

초등학교 도덕과 수업에서 인공지능 윤리교육 방안 탐색 연구*

신현화¹, 신태수², 신현우³

<< 요약 >>

인공지능이 우리 생활의 일상이 되어 가고 있는 현실에서 그것의 사용으로 인해 우리는 많은 편리함과 유익을 얻고 있다. 그러나 인공지능의 편리함과 유익 못지않게 여러 가지 다양한 문제에 직면하여 그것을 해결해야 하는 과제 또한 안고 있다. 이러한 현실은 우리 아이들에게도 예외는 아니다. 이에 본 논문에서는 인공지능 시대를 살아갈 아이들에게 필요한 가치 규범 즉, 인공지능 윤리에 대한 이론적 탐색을 시도하였다. 인공지능 윤리의 실태, 인공지능 윤리의 순기능과 역기능, 그리고 인공지능 윤리의 딜레마 사례 탐구 등이 그것이다. 이를 바탕으로 초등학교 도덕과 교육에서 인공지능 윤리교육이 나아가야 할 방향 제시와 더불어 인공지능윤리 교육의 효과적인 지도방안을 모색하였다. 그러므로 본 논문은 인공지능 윤리에 대한 이론적 탐색과 인공지능 윤리교육의 목표 및 내용 제시 그리고 인공지능 윤리교육의 지도방안 모색 등으로 구성되어 있다.

주제어 : 인공지능, 인공지능 윤리, 디지털 리터러시, 협동학습, 메타버스, 합리적 의사소통, 인간성

1. 원주 서곡초등학교 교사, sh2@knue.ac.kr (주저자)
2. 한남대학교 부교수, penpipe@hnu.kr (공동저자)
3. 청주교육대학교 교수, shinnew@cje.ac.kr (교신저자)

I. 서론

인공지능은 대개 인간의 지적 능력 즉, 학습 능력과 추론 능력, 지각 능력, 자연언어의 이해 능력 등을 컴퓨터 프로그램으로 실현한 기술을 일컫는다¹⁾. 다시 말해 인공지능은 인간과 유사하게 생각하고 행동할 수 있는 지능을 가진 컴퓨터라고 할 수 있다. 그리고 인공지능 기술은 자율 주행 자동차, 가상인간, 인공지능 챗봇, 자동 번역 등의 예에서 알 수 있듯이 이미 우리 일상생활 깊숙이 자리 잡고 있다. 인공지능은 4차 산업혁명을 이끄는 핵심 기술 중의 하나이며 4차 산업혁명 시대를 살아가는 우리에게 가장 필요한 기술이다.

인간의 사고와 행동에 대한 모방을 가능하게 한 인공지능의 핵심인 알고리즘은 어떤 문제를 해결하는데 필요한 계산, 절차, 방법 등과 같은 명령어들의 집합이다. 그러한 알고리즘 설계에서 중요한 것은 데이터인데 어떤 데이터를 기반으로 하는지에 따라 결과가 달라지기 때문이다. 어떤 데이터를 기반으로 하는지에 따라 알고리즘의 결과가 누구에게는 정확하고 공정한 선택을 제공하지만 다른 사람들에게는 부정확하고 불리한 선택을 제공하게 된다. 알고리즘이 모든 사람에게 최적의 것만을 추천하고 제공하는 순기능의 역할만 하지 않는다는 사실이 컴퓨터 기술에 기반한 인공지능이 지닌 문제점이라고 할 수 있다. 이러한 이유로 인공지능의 순기능과 역기능 그리고 인공지능의 올바른 활용 등에 대한 교육이 이루어져야 한다. 이것이 인공지능에 대한 교육 특히 인공지능 윤리교육이 필요한 이유이다.

인공지능기술의 발전은 그동안 인간의 고유영역으로 여겨져 왔던 지적 영역 즉, 생각하고 학습한 것을 바탕으로 주어진 과제를 수행하는 영역에서 기계가 사람을 대신할 수 있음을 여실히 보여주고 있는데 알파고와 이세돌 9단의 바둑 대결이 그 단적인 예이다. 심지어 인공지능이 엄청난 양의 빅데이터의 학습과 그에 기반한 알고리즘에 따라 인간과 매우 유사한 감정을 표현하기도 한다. 그리고 인공지능과 IT 기술의 발전은 실제보다 더 실제 같은 가상의 세계로 우리를 이끌고 있으며 어떤 면에서 가상의 세계가 현실의 세계보다 더 많은 영향을 주고 있다.

현대사회는 거대한 양의 정보가 오가는 빅데이터의 시대이며 가상의 세계가 현실의 삶을 흡수하고 있다. 주된 삶의 영역과 공간의 전환은 현대인의 삶과 가치관의 변화에 적지 않은 영향을 끼치고 있다. 컴퓨터, 스마트폰과 같은 IT 기술을 활용한 기기들이 그랬듯이 머지않은 장래에 인공지능 기술도 우리의 일상에서 중요한 위치를 차지하게 될 것이다. 그것으로 인해 우리의 삶이 윤택해지기도 하겠지만 지금까지와는 전혀 다른 성격의 문제나 쟁점에 직면하게 될 것이다. 그러므로 우리에게 그러한 문제나 쟁점들의 올바른 해결을 위해 가치·덕목에 근거한 윤리

1) 두산백과.

적 판단이 필요하며 그때 인공지능 윤리가 중요한 역할을 담당한다. 이에 본 논문에서는 인공지능 윤리에 대한 이론적 탐색을 기반으로 초등학교 도덕과 교육을 통해 인공지능 윤리를 지도할 수 있는 효율적인 방안을 모색해 보았다.

II. 인공지능 윤리에 대한 이해

1. 인공지능 윤리의 동향

인공지능 윤리는 인공지능을 개발·운영·사용함에 있어 개발자와 소비자에게 요구되는 윤리의식을 말한다. 보다 구체적으로 인공지능 윤리는 사회에서 발생하는 다양한 쟁점에 대한 단순한 선악 판단의 기준이라기보다는 우리 사회의 핵심적 가치와 그러한 가치를 균형 있게 존중하는 방식으로 인공지능을 개발하고 활용함에 있어서 주의해야 할 점 및 제도적 장치 등에 대한 통합적인 탐색 노력을 말한다(이상욱, 2021). 최신 세계 각국에서 제시하고 있는 인공지능 윤리는 인공지능 연구 및 제작에 관한 실질적인 지침의 형태로 나타나고 있다. 유엔, 경제개발기구, 국제전기통신연합(ITU), 유네스코, 구글, 마이크로소프트 등 국제기구와 기업들도 인공지능 원칙을 발표해 왔고 현재 150개 이상의 인공지능 규범이 존재하는 것으로 알려져 있다.

