성별 상호 작용 효과를 검증하기 위해 분산분석을 실시하였으며, 자료의 통계 처리는 SPSS Win 26.0을 활용하였다.

III. 연구 결과

1. 인공지능 윤리의식 전체 성차 비교

초등학교, 중학교, 고등학교 학생들 전체를 대상으로 여학생과 남학생 간의 인공지능 윤리의식(AIEA: Artificial Intelligence Ethics Awareness)을 비교한 결과는 Table 3과 같다. 초등학교, 중학교, 고등학교 학생 전체를 대상으로 한 인공지능 윤리의식 평균은 남학생(3.31)이 여학생(3.20)보다 높았으며, 이는 통계적으로도 유의미한 차이였다($p < .001$). 이는 전체적으로 보았을 때 인공지능 윤리의식에 있어서 초·중·고등학교의 남학생과 여학생 간에 성차가 존재한다는 것을 의미한다.

초등학교, 중학교, 고등학교 학생들 전체를 대상으로 여학생($n = 322$)과 남학생($n = 255$) 간의 인공지능

Table 1. Participants of the study

대상	학년	성별		합계
		여학생	남학생	
초등학교	5~6학년	98	100	198
중학교	1~2학년	166	99	265
고등학교	1학년	58	56	114
계		322	255	577

Table 2. Items of survey questionnaire and reliability

하위범주	문항 번호	신뢰도(Cronbach's α)			
		초등	중학	고등	계
책임성	1*, 9, 17	.633	.610	.717	.657
안정성, 신뢰성	2, 10, 18	.807	.808	.815	.810
차별 금지	3*, 11, 19*	.773	.633	.762	.723
투명성, 설명가능성	4, 12, 20	.774	.856	.725	.808
사람중심서비스	5, 13, 21	.703	.716	.707	.715
고용	6, 14, 22	.785	.792	.759	.789
허용과 한계	7*, 15, 23	.588	.515	.723	.616
로봇의 권리	8, 16, 24	.797	.801	.807	.797
총 문항 수	24	.812	.806	.807	.812

* 역채점 문항

윤리의식에 대한 하위 범주별 비교한 결과는 Table 4와 같다. 8개의 하위 범주 중 통계적으로 의미 있는 수준에서 남학생이 여학생보다 더 높은 점수를 보인 것은 안정성 및 신뢰성, 차별금지, 사람중심서비스, 허용과 한계, 로봇의 권리 범주 등 5개 범주로 나타났다.

책임성 범주는 남녀 학생 간 차이를 보이고 있지 않으며, 비교적 높은 점수를 남학생과 여학생 모두 보이고 있어서 인공 지능 개발로 인한 책임성을 명확하게 해야 한다는 의식을 가지고 있었다. 안정성 및 신뢰성 범주는 남학생은 중간값 정도를, 여학생은 중간값 이하의 점수를 보였는데 이것은 여학생이 남학생들보다 인공지능에 대한 신뢰가 낮다는 것을 보여주고 있다($p < .001$). 중간값 근처의 점수를 보이고 있다는 것은 어쩌면 아직 학생들이 인공지능 기술의 안정성 및 신뢰성에 대한 확실한 자기 입장을 가지고 있지 않은 유보적 태도로 말미암은 것일 수도 있다. 차별금지 범주의 경우 점수가 높을수록 차별이 없는 공평함을 의미하는데 남학생이나 여학생 모두 다른 하위 범주에 비해 낮은 점수를 보이고 있다. 이것은 학생들은 인공지능이 유발할 차별에 대해 우려하고 있으며 그 우려의 정도가 여학생이 남학생에 비해 높다는 것으로 해석할 수 있다($p < .001$).

투명성과 설명 가능성 범주는 다른 범주에 비해 점수가 높게 나타났으며, 남학생이나 여학생 모두 인공지능에 대해 설명 가능하고 모든 것이 투명해야 한다는 인식을 함께 하고 있음을 알 수 있다. 사람중심 서

비스 범주의 경우 중간값 정도의 반응을 보이고 있다. 사람의 편의성을 위해 인공지능이 사용되어야 한다는 의식은 중간 정도라는 것을 의미한다. 여기서 중간 정도의 의식이란 다른 말로 표현하면 아직 확고한 신념이나 의식보다는 유보적인 태도를 취하고 있음을 동시에 내포하고 있다고 생각하는 것이 옳을 것이다. 이 범주에서는 남학생이 여학생보다는 좀 더 사람의 편의성에 활용되어야 한다는 생각을 더 가지고 있음을 알 수 있다($p < .01$).

인공지능의 노동자 대체로 발생할 수 있는 고용 문제에 대해 비교적 남학생이나 여학생 모두 높은 점수를 보이고 있어, 인공지능에 의한 노동 대체 가능성을 받아들인다고 할 수 있다. 인공지능을 인간의 영역을 기준으로 어디까지 허용할 것인가에 대한 문제는 여학생의 경우 허용보다는 보류 혹은 유보적인 태도를 보인다고 해석할 수 있는 중간값 이하의 반응을 보였다. 남학생이 여학생보다는 통계적으로 의미 있는 수준에서 허용감이 높기는 하지만 중간값 근처의 값은 완전한 허용이기보다는 아직 자신들의 성향을 드러내지 않는 유보적인 태도에 가깝기 때문에 쉽게 허용 여부를 판단하기는 어렵다. 높은 수준의 허용 의식이라고 할 수 있다($p < .01$). 로봇에게 인간에 가까운 권리를 허용할 것인지에 대한 문제는 아직 학생들이 쉽게 동의하지 않고 있고 유보적인 태도를 취하고 있음을 알 수 있다. 이러한 태도의 정도는 여학생이 더 높다고 할 수 있다($p < .05$).