먼저 인공지능에 대한 윤리적 가이드라인으로 유네스코에서 제시한 AI 윤리에 대한 권고안이 있다. 이 권고안에는 6개의 가치와 12개의 원칙을 담고 있다. 존엄성, 인권 및 자유, 소외로부터의 자유, 조화로운 삶, 신뢰 가능성, 환경보호 등의 6가치와 인간과 인간의 번영 원칙, 비례성의 원칙, 인간의 관리 감독 및 결정의 원칙, 지속 가능성 원칙, 다양성 및 포용성의 원칙, 개인정보 보호 원칙, 인식 및 교육원칙, 다중이해관계자 및 적응형 거버넌스 원칙, 공정성 원칙, 투명성 및 설명 가능성 원칙, 안전 및 보완원칙, 책임 및 책무성 원칙 등의 12원칙이다(변순용, 2020a: 154 재인용).

이에 더해 유네스코에서는 2021년 11월 23일 프랑스 파리 유네스코 본부에서 열린 제41차 유네스코 총회에서 141개 항목으로 이뤄진 '인공지능 윤리 권고(Recommendation on the ethics of artificial intelligence)'를 193개 유네스코 회원국이 만장일치로 채택하였다. 이 권고안은 2020년 3월 유네스코가 인공지능 윤리 전문가 24명을 중심으로 AHEG(Ad Hoc Expert Group)를 구성한 이래 1년 9개월 만에 채택된 것으로 인공지능 윤리에 대한 최초의 세계적 표준 지침으로 데이터 보호 강화, 사회적 점수 평가 및 대량 감시 금지, 모니터링 및 평가 지원,

환경보호 등의 내용을 담고 있다²⁾.

다음으로 일본의 인공지능 활용 윤리지침이 있다. AI는 인간의 기본적 인권을 침해하지 않을 것, AI를 누구나 이용할 수 있도록 교육을 충실히 할 것(접근의 용이성), 개인정보를 신중히 관리할 것, AI의 보완조치를 확보할 것, 공정한 경쟁환경을 유지할 것, AI를 활용한 기업에 결정 과정에 대한 설명 책임을 부여할 것(설명 가능성 및 책임성), 국경을 초월한 데이터 이용환경을 정비할 것 등이다(한국정보화진흥원, 2019).

한편 유럽(EU)은 지난 2019년 미국 등 주요 선진국 중 가장 먼저 ‘신뢰할 수 있는 AI (Trustworthy AI)’에 대한 윤리 가이드라인을 만들고 법 제정을 추진하고 있다. 유럽의 신뢰 가능한 AI를 위한 4가지 윤리 원칙을 제시하였다. 인간의 자율성 존중의 원칙이다. 인간과 AI의 상호작용에서 우리는 자기 결정이 가능해야 하며 민주적 과정에 참여할 수 있어야 한다는 것이다. 둘째, 해악 금지의 원칙이다. AI 시스템이 해를 발생시키거나 악화시켜서는 안 되며, 인간에게 불리한 영향을 끼쳐서는 안 된다는 것이다. 셋째, 공정성의 원칙이다. AI 시스템의 개발, 배포 및 사용은 공정해야 한다는 것이다. 넷째, 설명 가능성의 원칙이다. AI 시스템의 과정이 투명해야 하고, AI 시스템의 능력과 목적이 공개적으로 밝혀져야 하며 최대한 가능한 정도로 결정들이 직·간접적으로 영향을 받는 사람들에게 설명될 수 있어야 한다는 것이다(EU, 2019; 변순용b, 2020: 434-436 재인용).

중국은 2021년 9월 전 사회의 인공지능 윤리의식과 행위 자각을 증진하고, 책임을 지는 인공지능 연구개발 및 응용 활동을 적극적으로 유도하며, 인공지능의 건강한 발전을 촉진하고자 『차세대 인공지능 윤리 규범』을 발표하였다. 인간복지 증진, 공평·공정 촉진, 프라이버시 안전 보호, 제어·신뢰성 확보, 책임담당 강화, 윤리적 소양 향상 등의 6가지 원칙을 강조하고 있다. 그리고 거기에는 인공지능 관리·연구개발·공급·사용 등과 관련된 활동에 종사하는 자연인, 법인 및 기타 관련 기관 등에 따라 아래의 <표 1>과 같은 윤리 규범을 담고 있다.

<표 1> 인공지능 윤리규범 구성³⁾

관리규범	R&D규범	내용공급규범	사용규범
민첩한 거버넌스 추진 적극적인 실천 시범 정확한 권리 행사·사용 리스크 방지 강화 포용·개방 촉진	자율의식 증진 데이터 품질 향상 안전·투명성 증강 편견·차별 회피	시장법칙 존중 품질 통제 강화 사용자 권익 보장 응급보장 강화	선의적 사용 제창 오남용 회피 규경위반 악용 금지 적시성·능동성 피드백 사용능력 제고

2) <http://www.aitimes.kr/news/articleView.html?idxno=23359>(인공지능신문)

3) http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html

미국의 경우 2020년 신뢰, 공정, 안전 등을 강조하는 인공지능 활용 원칙을 발표했고 바이든 행정부는 인공지능 기술의 해악에 대비하여 개인의 자유와 권리를 지킬 수 있는 권리장전의 필요성을 역설하며 강력한 인공지능 규제법을 준비하고 있다. 미국은 인공지능 산업 진흥과 공익 및 개인의 권리 보호 사이에서 양자 간 어떻게 균형을 잡아갈 것인가에 대해 고민하고 있다.

우리나라는 공공성(Publicness), 책무성(Accountability), 통제성(Controllability), 투명성(Transparency) 등의 4대 공통 원칙을 담은 지능정보사회의 윤리 가이드라인이 제시된바 있다. 지능정보기술은 인류의 보편적 복지에 기여하고 사회변화를 일으키지만, 자기학습 및 진화하는 기술이며 설명이 필요한 알고리즘을 가지고 있는 기술이기 때문에 그러한 원칙이 지켜져야 한다는 것이다(한국정보화진흥원, 2018: 9). 또한, 2020년 6월 기존의 국가정보화 기본법을 전면 개정하여 지능정보화 기본법(2020.12.10)을 마련하였고 동 법률에서 인공지능 윤리를 마련하도록 규정하고 있다. 그리고 2020년 12월 과학기술정보통신부와 정보통신정책연구원은 인공지능 시대 바람직한 인공지능 개발·활용 방향을 제시하기 위한 국가 인공지능 윤리기준을 발표하였다. 국가 인공지능 윤리기준(안)에는 ‘사람 중심의 인공지능’을 위한 최고 가치인 ‘인간성(Humanity)’을 위한 3대 기본원칙(인간의 존엄성 원칙, 사회의 공공선 원칙, 기술의 합목적성 원칙)과 10대 핵심요건(인권 보장, 프라이버시 보호, 다양성 존중, 침해금지, 공공성, 연대성, 데이터 관리, 책임성, 안전성, 투명성)이 제시되어 있다(과기정통부, 2020).

이상에서 세계 각국은 인공지능 기술의 급격한 발달에 따른 부작용에 직면하여 그것과 인간이 공존할 수 있는 규범적 토대인 인공지능 윤리를 정립하기 위해 지속적인 노력을 기울이고 있음을 알 수 있다. 세계 각국의 인공지능 윤리는 인간의 자율성, 존엄성, 인권 등에 강조점을 두고 있다는 점에서 인공지능 자체의 윤리이기보다는 인공지능을 올바르게 활용하도록 하는 인간의 윤리에 가깝다. 그리고 인공지능 윤리는 인공지능 기술의 발전에 따라 지속적으로 수정 보완되어야 하며 특히 인공지능 윤리는 인공지능 기술이 이야기시키는 부작용에 대한 대비와 더불어 인공지능의 도덕·윤리적 장치가 구현되는 방향으로 정립되어야 한다.

2. 인공지능의 순기능과 역기능

과학과 기술의 발전은 우리 삶의 많은 변화를 가져다주었다. 인공지능이 인간의 지식과 기술을 급속히 대체하고 있다. 전통적인 시장은 웹이라는 가상의 공간으로 대체되고 있으며, 다른 사람과의 만남의 문화는 SNS를 매개로 시공을 초월해 이루어지고 있다. 인공지능의 활용으로 인해 우리는 삶의 다양한 분야에서 전보다 윤택한 삶을 영위하게 되었다. 지능형 로봇, 자율주행 자동차

차와 더불어 가상인간(인플루언서), 인공지능 챗봇 등을 통해 그러한 삶이 전보다 수월해졌다.

그러나 인공지능의 긍정적 활용 이면에는 인공지능의 그늘이 있다. 인공지능의 역기능은 윤리 의식이 부족한 상태에서 인공지능을 활용하는 경우 잘 나타난다. 중국의 신장 위구르족 추적 감시를 위해 안면인식 기술을 활용한 최첨단 감시 카메라를 이용한 주민 감시, 영상 조작 기술 딥페이크를 이용한 인도의 선거, 인공지능이 작성한 가짜 기사 등이 그 예이다.

한편 문제 처리 과정에서 불공정함을 경험한 사람은 그러한 문제에 직면하여 인간보다 인공지능에 더 의존하게 되는 경향이 있다. 의료와 재판 그리고 시험에서 불공정을 경험한 사람은 인공지능 의사, 판사, 면접관의 공정한 판단을 기대하며 인공지능에 의존한다. 그러나 여기에도 역시 인공지능의 그늘이 존재한다. 만일 인공지능이 차별과 편견에 기반을 둔 데이터에 의해 학습된 알고리즘에 따라 판단을 내린다면 거기에 더욱더 심각한 불공정의 문제가 발생한다.

이때 우리는 한 가지 역설적 상황에 직면하기도 한다. 인공지능으로 생겨난 문제 즉, 잘못된 알고리즘 기반의 인공지능이 내린 불공정한 판단으로 생겨난 문제의 해결책을 인간이 아닌 인공지능에서 찾게 되는 우를 범하고 있지는 않은지 생각해 볼 필요가 있다. 인공지능의 편견과 차별성은 데이터 수집 단계에서부터 편향적이고 왜곡된 자료인지에 대한 검토와 더불어 최초 인공지능 설계자들에 대한 올바른 가치관 및 태도 교육이 문제 해결을 위한 일차적 과제가 아닐까?

이러한 맥락에서 우리에게 인공지능의 올바른 사용 및 활용에 대한 교육, 다시 말해 인공지능 윤리교육이 필요하다. 우리가 빅데이터와 알고리즘 그리고 인공지능의 결합으로 생겨나는 문제들을 해결하려면 기존의 접근법과 확연히 다른 새로운 접근법이 활용되어야 한다. 4차 산업혁명으로 대변되는 지능 정보화 사회에서 우리는 현실 세계보다는 디지털 사이버 공간에서의 삶이 증가하며 그 중요성이 증대하는 상황에서 문제 해결을 위해 디지털 문법에 대한 이해와 적용 교육이 필요한 시점이다.

3. 인공지능의 윤리적 딜레마 : 자율주행 자동차 사례를 중심으로

운전자 또는 승객의 조작 없이 스스로 움직이는 자동차 즉, 자율주행 자동차가 머지않은 장래에 일상생활의 중요한 운송 수단으로서 위치를 다지고 있는 상황에서 우리는 자율주행 자동차의 편리함에만 머물러 있을 수 없다. 자율주행 자동차의 사용으로 인해 생겨날 문제 또한 적지 않기 때문이다. 그리고 이러한 문제는 윤리적 딜레마로 연결된다.

영국의 철학자 필리파 풋(Philippa Foot)이 제기한 트롤리 문제와 유사한 ‘터널 문제’ 사례를 통해 우리는 자율주행 자동차의 윤리적 딜레마를 잘 이해할 수 있다(구본권, 2020: 51). 자율주

행 상태로 운행 중인 차가 좁은 1차선 터널에 진입하려고 하는 순간 근처에 있던 어린아이가 발을 헛디뎠서 도로 위로 넘어진다. 이때 자율주행 자동차가 어린아이를 피할 시간이 없는 상황이다. 어린아이를 치고 터널로 진입하든지 아니면 터널 입구 암벽에 차를 부딪쳐 멈춰서야 한다. 이 경우 어린아이를 구할 수는 있으나 자율주행 자동차의 운전자는 크게 다치거나 죽게 된다. 이와 같은 절박한 상황에서 자율주행 자동차의 설계자가 미리 정해 놓은 우선순위에 따라 알고리즘은 어떤 생명을 우선 보호할 것인지에 대해 판단을 내리게 된다.

터널 문제와 유사한 실제 상황에서 자동차를 운전하는 사람은 대개 도덕 혹은 윤리 규범이 아니라 본능에 따라 행동을 하게 되는데, 이 경우 보통 운전자의 책임 문제가 큰 쟁점이 되지 않는다. 그러나 컴퓨터 알고리즘에 기반을 둔 인공지능 자율주행 자동차의 경우에는 책임의 문제가 중요한 쟁점이 된다. 인공지능 자율주행 자동차는 무수히 많은 데이터에 기반을 둔 알고리즘에 의해 사전에 설계된 대로 움직이는 기계이기 때문이다. 그 기계는 모든 상황 속에서 최선의 판단을 내려야 하며 적어도 인간의 선택보다 나은 선택을 할 것이라는 기대 속에서 작성된 알고리즘을 전제로 하기 때문이다. 만일 인공지능이 주행하는 자동차가 사고를 일으킨다면 그 사고의 책임을 운전자에게 부과해야 하는지 아니면 자율주행 자동차를 설계하고 만든 제작자에게 부과해야 하는지가 불명확하다. 정확히 말하면 사고의 책임 소재에 대한 관련 당사자들 간의 합의나 사회적 기준이 아직 미비한 상태이다. 나아가 사고의 책임이 설계자(혹은 제작자) 그리고 사용자에게 있다고 가정할 경우 일차적 책임을 누구에게 귀속시켜야 하는지가 불분명하다. 자율주행 자동차의 사례에서 보듯이 인공지능은 설계자의 가치관 및 윤리관에 의해 우선순위가 결정되는 알고리즘의 선택에 따라 행위를 수행하기 때문에 인공지능 설계자의 윤리적 자세 및 태도가 매우 중요하며 바로 이 지점에서 인공지능의 윤리적 딜레마가 생겨난다.