Table 3. The result of t -test on AIEA by gender

집단	평균	표준편차	t	p
여학생($n = 322$)	3.20	.299	-3.828	.000***
남학생($n = 255$)	3.31	.419		

*** $p < .001$

Table 4. The result of *t*-test on AIEA in 8 categories

영역	집단	평균	표준편차	<i>t</i>	<i>p</i>
책임성	여학생	3.63	.660	.201	.841
	남학생	3.62	.695		
안정성 및 신뢰성	여학생	2.74	.694	-4.355	.000***
	남학생	3.03	.887		
차별금지	여학생	2.26	.690	-3.550	.000***
	남학생	2.48	.761		
투명성, 설명가능성	여학생	4.10	.669	-.054	.957
	남학생	4.10	.726		
사람중심서비스	여학생	3.02	.611	-3.059	.002**
	남학생	3.20	.785		
고용	여학생	4.02	.686	1.359	.175
	남학생	3.93	.764		
허용과 한계	여학생	2.88	.589	-3.095	.002**
	남학생	3.04	.680		
로봇의 권리	여학생	2.92	.812	-2.578	.010*
	남학생	3.11	.915		

* $p < .05$, ** $p < .01$, *** $p < .001$

2. 인공지능 윤리의식 학교급별 성차 비교

인공지능 윤리의식에서 성에 따른 차이가 존재하는지를 알아보기 위해 학교급별로 남학생과 여학생의 성차 분석을 실시하였다. 본 연구는 횡단적 연구방식이어서 성차 발생 시기를 종단적 연구처럼 직접적으로 알기는 어렵지만 간접적으로 시사 하는 바를 얻을 수 있다.

1) 초등학교

초등학교 여학생($n = 98$)과 남학생($n = 100$) 등 모두 198명을 대상으로 인공지능 윤리의식의 하위 범주별 성에 따른 차이를 알아본 결과는 Table 5와 같다. 초등학교 5, 6학년 남녀 학생들을 대상으로 한 인공지능 윤리의식 평균은 남학생(3.22)이 여학생(3.13)보다 조금 더 높았으나 이는 통계적으로는 유의미한 차이는 아니다($p > .05$). 이는 인공지능 윤리의식에 있어서 초등학교 남학생과 여학생 간에는 비교적 아직 성에 따른 차이가 거의 나타나고 있지 않음을 시사 하는 것이라고 할 수 있다.

8개의 범주 중 차별금지 범주에서는 남학생(2.50)이 여학생(2.28)보다 높은 점수를 보이고 있는 데 이 범

주에서만 유일하게 성차가 발생했다. 여학생의 경우 중간값 이하의 범주는 안정성 및 신뢰성, 차별금지, 사람중심 서비스, 허용과 한계, 로봇의 권리 등 5개 범주였다. 이에 비해 남학생의 경우는 안정성 및 신뢰성, 차별금지, 허용과 한계 등 3개의 범주였다.

중간값 이하를 보이는 범주들을 보면, 안정성 및 신뢰성의 경우 초등학교 남녀학생 모두 중간값 아래의 점수를 보여주고 있어 인공지능에 대한 신뢰가 아직 굳건하지는 않고 유보적인 태도를 취하고 있다는 것을 알 수 있다. 차별금지 범주에서 볼 수 있는 것처럼 인공지능으로 인한 차별에 대한 우려가 남학생과 여학생 모두 가지고 있지만 여학생의 우려가 남학생의 우려보다도 더 크다는 것을 알 수 있다($p < .05$). 비록 성차가 있다고 하더라도 남학생이나 여학생 모두 그 값이 아주 낮거나 높지 않은 중간값 정도인 것으로 보아 차별금지에 대해 유보적 입장을 지니고 있다고 할 수 있다. 인공지능의 허용과 한계 범주도 남녀학생 모두 중간값 이하를 보이고 있다. 아직은 예술 활동 등에서 인간의 영역에 들어오는 것을 쉽게 초등학교들은 받아들이지 못하고 유보적인 태도를 보이고 있다고 할 수 있다.

중간값 이상을 보이는 범주들 중 책임성 범주는 비교적 높은 점수를 보이고 있어 인공지능 개발로 인한 책임성을 명확하게 해야 한다는 의식이 남녀 학생

Table 5. The result of *t*-test on AIEA in primary school students

영역	집단	평균	표준편차	<i>t</i>	<i>p</i>
책임성	여학생	3.58	.608	.500	.618
	남학생	3.53	.652		
안정성 및 신뢰성	여학생	2.70	.717	-1.040	.300
	남학생	2.81	.847		
차별금지	여학생	2.28	.712	-2.031	.044*
	남학생	2.50	.814		
투명성, 설명가능성	여학생	3.93	.646	-1.236	.218
	남학생	4.05	.696		
사람중심서비스	여학생	2.91	.560	-1.291	.198
	남학생	3.03	.752		
고용	여학생	3.86	.744	-.180	.858
	남학생	3.88	.779		
허용과 한계	여학생	2.87	.576	-.032	.975
	남학생	2.87	.608		
로봇의 권리	여학생	2.88	.792	-1.376	.170
	남학생	3.05	.897		
전체	여학생	3.13	.290	-1.789	.075
	남학생	3.22	.409		

**p* < .05

차이 없이 어느 정도 세워진 것으로 여겨진다. 투명성 및 설명 가능성 범주도 비교적 높은 값을 나타내는데 이는 초등학생들은 남녀 학생 관계없이 인공지능에 의한 다양한 사례들이 투명하게 공개해야 한다는 데 동의하고 있음을 알 수 있다. 인공지능으로 인한 고용의 문제도 비교적 인공지능으로 대체되는 고용 문제를 받아들이고 있다는 것을 알 수 있다. 로봇의 권리 문제도 아직 여학생의 경우는 중간값 이하를 나타내고 있고, 남학생도 중간값 정도를 부여하고 있어 로봇의 권리를 인간의 권리처럼 부여하는 것을 쉽사리 받아들이지 않고 입장을 완전하게 정하고 있지 않는 것을 알 수 있다.