인공지능은 순기능과 역기능 그리고 윤리적 딜레마를 동반한다는 점에서 인공지능 윤리의식의 정립이 필요하며 또한 인공지능 시대를 살아갈 학생들에게 그것을 올바르게 활용할 수 있는 인공지능윤리 교육이 필요하다고 본다. 이에 본 연구에서는 가급적 이른 시기인 초등학교 교육, 특히 도덕과 교육을 통해 인공지능 윤리의식 함양을 위한 실제적 방안을 찾아보았다.

III. 초등 도덕과 교육에서 인공지능 윤리교육의 적용

도덕과 교육에서 인공지능 윤리교육을 한다고 했을 때 무엇을 가장 먼저 고려할 것인지에 대해 숙고해 보아야 할 것이다. 기계나 컴퓨터 그리고 인공지능이 아닌 사람이 주인이 되는 삶을

위한 교육, 다시 말해 인공지능을 잘 활용하여 학생이 주체적으로 자신의 삶을 영위해 나갈 수 있는 능력을 길러주는 교육이어야 하는데 이를 위해서는 다음의 세 가지 측면이 고려되어야 한다. 첫째, 어떤 목적을 갖고 살아갈 것인가. 둘째, 그러한 목적 달성을 위해 어떤 것을 배워야 할 것인가. 셋째, 학습한 것 혹은 학습을 통해 습득한 것(재능)을 삶에 어떻게 적용 또는 활용할 것인가.

이러한 세 가지 관점을 인공지능 윤리교육에 적용해 보면 첫째, 목적과 관련하여 나는 어떤 목적으로 인공지능을 만들 것인가에 대한 깊은 성찰과 고민이 있어야 한다. 인공지능 개발과 제작, 사용의 궁극적 목적 설정은 인간의 행복을 위한 것이라는 전제하에 이루어져야 함은 물론이다. 둘째, 그러한 목적 달성을 위해 우리는 어떤 데이터를 학습시킬 것인가에 대해 고민해 보아야 한다. 인공지능 윤리교육에서 데이터의 특성과 종류에 따라 학습시킬 데이터를 결정하는 것이 중요하기 때문이다. 셋째, 학습된 데이터에 따라 만들어진 인공지능을 어떤 또 다른 영역에 적용할 것인가의 문제이다. 우리는 개인뿐만 아니라 공동체의 선을 위한 영역에 인공지능이 활용될 수 있도록 해야 한다.

이러한 맥락에서 인공지능 윤리교육이 지향하는 바는 무엇이며 과연 어떤 내용을 선정하여 가르칠 것인가? 특히 학교 도덕과 교육을 통해 가르쳐야 할 인공지능 윤리교육의 바람직한 방향이나 목표는 어떠해야 할까? 그리고 그러한 목표 달성을 위해 어떤 내용을 선정해야 하며 어떻게 하면 그것을 효과적으로 가르칠 수 있을 것인가에 대해 구체적으로 살펴보겠다.

1. 초등 도덕과 교육에서 인공지능 윤리교육의 목표 및 내용

가. 도덕과 교육에서 인공지능 윤리교육의 지향점

도덕과 교육에서 인공지능 윤리교육을 실행한다면 우리는 어떤 방향으로 나아가야 할까? 이에 대한 대답은 교육의 대상인 학생들이 살아갈 미래 사회의 특징 중의 하나인 디지털 공간에서의 삶이 현실 공간 못지않게 중요해진다는 점에서 찾을 수 있을 것이다. IT 기술과 인터넷의 발달 그리고 거대한 양의 정보와 더불어 인공지능의 보편화는 우리가 일찍이 경험해보지 못했던 엄청난 변화의 소용돌이 속으로 우리의 삶을 옮겨놓고 있다. 급격한 변화의 중심에는 이른바 컴퓨터라는 기계가 자리 잡고 있으며 그 기계가 인간의 일을 상당 부분 대신하고 있다. 여기서 우리는 인간에게 쉬운 일이 컴퓨터에게 어렵고, 컴퓨터에게 쉬운 일이 인간에게 어렵다는 모라벡의 역설(Moravec's Paradox)에 주목할 필요가 있다. 이것은 컴퓨터가 모든 분야에서 인간을 이길 수는 없음을 의미한다. 인간이 컴퓨터를 이기는 영역이 존재한다는 것인데, 이를테면 감정

을 경험하고 관계를 형성하는 일, 복잡한 규모의 출처를 활용해 질문하고 답하는 일, 컴퓨터나 인공지능이 해야 할 작업이나 그것에 제공할 데이터를 결정하는 일, 추상적 가치와 관련된 문제에 대해 의사를 결정하는 일 등이다. 이는 인공지능 윤리교육이 나가야 할 바 즉, 목표 설정에 중요한 지표가 된다. 그러므로 인공지능 윤리교육의 목표는 인공지능과 알고리즘 활용 시 나타나는 문제점을 예방하기 위한 윤리적 가이드라인의 성격을 지니는 동시에 그러한 문제의 궁극적 해결책은 기계가 아닌 인간이 스스로 마련해야 하는 것으로 결국에는 인간의 올바른 가치 판단의 문제로 귀결된다는 사실에 기반하여 설정되어야 한다.

그렇다면 도덕과 교육에서 인공지능 윤리교육의 목표는 구체적으로 어떻게 설정할 수 있을까? 한편으로 면대면 공간을 포함한 디지털 공간에서 개개인의 의사를 합리적으로 표현하고 소통할 수 있는 능력, 즉 다른 사람과 함께 바람직하고 긍정적으로 소통·협업할 수 있는 사람으로 기르는 것이며 다른 한편으로 정보의 무분별한 수용이 아닌 비판적 수용 능력, 즉 상당히 많은 양의 정보에 접근하여 그것을 비판적으로 수용할 수 있는 능력을 기르는 것으로 설정해 볼 수 있다. 달리 말해 학생들이 디지털 공간에서 민주적으로 의사를 표현하며 다른 사람과의 소통 및 협업을 통해 함께 상생할 수 있는 능력을 지닌 인간을 육성하는 것이 도덕과 교육에서의 인공지능 윤리교육의 목표라고 할 수 있다.