2) 중학생

중학교 여학생(*n* = 166)과 남학생(*n* = 99) 등 모두 265명을 대상으로 인공지능 윤리의식의 하위 범주별 성에 따른 차이를 알아본 결과는 Table 6과 같다. 중학교 남녀 학생들을 대상으로 한 인공지능 윤리의식 평균은 남학생(3.34)이 여학생(3.21)보다 더 높았으며 이는 통계적으로는 유의미한 수준의 차이이다(*p* < .01). 이는 인공지능 윤리의식에 있어서 초등학교 때에 나타나지 않은 성차가 중학생에서 나타난 것으로 사춘기를

겪으면서 함께 따라온 윤리의식의 변화로 추정된다.

8개의 범주 중 중학교 남녀 학생 사이에서 통계적으로 유의미한 성차를 보인 것은 안정성 및 신뢰성, 사람중심 서비스, 허용과 한계 등 3개의 범주였다. 초등학교에서 성차를 유일하게 보였던 것이 1개였던 것에 비해 3개로 늘어났다. 다만 초등학교에서 성차를 보였던 차별금지 범주의 경우 중학생에서는 성차를 보이지 않았다.

안정성 및 신뢰성 범주에서 여학생은 중간값 아래의 점수(2.77)를 보이면서 남학생보다 인공지능에 대한 신뢰성이 낮음을 알 수 있다. 사람중심 서비스의 경우 남학생과 여학생 모두 중간값 이상을 보이고 있지만 남학생이 여학생보다 통계적으로 유의미한 수준에서 더 높은 점수를 보이고 있다. 이것은 사람의 편의성을 위해 인공지능을 활용하는 것에 남학생이 여학생보다 더 우호적이라는 것을 의미한다. 허용과 한계 범주에서 여학생은 중간값 이하의 점수를 보이고 있어 인간의 영역에 인공지능이 넘어서는 것을 남학생보다 훨씬 허용해서는 안 된다는 생각을 지니고 있음을 알 수 있다. 그러나 중간값 정도를 보이고 있는 이들 모두에 대해 아직 완전한 자기 성향이 정해졌다고 해석하기보다는 유보적 태도를 지니고 있다고 해석하는 것이 더 타당하다고 여겨진다.

Table 6. The result of *t*-test on AIEA in middle school students

영역	집단	평균	표준편차	<i>t</i>	<i>p</i>
책임성	여학생	3.53	.617	-.707	.480
	남학생	3.58	.645		
안정성 및 신뢰성	여학생	2.77	.675	-3.471	.001**
	남학생	3.10	.861		
차별금지	여학생	2.39	.674	-1.638	.103
	남학생	2.51	.535		
투명성, 설명가능성	여학생	4.09	.698	.024	.981
	남학생	4.08	.762		
사람중심서비스	여학생	3.04	.602	-2.662	.009**
	남학생	3.28	.770		
고용	여학생	3.98	.653	.474	.636
	남학생	3.94	.732		
허용과 한계	여학생	2.89	.550	-2.202	.029*
	남학생	3.04	.560		
로봇의 권리	여학생	3.01	.757	-1.390	.166
	남학생	3.15	.891		
전체	여학생	3.21	.293	-2.972	.003**
	남학생	3.34	.395		

p* < .05, *p* < .01

비록 성차는 나타나지 않지만 차별 금지 범주에서 남학생과 여학생 모두 낮은 점수를 나타내고 있는 중학생들도 초등학생들과 마찬가지로 인공 지능이 야기할 수 있는 차별을 여전히 우려하고 있음을 알 수 있다. 인공지능으로 인해 발생한 윤리적, 법적 책임성에 대해서도 중학생은 비교적 개발자나 사용자에게 있다는 생각을 지니고 있음을 알 수 있다. 인공지능에 대한 투명한 설명과 공개에 대해 중학생들도 초등학생과 마찬가지로 높은 윤리의식 수준을 지니고 있음을 알 수 있다. 고용의 문제도 초등학생들과 마찬가지로 비교적 인공지능으로 대체되는 것을 받아들인다고 할 수 있다. 로봇의 권리 범주의 경우 비록 점수 향상은 있었으나 초등학생들과 마찬가지로 아직 선뜻 인간의 권리처럼 부여하는 것을 받아들이고 있지 않다는 것을 알 수 있다.

3) 고등학생

고등학교 여학생(*n* = 58)과 남학생(*n* = 56) 등 모두 114명을 대상으로 인공지능 윤리의식의 하위 범주별 성에 따른 차이를 알아본 결과는 Table 7과 같다. 고등학교 남녀 학생들을 대상으로 한 인공지능 윤리

의식 평균은 남학생(3.45)이 여학생(3.27)보다 더 높았으며 이는 통계적으로는 유의미한 수준의 차이이다 (*p* < .05). 중학생에서 나타난 인공지능 윤리의식에 있어서 성차가 여전히 존재하고 있음을 시사 한다.