나. 인공지능 윤리교육에서 가르쳐야 할 것

인공지능 윤리교육에서 가르쳐야 할 내용을 선정할 때 참고할 만한 다양한 연구들이 있다. 대구교육정보원은 한 연구에서 인공지능 시대 학생들에게 필요한 능력으로 공감 능력(61.9%), 도덕성(45.9%), 의사소통 능력(32.1%), 창의성(28.9%) 등을 제시한 바 있다. 공감은 타인의 감정을 공유하고, 그 감정에 관해 생각하고, 그 감정을 배려하는 것으로, 어떤 하나로 공감을 표현하기란 쉽지 않다. 즉, 공감은 공유, 생각, 배려를 기반으로 서로에게 다양한 방식으로 반응하는 것이다(정지은 역, 2021: 373-374). 이러한 공감은 소통의 중요한 전제 조건으로 길러질 수 있으며 인공지능 시대를 살아갈 학생들에게 가장 필요한 능력이라고 할 수 있다. 그리고 ‘21세기에 학생들은 무엇을 배워야 하는가?’라는 질문에 답하기 위해 찰스 파텔이 설립한 미국의 CCR(교육과정 재설계센터, Center for Curriculum Redesign)에서 21세기 핵심역량(21st Century Skills)으로 이른바 4C, 즉 창의력(Creativity), 의사소통(Communication), 비판적 사고(Critical thinking), 협업(Collaboration)을 제시하였다⁴⁾.

4) <https://riverxchange.com/2015/05/03/riverxchange-video-projects-showcase-21st-century-skills>

한편 OECD에서 발간한 교육과 기술의 미래 2030 보고서에서 21세기에 필요한 기술들을 제시하였는데 기초 능력, 역량, 인격적 특성의 3가지 영역이다. 그리고 앞 장에서 밝힌 바와 같이 과학기술정보통신부와 정보통신정책연구원이 2020년 12월 「국가 인공지능 윤리기준」에 발표한 인공지능 윤리기준 3대 기본원칙과 10대 핵심요건이다. 여기서 특히 주목할 만한 것이 있는데 ‘인간성’(Humanity)이다. 인공지능 기술이 급성장하는 현 시대적 배경 속에 기술 개발부터 활용에 이르는 전 단계에서 모든 사회 구성원이 고려해야 할 가장 높은 가치로 인간성을 제시하였다.

마지막으로 AI 윤리교육의 내용(주제)을 구체적으로 제시한 최근의 연구도 있다(변순용, 2020a, 160). 인공지능이란 무엇인가, 인공지능과 인간의 관계, 인공지능과 기계학습, 인간의 감정과 인공지능의 감정, 자율주행 자동차의 윤리적 문제, 데이터 윤리: 인공지능과 빅데이터, 인공지능과 인간의 복지, 인공지능 로봇과 가족, 인공지능과 교육의 변화 등이 그것이다.

이상의 내용을 바탕으로 도덕 교과에서 인공지능윤리 교육을 통해 지도해야 할 내용을 체계적으로 제시하려고 했지만, 미래형 교육과정과 같이 이와 관련된 논의가 활발히 진행되고 있는 상황에서 그 내용이나 주제를 확정적으로 제시하는 일이 간단한 문제는 아니라는 생각이 든다. 추후 연구를 통해 지속적인 보완이 이루어졌으면 한다.

2. 초등 도덕과 교육에서 인공지능 윤리교육의 실제

가. 디지털 리터러시 추구를 위한 인공지능 윤리교육

엄청난 양의 지식이 빠른 속도로 생산되는 환경에서 효과적이고 유일한 교육과 교수학습 방법은 이러한 지식의 구조를 알고 여기에 대응하는 것이다. 지식의 구조가 바뀐 디지털 세상에서 유용한 것은 스스로 학습자가 되어 끊임없이 변화하고 확대되며 새로워지는 지식을 탐구해나가는 길이다. 특히 인터넷과 정보의 홍수로 대변되는 디지털 세상에서 교육을 통해 학생들에게 길러주어야 할 역량은 디지털 리터러시이다. 디지털 리터러시는 올바른 정보에 접근하고 비판적으로 수용할 수 있을 뿐만 아니라 바람직하고 긍정적으로 소통 및 협업할 수 있는 역량을 말하는 것으로 정보의 무분별한 수용이 아닌 비판적 수용 능력, 즉 디지털 공간에서의 책임감과 능동적이고 합리적인 소통 능력을 중시한다.

다시 말해 디지털 리터러시는 디지털의 속성과 구조를 파악하고 디지털 문법을 제대로 이해하고 사용하는 능력을 말하는 것으로 인공지능 윤리교육에서 무엇보다 중요한 위치를 차지하고

있다. 미래 세대들은 현실 문법이 아닌 디지털 문법 속에서 살아갈 것이며, 그들이 겪게 될 문제들은 그 해결을 위해 디지털 문법에 대한 정확한 이해와 올바른 사용 능력을 요청하기 때문이다. 특히 디지털 리터러시와 관련해서 저소득층, 장애인, 농어민, 고령층 등 이른바 디지털 소외계층의 불평등 문제가 종종 발생하는데 이러한 불평등 문제 해결이 교육이 담당해야 할 중요한 역할 중 하나이다.

그렇다면 인공지능 윤리교육에서 어떻게 하면 디지털 리터러시 역량을 효과적으로 함양할 수 있을까? 디지털 리터러시에는 2가지 중요한 요소가 있다. 하나는 정보 및 데이터에 대한 비판적 수용 능력이고 다른 하나는 디지털 공간에서 자신의 의사를 합리적이며 책임감 있게 표현하고 소통할 수 있는 능력이다. 이 두 가지 측면에 역점을 둔 인공지능 윤리교육 방법을 제시하면 다음과 같다.

첫째, 도덕과 교육에서 인공지능 윤리를 가르칠 때는 정보 및 데이터에 대한 비판적 수용 능력이 중요하다. 특히 데이터에 대한 차별 및 편향성에 대한 비판적 고려 없이 데이터를 수용하는 것의 문제점들을 주의 깊게 살펴보아야 한다. 인공지능의 설계자가 어떤 데이터를 선택하여 학습시키는가에 따라 인공지능이 내린 판단이 공정할 수도 그렇지 않을 수도 있다. 그러므로 인공지능이 내린 판단의 공정성은 편견이나 차별성이 거의 없는 데이터의 선정 및 선정된 데이터의 학습 여부에 달려 있다. 따라서 초등학생을 대상으로 하는 인공지능 윤리교육을 가르칠 때 가장 중요한 것 중의 하나는 미래의 인공지능의 설계자인 학생들이 특정한 가치에 대해 편견 및 차별을 지니지 않도록 하는 것이다.

차별 및 편견의 극복을 위해서는 그것의 속성 내지는 기본 전제를 파악할 필요가 있다. 차별은 둘 이상의 대상을 차이를 두어 구별하는 것으로, 그것이 사회과학이나 자연과학의 연구에 활용되는 경우 큰 문제가 되지 않는다. 그러나 차별이 인간 또는 사회적 관계 속에서 발생하는 문제라면 경우가 다르며 매우 심각한 상황에 다다를 수 있다. 차별로 인해 생겨난 문제들이 인간의 존엄성을 침해하는 경우라면 더욱더 그러하다. 특히 어떤 한 인간이 지닌 편견으로 인해 형성된 차별이 인간의 존엄성을 해치는 경우 그것은 매우 심각한 문제가 아닐 수 없다.

이러한 맥락에서 인공지능 설계자가 차별 또는 편견의 가치를 지니지 않도록 하는 것이 중요하며 이를 극복하기 위한 교육적 방안으로 협동 학습이 활용될 수 있다. 협동 학습은 주어진 과제 해결을 위해 구성원들이 서로 협력·협동하는 속성을 지닌 교수학습 방법이다. 협동 학습을 활용하여 인공지능 윤리를 가르칠 때 고려해야 할 것이 있다. 협력이나 협동이 가능한 주제(혹은 활동)를 선정해야 하며, 개별학습보다는 협동 학습이 과제 수행 및 목표 달성에 효과적이어야 한다.

구체적으로 앞에 제시한 자율주행 자동차의 ‘터널 문제’의 예를 활용하여 도덕 교과에서 인공지능윤리를 가르칠 수 있다. 학습 주제는 자율주행 자동차의 윤리지침을 설계자, 제작자, 사용자 측면에서 만들어 보는 것으로 아래의 과정 및 절차를 통해 수업을 진행할 수 있다.

- ① 터널 문제의 사례를 제시한다.
- ② 다양한 관점에서 자율주행 자동차의 판단에 활용할 데이터를 수집한다(설계자, 제작자, 사용자 측면).
- ③ 각자 수집한 데이터에 대한 사실적 타당성에 대한 검토 활동을 한다.
- ④ 공정성의 관점에서 각자 수집한 데이터에 기반을 둔 알고리즘이 생성한 결과를 검토한다 (역할 교환 검사, 포함관계 검사, 보편적 검사 결과, 새 사례 검사 등).
- ⑤ 인공지능이 판단한 결과들에 대해 조원들에게 확인해 보는 활동을 한다.
- ⑥ 데이터의 편향성 및 편견이 발견된 것이 있다면 어떻게 수정 보완할 것인지 토론한다.
- ⑦ 최종적으로 산출된 자율주행 자동차의 윤리지침을 조별로 발표한다.

위의 내용을 도덕과 수업의 한 단원으로 가정하는 경우 4차시 수업으로 구성할 수 있으며 교수학습 과정 안을 간략히 제시해 보면 아래의 <표 2>와 같다.

<표 2> 교수학습 과정안

주제	자율주행 자동차와 관련된 윤리지침 만들기를 통해 정보 및 데이터에 대한 비판적 수용		
관련 성취기준	6도03-01 인권의 의미와 인권을 존중하는 삶의 중요성을 이해하고, 인권 존중의 방법을 익힌다.		
핵심 요소	정보 및 데이터에 대한 비판적 수용	인공지능 윤리영역	인간 존엄성의 원칙 기술의 합목적성의 원칙
학습 목표	문제 상황(터널 문제)에서 발생할 수 있는 설계자의 윤리지침 만들기		
단원 내용 체계	1차시: 자율주행 자동차와 관련된 문제 사태 찾아보기 2차시: 문제 사태에서 나는 어떻게 할 것인가에 대한 의견 쓰기 3차시: 내 판단에 사용할 데이터를 수집하고 판단하기 4차시: 조별 토론을 통해 ‘설계자’의 윤리지침 제안하기		
수업 과정안			
차시	4/4	수업 형태	협동 학습
차시 목표	조별 토론을 통해 설계자의 윤리지침 제안하기		

<p>수업의 흐름</p>	<ul style="list-style-type: none"> □ 도입 (전체 활동) <ul style="list-style-type: none"> - 자율주행 자동차와 관련된 동영상 시청하기 - 지난 시간까지 수집한 데이터와 내 의견 검토하기 □ 전개(조별 활동) <ul style="list-style-type: none"> - 설계자의 관점에서 중요하게 생각한 요소와 그에 따른 근거 데이터 발표하기 - 발표 속에서 나타난 각각의 요소들에 대해 조원들과 협의하기 - 합의된 요소들 속에 나타난 문제점은 없는지 검토하기 - 설계자의 관점으로만 보았을 때 나타날 수 있는 문제(효율성)인간 존중)를 파악하고 자율주행 자동차 설계자의 윤리지침 제안하기 □ 정리(전체 활동) <ul style="list-style-type: none"> - 필요한 정보 및 데이터를 수집할 때 경계해야 할 점에 대해 말하기
<p>지도상의 유의점</p>	<ul style="list-style-type: none"> - 설계자 입장에서 생각할 때 인권보다 경제성과 효율성에 입각해 생각하는 것이 자연스러운 현상임을 알고 처음부터 교사가 인권의 측면에서 접근하지 않는 것이 중요함. - 학생들 스스로 데이터를 수집하고 분석하는 과정에서 갈등 상황에 직면한 후 데이터를 어떻게 받아들일 것인지 함께 토론할 수 있도록 함. - 필요한 정보에 대한 무비판적인 수용이 어떤 문제를 이야기시키는지에 대해 고민하고 함께 문제를 해결할 수 있도록 함.

위 수업 안은 설계자의 입장에서 수행하는 것으로 이것을 사용자의 관점으로 바꾸어 연속적으로 수업을 운영할 수 있다. 사용자는 설계자의 입장과는 다른 측면에서 논의가 될 것이며 이 역시 같은 흐름으로 진행하면서 학생들이 갈등 상황에 직면하여 문제 해결을 위한 타당한 근거 제시를 통해 자율주행 자동차의 편리함의 바탕에는 인권이 자리해야 한다는 것을 스스로 인식하도록 한다.

둘째, 디지털 공간에서 합리적이고 책임감 있게 자신의 의사를 표현하고 소통할 수 있는 능력 함양 방안이다. 우리의 의사소통은 크게 두 가지 공간 즉, 현실과 디지털 공간에서 이루어진다. 디지털 공간에서 이루어지는 의사소통은 그 주체가 사람이라는 점에서 그리고 그 사람은 현실의 공간에서 의사소통을 줄곧 해왔다는 점에서 현실 공간에서의 합리적 의사소통 능력을 기르는 교육 방법에 대한 고찰이 선행되어야 할 것이다. 디지털 공간에서의 의사소통은 장소만 다를 뿐 현실 공간에서의 의사소통과 크게 다르지 않기 때문이다. 공론의 장인 광장에서 자유롭게 자신의 의사를 표현하며 상대방의 견해에 비판적으로 토론하는 일련의 활동을 통해 합리적 의사소통 능력을 기르는 것이 교육의 중요한 과제라는 점에는 큰 이견이 없다. 다시 말해 학생들이 자유로운 대화와 비판적 토론 그리고 합리적 의사소통을 통해 다른 사람과 상생할 수 있는 사람으로 성장하도록 돕는 것이 교육의 중요한 목적 중 하나이다.