8개의 범주 중 고등학교 남녀 학생 사이에서 통계적으로 유의미한 성차를 보인 것은 안정성 및 신뢰성, 차별 금지, 고용, 허용과 한계, 로봇의 권리 등 5개의 범주였다. 초등학교에서 1개, 중학교에서 3개의 성차를 보였던 것에 비해 고등학교에서는 5개로 늘어났다.

먼저 성차를 보인 5개의 범주를 보면, 안정성 및 신뢰성 범주에서 여학생은 중학교 여학생 점수(2.77)와 마찬가지로 여전히 중간값 아래의 점수(2.73)를 보이고 있으며, 남학생의 경우는 중학생(3.10)보다 0.22점 상승하였다. 인공지능에 대한 신뢰가 남학생의 경우 어느 정도 향상되었지만 여학생의 경우 남학생에 비해 인공지능에 대해 여전히 신뢰하지 못하고 유보적인 태도를 취하고 있음을 알 수 있다. 차별금지의 경우 고등학교 여학생의 점수는 1.87로 모든 범주의 점수 중 가장 낮았다. 물론 남학생의 경우도 중간값 이하를 보이고 있지만 여학생의 경우 인공 지능으로 인해 발생할 가능성이 있는 차별을 매우 우려하고 있음을 알 수 있다. 고용의 문제도 초등학생이나 중학생과

Table 7. The result of *t*-test on AIEA in high school students

영역	집단	평균	표준편차	<i>t</i>	<i>p</i>
책임성	여학생	4.01	.728	1.233	.220
	남학생	3.83	.814		
안정성 및 신뢰성	여학생	2.73	.719	-3.795	.000***
	남학생	3.32	.920		
차별금지	여학생	1.87	.551	-3.361	.001**
	남학생	2.37	.976		
투명성, 설명가능성	여학생	4.44	.484	1.704	.091
	남학생	4.24	.706		
사람중심서비스	여학생	3.14	.696	-1.581	.117
	남학생	3.37	.823		
고용	여학생	4.37	.550	2.784	.006**
	남학생	4.01	.796		
허용과 한계	여학생	2.90	.718	-3.176	.002**
	남학생	3.37	.866		
로봇의 권리	여학생	2.73	.961	-2.090	.039*
	남학생	3.12	.998		
전체	여학생	3.27	.309	-2.541	.012*
	남학생	3.45	.439		

* $p < .05$, ** $p < .01$, *** $p < .001$

마찬가지로 인공지능으로 대체되는 것을 대체로 받아들이고 있다. 다만 고등학생의 경우 여학생이 남학생에 비해 훨씬 유연한 태도를 보이고 있다는 점이다. 허용과 한계 범주에서 여학생은 중학생 점수(2.89)와 마찬가지로 여전히 중간값 이하의 점수(2.90)를 보이고 있다. 남학생의 경우 중학생의 경우(3.04)보다 0.33점 비교적 많이 상승한 점수를 보여주고 있다. 이는 고등학교 여학생의 경우도 인간의 영역에 인공지능이 넘어서는 것을 남학생보다 훨씬 허용해서는 안 된다는 생각을 여전히 지니고 있거나 최소한 인간의 영역에 넘어오는 것에 대해 유보적인 태도를 가지고 있다고 할 수 있다. 이는 로봇의 권리 범주에서도 여실히 드러난다. 여학생의 경우 로봇의 권리에 대해 중간값 이하의 점수를 보이고 있어 남학생에 비해 인간의 권리 수준으로 생각하기 어렵다는 것을 보여주고 있다.

비록 성차는 나타나지 않지만 책임성, 투명성 및 설명 가능성 범주에서 여학생이 남학생보다 더 높은 점수를 보이고 있으며, 사람 중심 서비스 범주에서는 중학생과 마찬가지로 남학생과 여학생 모두 중간값 이상을 보이고 있고 남학생이 여학생보다 더 높은 점수

를 보이고 있다. 고등학생들도 사람의 편의성을 위해 인공지능을 활용하는 것을 일정 정도 받아들이고 있다고 할 수 있다. 고등학생들도 남녀학생 모두 인공지능으로 인해 발생한 윤리적, 법적 책임성이 개발자나 사용자에 있다는 생각을 지니고 있고, 인공지능에 대한 투명한 설명과 공개를 해야 한다는 생각을 가지고 있음을 알 수 있다.

초등학교에서 거의 나타나지 않던 인공지능 윤리의식에 대한 성차가 학령이 올라가면서 점점 더 많이 나타나는 것은 생물학적 성숙이라기보다는 사회문화적 관점에서 접근하는 것이 옳다(Park & Woo, 2008). 과학에서의 성차가 과학을 남성 중심적이거나 편향적 경향 때문이라는 주장(Barton, 1998)처럼 인공지능에 대한 담론이 혹시 남성 편향적이기 때문에 성에 대한 의식이 깨어나는 것과 비례하여 성차가 더 많이 발생하는 것은 아닌지 돌아봐야 한다. 2021년의 'AI 이루다'의 혐오 발언이나 AI 딥페이크 등이 남학생들보다 여학생들에게 더 큰 영향을 미쳤을 것이다. 과학과 정보통신기술에 대한 위험 인식에 대해 여성은 남성보다 더욱 민감하게 반응하는 것으로 조사되었다(Yoo, 2018)는 사실은 여성과 남성에 미치는 영향에 차이가