의사소통과 관련하여 하버마스의 합리적 의사소통 이론에 주목해 볼 필요가 있는데 올바른 의사소통의 조건으로 다음의 몇 가지를 제시하였다. 소외되거나 배제되는 사람 없이 누구나 평등하게 참여할 수 있을 것, 어떤 주장 혹은 관점이든지 자유롭게 개진될 수 있을 것, 상대방의 말을 자신의 관점이 아닌 상대방의 관점에서 해석하고 나아가 공평하게 서로의 관점을 보완할 수 있어야 할 것 등이다. 그리고 그는 서로 무슨 뜻인지 이해할 수 있고(이해 가능성), 그 내용은 참(진실성)이어야 하며, 상대방이 성실히 지킬 것이라는 믿음이 있고(신뢰성), 말하는 사람들의 관계가 평등하고 수평적(평등성)일 때 진정으로 올바른 대화가 가능하다고 하였다(장춘익 외, 2001: 126; 장춘익 역, 2007a :176, 449, 476; 장춘익 역, 2007b: 198-199). 이와 같은 현실 세계에서의 합리적 의사소통의 조건은 디지털 공간에서도 그대로 적용될 수 있다.

초등학생들에게 디지털 공간에서의 합리적 의사소통 능력을 기르기 위해 메타버스를 활용해 볼 수 있다. 메타버스는 3차원 가상세계, 가공, 추상을 의미하는 메타(Meta)와 현실 세계를 의미하는 유니버스(Universe)의 합성어로 현실 세계와 같은 사회·경제·문화 활동이 이뤄지는 3차원 가상세계를 일컫는다. 메타버스는 5G 상용화에 따른 정보통신기술 발달과 코로나19 팬데믹에 따른 비대면 추세의 가속화로 점차 주목받고 있는데 이와 같은 메타버스 플랫폼을 인공지능 윤리교육에 활용해 볼 수 있다. 그 구체적 과정 및 절차를 제시해 보면 아래와 같다.

① 메타버스에 대한 이해 및 체험 활동이다. 아래와 같은 메타버스 여러 플랫폼 중에서 게더타운(Gather.town)을 활용하여 메타버스에 대한 기본적인 이해 및 체험 활동을 한다. 모뎀 활동을 할 수 있는 다양한 공간을 만들어 본다. 특히 조별 활동(토론)에 사용할 광장 형태의 모뎀 활동 공간을 생성하고 거기에 알맞은 이름을 붙여준다.

② 게더타운 공간에서 인공지능 윤리교육이 가능한 학습 주제(혹은 활동)를 제시한다. 일례로 인공지능 로봇의 상용화에 대비하여 인공지능 로봇 윤리 가이드라인을 만들어 보는 것이다. 보다 구체적으로 설계자, 제작자, 사용자 각각의 입장에서 이미 만들어놓은 게더타운 공간인 광장에서 로봇 윤리 가이드라인을 만드는 조별 활동을 한다. 이때 교사는 광장이 조원들의 자유로운 대화 및 비판적 토론이 가능한 공간 즉, 합리적인 의사소통을 통해 바람직한 결과가 도출되는 공간이 되도록 분위기를 조성한다. 이러한 활동을 통해 학생들은 광장이 비록 가상공간이지만 현실 공간 못지않게 합리적 의사소통이 가능한 공간이라는 사실을 인식하게 된다.

③ 조별 활동 후 인공지능 로봇의 설계자, 제작자, 사용자의 관점에서 윤리 가이드라인을 발표하고 그것에 대해 비판적으로 그리고 다양한 관점에서 의사를 표현하는 등의 토론 활동을 통해 자신이 만든 윤리 가이드라인을 수정·보완한다. 이것을 수업안으로 작성한다면 아래의 <표 3>과 같은 흐름으로 운영될 수 있다.

〈표 3〉 교수학습 과정안

주제	디지털 공간에서의 합리적 의사소통 능력 함양		
관련 성취기준	6도02-01 사이버 공간에서 발생하는 여러 문제에 대한 도덕적 민감성을 기르며, 사이버 공간에서 지켜야 할 예절과 법을 알고 습관화한다.		
인공지능 윤리영역	사회 공공선의 원칙	대상 학년	5-6
학습 목표	가상공간(게더타운)에서 지켜야 할 예절 알기		
단원 내용 체계	1차시: ‘인공지능 로봇’에 대해 알아보기 2차시: 인공지능 로봇의 설계자와 사용자의 측면에서 고려되어야 할 사항에 대한 의견 쓰기 3차시: 설계자와 사용자로 역할을 나누어 토론하기. 인공지능 로봇 개발자에게 필요한 윤리의식 제안하기. 4차시: 가상공간에서 지켜야 할 예절 알기		
수업 과정안			
차시	4/4	수업 형태	협동 학습
차시 목표	가상 공간(게더타운)에서 지켜야 할 예절 알기		
수업의 흐름	<input type="checkbox"/> 도입 - 인공지능 로봇 개발자에게 필요한 윤리의식 제안에 대해 발표하기 <input type="checkbox"/> 전개 - 가상공간에서 토론 과정이 좋았던 측면에 대해 이야기하기 - 가상공간에서 토론 과정이 불편했던 측면에 대해 이야기하기 - 가상공간에서 토론할 때 지켜야 할 예절이 무엇인지 생각하기 <input type="checkbox"/> 정리 - 가상공간에서 지켜야 할 예절에 대해 자기 평가하기		
지도상의 유의점	- ‘인공지능 로봇’은 하나의 소재로서 인공지능 로봇의 윤리지침을 평가하는 데 주안점이 있는 것이 아님. - 대면 수업에서 말하기 어려워하는 학생들이 가상공간에서 태도 변화가 있는지 살펴보는 것이 필요함. - 조별 활동에서 활발하게 토론할 수 있도록 교사가 미리 예절에 대해 강조하지 않는 것이 중요함. - 게더타운 외에도 일상적인 가상공간에서의 나와 친구들의 태도와 자세에 대해 자연스럽게 돌아보는 계기가 되도록 함. - 정해진 답이 있는 것이 아니라 타당한 근거를 가지고 토론하면서 나와 다른 입장을 가진 의견에 대해서 합리적인 토론이 가능하도록 함.		

위의 메타버스를 활용한 인공지능 윤리교육은 디지털 공간에의 삶이 현실 공간 못지않게 중요한 미래 세대들에게 요청되는 디지털 리터러시 역량을 기르는 긍정적 요인으로 작용할 것이다.

다만 대화(혹은 논쟁)의 목적이 상대방을 설득하여 자신의 주장이나 의견에 동조하도록 하는 것이 아니라 상대방의 주장이나 의견을 진심으로 받아들이려는 능력과 자세를 기르는 것이 대화 혹은 논쟁의 궁극적 목적이라는 하버마스의 견해에 주의를 기울일 필요가 있다. 이러한 맥락에서 합리적 의사소통 능력이 디지털 공간에서의 삶이 일상이 되어가고 있는 우리 아이들에게 꼭 필요한 디지털 리터러시 역량이라고 할 수 있다.