날 수 있음을 시사하기도 한다. 한편으로 20세기 후반의 남녀 차이의 오래된 성차 연구들이 수학이나 과학 능력에 성차가 있다는 연구를 보여 주었지만 최근의 연구들은 남녀 사이의 유사성에 대해 연구하는 경향이 있다(Yoon, 2016)는 사실에 주목할 필요가 있다. 수학이나 과학처럼 알고리즘이 많이 포함된 분야인 인공지능의 경우도 20세기 후반의 연구를 답습하지 말고 남녀의 유사성에 방점을 찍고 있는 최근의 연구 경향과 궤를 같이 해야 할 것이다. 이러한 접근은 결국 인공지능 윤리의식에 성차를 유발하는 상황이 아닌 성차가 없는 상황으로 연결될 것이기 때문이다. 성차가 이미 발생한 경우는 성차를 줄이거나 없애는 방향으로 국가적 사회적 노력을 해야 할 것이며, 수학이나 과학에서의 성차 극복의 관점도 참고할만하다(Lee & Lee, 2011; Park *et al.*, 2007).

4) 학교급 및 성별 상호 작용

학교급별로 분리하여 성별로 나타나는 차이 분석에서 더 나아가 학교급과 성별의 상호작용 효과를 확인하기 위하여 이원분산분석을 실시하였다. 그 결과 인공지능 윤리의식 8개 하위 범주 중에서 학령과 성별의 상호작용 효과가 통계적으로 유의미하게 나타난 것은 안정성 및 신뢰성 범주(Table 8)와 허용과 한계 범주(Table 9)였다.

안정성 및 신뢰성의 경우, 학교급별 차이와 성별 차이, 그리고 상호작용 효과가 모두 통계적으로 유의미

하였다. 가장 낮은 점수는 초등학교 여학생(2.70)이었으며, 가장 높은 점수를 보인 것은 고등학교 남학생(3.32)이었다. 여학생의 경우 초 - 중 - 고의 점수가 2.70 - 2.77 - 2.73 정도였으며, 이는 여학생들은 대체로 인공지능의 안정성과 신뢰성에 대해 동의하지 못하고 유보적인 태도를 취하고 있음을 알 수 있다. 남학생의 경우는 초 - 중 - 고의 점수가 2.81 - 3.10 - 3.32의 분포를 보이고 있으며, 학령이 올라갈수록 신뢰성과 안정성의 증가를 확인할 수 있었다. 전체적으로 보면 학령의 증가와 성별은 서로 상호 작용을 하여 인공지능의 안정성과 신뢰성 형성에 영향을 미치고 있음을 알 수 있다.

허용과 한계 범주 경우도 안정성과 신뢰성과 마찬가지로 학교급별, 성별, 상호작용 효과가 모두 통계적으로 유의미한 차이였다. 가장 낮은 점수는 초등학교 남학생과 여학생의 점수로 2.87이었으며, 가장 높은 점수는 고등학교 남학생이 3.37을 보여주고 있다. 여학생의 경우 초 - 중 - 고의 점수분포를 보면 2.87 - 2.89 - 2.90으로 거의 점수 차이가 나타나고 있지 않았다. 남학생의 경우 초 - 중 - 고의 점수분포는 2.87 - 3.04 - 3.37로 학령이 올라갈수록 여학생에 비해 상대적으로 큰 폭으로 증가하고 있음을 알 수 있다. 남학생의 이러한 증가는 학령과 성별 차이의 상호 작용의 효과로 나타난 것이라고 할 수 있다. 특히 여학생이 더 유보적인 태도를 보이는 것은 인터넷 상에서의 성의 상품화, AI 답페이크, AI 이루다와 같은 챗봇의 혐오 발언 등이 영향을 미쳤을 것으로 판단된다. 그러나

Table 8. The result of two way ANOVA in 'Stability and Reliability' category

변량원	제곱합	자유도	평균제곱	F	p
학교급별	6.041	2	3.021	5.004	.007**
성별	14.917	1	14.917	24.712	.000***
학교급별×성별	4.050	2	2.025	3.355	.036*
오차	344.668	571	.604		

*p < .05, **p < .01, ***p < .001

Table 9. The result of two way ANOVA in 'Permit and Limit' category

변량원	제곱합	자유도	평균제곱	F	p
학교급별	5.047	2	2.523	6.546	.002**
성별	5.557	1	5.557	14.415	.000***
학교급별×성별	4.005	2	2.002	5.194	.006**
오차	220.113	571	.385		

*p < .05, **p < .01, ***p < .001

이 두 범주에서 유독 이러한 결과가 나오는 것은 남학생과 여학생이 성장하는 과정에서 인공지능에 대한 사고에 영향을 미치는 요인들에 차이가 있는지 즉, 사회문화적 접근과 교육 행태 등에 차이가 있는지 확인해볼 필요가 있다.

IV. 결론 및 제언

우리 사회에서 인공 지능의 활용은 광범위하게 활용되고 있으며 그 영역은 앞으로도 계속 확장될 전망이다. 인공지능의 확장성은 계층과 노소와 남녀와 관계없이 펼쳐지지만 경우에 따라서는 다른 방향으로 본의 아닌 차별을 불러일으킬 수 있다. 이러한 차별은 인공지능에 대한 윤리의식에도 영향을 미칠 수 있다. 이런 관점에서 본 연구는 인공지능 윤리의식에 대한 남학생과 여학생 사이의 차이를 알아보기 위해 실시한 것이다. 본 연구는 사회문화적 배경이 동일한 한 지역을 골라(I광역시) 초등학교, 중학교, 고등학교 학생 577명을 대상으로 진행되었다. 본 연구를 통해 얻어진 결론은 다음과 같다.

첫째, 전체 남학생과 여학생 사이의 인공지능 윤리의식의 성차가 확인되었다. 본 연구에서 활용한 검사 도구는 인공지능 윤리의식을 8개 범주로 나누어 조사하도록 되어 있다. 8개의 하위 범주 중 안정성 및 신뢰성, 차별금지, 사람중심서비스, 허용과 한계, 로봇의 권리 범주 등 5개 범주에서 성차가 나타났으며, 통계적으로 의미 있는 수준에서 남학생이 여학생보다 더 높은 점수를 보였다.

둘째, 인공지능 윤리의식의 성차는 초등학교에서 중학교, 고등학교로 학령이 높아질수록 확고해지는 경향이 있었다. 인공지능 윤리의식 전체 점수를 기준으로 초등학교 단계에서는 성차가 나타나지 않았으나 중학교와 고등학교 단계에서는 통계적으로 의미 있는 수준에서 성차가 나타났다. 8개 범주 중에서 초등학교 단계에서는 차별금지 범주 1개에서만 성차가 나타났고, 중학교 단계에서는 안정성 및 신뢰성, 사람중심서비스, 허용과 한계 등 3개의 범주로 증가하였다. 고등학교 단계에서는 중학교 단계보다 차별 금지와 로봇의 권리 범주 2개가 더 늘어나 5개 범주에서 성차가 나타났다.

셋째, 초등학생, 중학생, 고등학생 모든 단계에서 여학생은 아직 인공지능에 대한 신뢰성이 그리 높지 않고, 차별 금지에 대한 우려가 남학생에 비해 크며, 예술 등의 분야에서 인공지능의 인간 영역에의 진입 허용에 대해 부정적인 입장을 가지고 있는 것으로 파악되었다. 이러한 여학생의 관점은 인터넷 상에서의 성의 상품화, AI 답페이지, AI 이루다와 같은 챗봇의

혐오 발언 등이 영향을 미쳤을 것으로 판단된다.

넷째, 학령과 성별의 상호 작용 효과를 검증해본 결과 안정성과 신뢰성, 허용과 한계 범주에서 상호 작용 효과가 있음이 확인되었다. 특히 여학생은 학령이 올라가도 인식의 변화는 거의 없으나, 남학생의 경우 인식의 변화가 크다는 것에 주목해야 하며, 이러한 결과를 가져온 사회문화적 영향을 면밀히 분석해볼 필요가 있다.

본 연구 결과는 사회문화적 배경이 비슷한 특정 지역에서 조사한 내용을 바탕으로 이루어졌다는 한계가 있기는 하지만 그 점을 감안하면서 연구 결과를 기반으로 다음과 같은 몇 가지 제언을 하고자 한다. 첫째, 인공지능 윤리의식의 성차를 가져온 근본 원인에 대한 총체적인 연구가 필요하다. 성차 극복을 위한 정확한 방안을 세우기 위해서 필요하다. 둘째, 인공지능 윤리의식은 학생들에게만 적용되어서는 안 된다. 향후 일반 성인들에 대해서도 인공지능 윤리의식을 조사해볼 필요가 있다. 셋째, 인공지능 윤리의식의 성차 분석을 기초 자료로 하여 인공지능 윤리의식을 고양하기 위한 교수학습 자료 개발 시 양성 평등의 관점을 적용하도록 추진하는 방안이 필요하다.

국 문 요 약

이 연구의 목적은 인공지능 윤리의식에 대한 초등학교, 중학교, 고등학교 학생들의 성차를 분석하는 것이다. 과연 인공지능 윤리의식에 성차가 존재하는지, 존재한다면 언제부터 성차가 발생하는지를 알아보기 위한 연구이다. 본 연구는 I 광역시 초등학교 198명(여학생 98명, 남학생 100명), 중학생 265명(여학생 166명, 남학생 99명), 고등학생 114명(여학생 58명, 남학생 56명)을 대상으로 실시하였다. 본 연구 결과는 다음과 같다. 첫째, 전체 남학생과 여학생 사이의 인공지능 윤리의식의 성차가 확인되었다. 둘째, 인공지능 윤리의식의 성차는 초등학교에서 중학교, 고등학교로 학령이 높아질수록 확고해지는 경향이 있었다. 셋째, 초등학생, 중학생, 고등학생 모든 단계에서 여학생은 아직 인공지능에 대한 신뢰성이 그리 높지 않고, 차별 금지에 대한 우려가 남학생에 비해 크며, 예술 등의 분야에서 인공지능의 인간 영역에의 진입 허용에 대해 부정적인 입장을 가지고 있는 것으로 파악되었다. 넷째, 학령과 성별의 상호 작용 효과는 안정성 및 신뢰성, 그리고 허용과 한계 범주에서 확인되었다. 이러한 결과들을 종합해 볼 때, 인공지능 교육 활동 시 인공지능 윤리의식에 성차가 생기지 않도록 교

육 프로그램을 양성 평등적으로 접근하는 교육적 방안이 필요하다고 할 수 있다.

주제어: 인공지능 윤리의식, 성차, 인공지능 윤리의식 검사도구, 횡단적 연구

References

- Bang, J. M. (2021). The transition of regulatory governance on AI algorithms - Focused on US algorithmic regulations and AI ethical principles. *Public Law*, 49(3), 375-406.
- Barrat, J. (2013). *Our final invention: Artificial intelligence and the end of the human era*. New York, NY: Thomas Dunne Books/St Martin's Press.
- Barton, A. C. (1998). *Feminist science education*. Teachers College Press.
- Byun, S. Y. (2019). A study on the ethics certification program based on the morality types of AI robots. *Journal of Ethics*, 126, 73-90.
- Byun, S. Y. (2020). A study on the necessity of AI ethics education. *The Journal of Korea elementary education*, 31(3), 153-164.
- Heo, E. S., Lee, Y. H., & Shin, J. W. (2020). Why ethics is: A landscape of modern AI ethics debate, Its features and limitations. *Human Beings, Environment and Their Future*, 24, 165-209.
- Ji, H. A. (2020). *A study on the application of roboethics to moral education* (Doctoral Dissertation). Seoul National University.
- Kim, D. H. (2016). Forming and indicating a christian theological discourse on AI. *Theological Studies*, 68, 35-60.
- Kim, G. S., & Shin, Y. J. (2021). Study on the development of test for artificial intelligence ethical awareness. *Journal of The Korean Association of Artificial Intelligence Education*, 2(1), 1-19.
- Kim, H. G., & Kim, Y. S. (2020). Meta-analysis of gender difference in performance on the information subtest of the Korean-Wechsler intelligence scale. *The Korean Journal of Rehabilitation Psychology*, 27(4), 151-163.
- Kim, J. H. (2010). Are women more sensitive than men to the risk of cyber victimizations?: Health communication perspective. *Health Communication Research*, 2(2), 155-180.
- Kim, J. M. (2020). Artificial intelligence algorithm regulation status, recent trends, and legal implications - Focusing on the issue of artificial intelligence bias. *The Digital Ethics*, 4(2), 27-42.
- Kim, M. J. (2017). The necessity of artificial intelligence ethics and trends in Korea and abroad. *Journal of The Korean Institute of Communication Sciences*, 34(10), 45-54.
- Kline, R. (2010). Cybernetics, automata studies, and the Dartmouth conference on artificial intelligence. *IEEE Annals of the History of Computing*, 33(4), 5-16.
- Kurzweil, R. (2005). *The singularity is near*. New York, NY: Viking.
- Lee, A. R., Lee, Y. J., & Yang, H. I. (2014). The effects of cyber bullying and bullied experience on upper elementary students' aggression and verbal aggression. *Korea Journal of Counseling*, 15(6), 2437-2450.
- Lee, C. S., & Lee, H. J. (2019). Expectations and anxieties affecting attitudes toward artificial intelligence revolution. *Journal of the Korea Contents Association*, 19(9), 37-46.
- Lee, E. J., & Lee, K. H. (2011). A study on the factors influencing gender differences changes of Korean students in PISA mathematics assessment. *Journal of Educational Research in Mathematics*, 21(4), 313-326.
- Lee, J. W. (2019). Can we impose responsibilities on artificial intelligence? To seek accountability-oriented ethics for artificial intelligence. *Korean Journal for the Philosophy of Science*, 22(2), 70-104.
- Lee, K. S., & Park, I. Y. (2015). Characteristics on gender difference of Korean students in TIMSS mathematics assessment. *The Journal of Curriculum and Evaluation*, 18(1), 155-183.

- Ministry of Education [MOE]. (2020). *Educational policy directions and key tasks in the age of artificial intelligence*. Press Release, Nov., 20. 2020,
- Ministry of Science and ICT [MSIT]. (2020). *Human-centered ethical standards for artificial intelligence*. Press Release, Dec., 23. 2020,
- Organisation for Economic Co-operation and Development [OECD]. (2019). *Trends shaping education 2019*. Paris: OECD Publishing, DOI: 10.1787/trends_edu-2019-en.
- Oh, T. W. (2020). EU policy for AI ethics from white paper on artificial intelligence. *The Digital Ethics*, 4(1), 22-31.
- Park, A. C., & Woo, C. Y. (2008). The relationship among high school students' attachment for their parents and peers depending on gender, self-identity, and career decision-making levels. *The Korean Journal of Educational Psychology*, 22(1), 69-85.
- Park, C. J., Dong, H. K., & Shin, Y. J. (2007). An analysis of preferences for science and the role gender differences plays in determining preferences for It amongst elementary school students. *Journal of Korean Elementary Science Education*, 26(2), 216-225.
- Park, J. Y. (2018). Trend analysis of artificial intelligence technology using patent information. *Journal of the Korea society of computer and information*, 23(4), 9-16.
- Song, M. Y., Im, H. J., Rim, H. M., Park, H. Y., & Ku, J. O. (2015). Educational factors influencing the gender difference in PISA. *The journal of Educational Studies*, 46(4), 99-122.
- Sung, S. J. (1999). Cyberface and gender difference ideology. *Journal of Women Studies*, 10, 101-116.
- Yoo, E. H. (2018). Online privacy, technologies, and perceived risk. *The Journal of Social Science*, 25(2), 82-100.
- Yoon, S. H. (2016). *The shaping and changes of scientific discourses: Focusing on the discourses on sex/gender differences in neuroscience* (Doctoral Dissertation). Seoul National University.
- Yun, S. J. (2020). The allegory of AI and empathy in the movie Her. *The Journal of Image and Cultural Contents*, 19, 213-236.

저 자 정 보

김 귀 식

(경인교육대학교
융합인재교육센터 선임연구원)

신 영 준

(경인교육대학교 교수)

<부록> 인공지능 윤리의식 검사 도구

문항 번호	내용	범주
1*	나는 인공지능이 운전하여 움직이는 자율주행차가 피할 수 없는 사고가 일어났을 때, 보행자보다 운전자를 최우선으로 보호해야 한다고 생각한다.	책임성
2	나는 인공지능 기술은 대부분 믿을 수 있고, 안전하다고 생각한다.	안정성, 신뢰성
3*	나는 인간처럼 스스로 생각과 판단을 하는 강한 인공지능에 의해서도 또 다른 형태의 차별이나 불공평한 일이 생길 수 있다고 생각한다.	차별금지
4	나는 국가나 단체, 사회, 모든 사람의 이익을 위해서 인공지능기술을 개발하고 데이터를 모으는 일을 할 때도, 개인정보와 사생활은 보호되어야 한다고 생각한다.	투명성 설명 가능성
5	나는 빠른 속도로 인공지능프로그램 기술을 사용하는 사회가 되어가는 현상으로 인해 우리 사회는 이웃이나 단체보다는 자신을 선호하는 개인화가 빠른 속도로 진행될 것으로 생각한다. 이 때문에 우리 사회는 서로 소통하고 교제하는 일이 줄어들 것으로 생각한다.	사람 중심 서비스
6	나는 앞으로 인공지능 기술로 인해 인간의 일자리의 많은 부분이 인공지능 로봇이나 인공지능프로그램으로 대체될 수 있다고 생각한다.	고용
7*	나는 종교, 예술, 출산, 문화 등의 분야는 기계가 대체할 수 없는 인간 고유의 영역에 해당하는 것이므로, 이것에 대한 의미나 가치를 인공지능기술로부터 허가받거나 침해 당해서는 안 된다고 생각한다.	허용과 한계
8	나는 인간처럼 스스로 느끼고 생각하는 인공지능 로봇을 위해 법을 만드는 것에 찬성한다.	로봇의 권리
9	나는 인공지능개발자가 가장 우선 깊게 생각해야 할 원칙은 경제적 이익이나 기술적 발달보다 윤리적인 내용이라고 생각한다.	책임성
10	나는 인공지능 질병 검진, 로봇수술, 로봇 판사의 판단을 신뢰하고 따를 수 있다.	안정성, 신뢰성
11	나는 인공지능의 발달은 우리 인간에게 공평한 기회와 나눔을 줄 것이라고 생각한다.	차별금지
12	나는 인공지능 자동차의 사고가 일어난다면, 인공지능 자동차를 만든 회사가 인공지능 자동차의 기술에 대해 특별한 권리를 가지고 있다고 하더라도, 그 사고의 원인이나 진실을 밝히는 데 필요하다면 인공지능 자동차의 기술에 대해서 시민에게 공개해야 한다고 생각한다.	투명성 설명 가능성
13	나는 우리 집에 인공지능 장치 및 기술이 더 많이 도입되면 우리 가정의 삶이 더 편해지고 나아질 수 있는 장점이 많아서, 인공지능 프로그램과 로봇이 많으면 많을수록 좋다고 생각한다.	사람 중심 서비스
14	나는 노동자들에게는 인공지능 로봇에 의한 혜택보다는 손실이 더 클 것이라고 생각한다.	고용
15	나는 인공지능 기술이 발달함에 따라 윤리성, 인간성, 문화성, 예술성에 영향을 끼치는 효과가 클 것이라고 생각한다.	허용과 한계
16	인공지능 로봇에게 직업을 선택할 수 있는 권리를 주어도 된다고 생각한다.	로봇의 권리
17	나는 인공지능 로봇에 의한 가짜뉴스나 가짜 정보를 만들어 널리 퍼트려서 생긴 사고에 대해서, 그 프로그램을 만든 개발자와 그 기업, 그리고 사용자에게도 법적 책임이 있다고 생각한다.	책임성
18	나는 인공지능 로봇이 안전관리자 없이 단독으로 학생들과 함께 있으며, 수업을 진행해도 된다고 생각한다.	안정성, 신뢰성
19*	나는 인공지능이 발전한다면 인간이 인공지능 로봇의 위치보다 더 낮은 위치에서 대우 받는 사람들도 생길 것이라고 생각한다.	차별금지
20	나는 인간처럼 생각하고 판단 가능한 인공지능 로봇이 자신과 유사한 인공지능 로봇이나 더 발전시킨 인공지능 로봇을 만들고 개발할 때에도, 인간에게 인공지능 기술에 대해 충분한 설명을 할 수 있는 투명함을 가져야 한다고 생각한다.	투명성 설명 가능성
21	나는 현재 선생님께 배우는 학교 수업 방식보다 최신의 기술이 적용된 인공지능 로봇 학습 환경을 갖춘 학교 수업 방식이 학생에게 훨씬 더 학습효과를 높이고 효율적이라고 생각한다.	사람 중심 서비스
22	나는 인공지능이 인간의 직업에 큰 영향을 끼칠 것이라고 생각한다.	고용
23	나는 인공지능 로봇에 의한 그림, 소설 등의 제작물이나 창작물에 대해 예술의 한 형태로 인정할 수 있다고 생각한다.	허용과 한계
24	나는 인간처럼 느끼고 생각하고 판단이 가능한 로봇도 일에 있어 파업도 할 수 있으며, 이것을 인정하고 허용해야 한다고 생각한다.	로봇의 권리

* 역 채점 문항