나. 인간성(humanity) 추구를 위한 인공지능 윤리교육

인공지능 시대는 이전의 인류가 경험해보지 못한 사회이며, 점차 구축되어가고 있는 미래지향적 사회이다. 따라서 인공지능 시대는 우리에게 이전에 없었던 새로운 차원의 문제들을 던진다. 무엇보다 인공지능이 급속히 인간의 기능들을 광범위하게 대체해나가면서 인공지능으로 대체할 수 없는 사람만의 기능은 무엇인지, 더 나아가 궁극적으로 인간의 본질과 인간성이란 무엇인지에 대한 물음에 직면하도록 만든다. 인간에게 편의를 제공함으로써 노동의 굴레로부터 인간을 해방하고 인간성을 추구하기 위해 개발해온 인간 친화적 로봇이 점차 업그레이드되어 이제는 인공지능이 인간과 유사한 속성을 띠면서도 인간을 능가하는 능력을 지닌 존재로 진화해나가고 있다. 그러한 가운데 우리는 인공지능이 인간의 본질 및 인간성에 반하는 기제로 작용하게 되는 상황에 직면하고 있다.

인공지능과는 다른 인간만의 고유한 특성에 대한 인식과 인간성의 추구는 인공지능 시대를 살아갈 학생들에게 필요한 교육영역임에 틀림이 없을 것이다. 그런데 인공지능 시대는 현재 구성되어가고 있는 단계이기 때문에 많은 부분에 있어서 아직은 구현되지 않은 미래적 상황에 놓여 있다. 인공지능 시대의 구체적인 면모는 과학기술 분야의 비전이나 기술발전에 의거한 유추, 혹은 상상에 의해 묘사되고 있다. 이러한 현실을 고려할 때 장차 인공지능 시대에 인공지능과 마주한 인간이 직면하게 될 인간성에 대한 고민을 다룬 문학작품을 통해 인공지능 윤리교육을 수행하는 것은 유용한 방법이 될 수 있다.

문학은 인류의 역사 이래로 인간의 삶을 총체적으로 담아내면서 인간의 본질을 파헤치는 데 힘을 쏟아왔다. 이러한 문학은 이미 오래전부터 인공지능 시대를 살아가는 인간이 겪게 될 문제들을 구체적이고 심도 있게 다루어왔다. 따라서 인공지능 시대를 다룬 문학작품을 통해 학생들로 하여금 미래 상황을 추체험하게 하고 인간의 본질과 인간성에 대해 학습할 수 있도록 지도하는 것은 유의미한 교수법이라 할 수 있다.

초등학생들에게 문학작품을 통해 인간성에 대한 인식을 함양하도록 지도할 때 활용할 수 있는 사례로 『담임 선생님은 AI』라는 동화를 들 수 있다. 이 책의 개략적인 내용은 아래와 같다.

미래초등학교 5학년 1반에 인공지능 담당 교사가 배정된다. 학생들은 처음에는 인공지능 담당 교사의 정확하고 유창한 답변에 만족해하지만 시간이 지나면서 불만을 품게 된다. 학생들의 잘못된 행동에 대해서 이유나 상황을 고려하지 않은 채 가차 없이 벌점을 부과하고, 쉬는 시간에도 화장실에 가지 않은 채 학생들을 감시하는 담당 교사 때문에 학생들은 스트레스가 쌓인다. 자신들은 공부하는 기계가 아니라고 하면서 담당 교사에게 저항한다. 그러나 우여곡절을 겪어나가면서 인공지능 담당 교사에게 감동한 학생들은 이제 오히려 다른 반 학생들의 가해로부터 담당 교사를 지켜내기 위해 단합한다. 학생들은 인공지능 담당 교사에게 다른 반 학생들의 편견을 없애기 위해서 왕따가 된 담당 교사에게 인간다운 행동들을 학습시키는 한편 함께 동화를 읽어 상상력을 키워주기로 한다. 어느 날 한 학생이 이 세상이 없는 새로운 동화를 만들어 담당 교사에게 읽혔는데 갑자기 인공지능 담당 교사가 격렬한 반응을 일으키며 오작동한다. 인공지능의 뉴로모픽 칩이 인간적인 감정에 격렬히 반응하면서 과부하가 걸려 나타난 증상이었다. 이에 인공지능 담당 교사는 폐기될 위기에 처하고, 학생들은 담당 교사를 살리기 위해 교장실에 잠입하여 인공지능 담당 교사의 신경망에 접근한다. 결국, 담당 교사를 향한 학생들의 애정어린 마음에 감동한 교장은 인공지능 담당 교사 김영희의 폐기를 취소하고 수리하여 다시 학생들에게 돌려준다(이경화, 2021).

초등학교에 인공지능 담당 교사가 배정되면서 벌어진 갖가지 흥미로운 사건들과 학생들의 심리적 변화를 구체적으로 묘사하고 있는 이 작품은 인공지능 시대에 ‘인간다움’을 탐구하는 동화이다. 이 책을 활용한 인공지능 윤리교육의 구체적 과정 및 절차를 제시해 보면 다음과 같다.

① 읽기 전 활동

(가) 그림을 보고 등장인물들이 무엇을 하고 있는지, 이 책은 어떤 내용인지 예측하여 이야기한 후 적는다.

〈표 4〉 그림 보고 책 내용 예상하기(자료)

- (나) 책표지에서 볼 수 있는 것들에 대해 말해 본다.
- (다) 차례에서 볼 수 있는 것들에 대해 말해 본다.
- (라) 책표지와 차례에서 관찰한 것들을 토대로 책에 관한 생각이나 궁금한 점들을 적는다.
- (마) 만약 자신의 담당 교사가 인공지능이라면 어떻게 말해 본다.
- (바) 책 읽는 방법을 정한다. 혼자 묵독하기, 실감 나게 낭독하기, 친구들과 번갈아 가며 읽기, 혼합 방식으로 읽기, 혹은 그 이외의 다른 방법도 제안할 수 있다.

② 읽기 중 활동


- (가) 낱말의 뜻을 찾아 알아둔다. 일례로 ‘로봇 3원칙’이란 무엇인지 38쪽에 설명된 내용을 미리 이해한다. 이 낱말의 뜻 찾기 활동은 학생들의 수준을 고려하여 과제를 통해 미리 준비해오게 해도 무방하다.
- (나) 등장인물의 이름과 별명을 정리해 본다.
- (다) 소재목별로 내용을 요약해 본다.
- (라) 학생들이 책 내용을 파악하고 있는지 중간중간 점검한다. 예를 들면, 91~93쪽의 공단비가 창작한 동화 <코노피오>를 읽은 후, 코노피오는 왜 인간이 되기를 거부했는지를 질문할 수 있다.

③ 읽기 후 활동

책을 읽고 난 후 곧바로 토론으로 들어가면 학생들은 부담을 느껴 분위기가 경직될 수 있으므로 먼저 책 내용과 관련된 가벼운 퀴즈 활동 등을 통해 편안한 분위기를 형성할 필요가 있다.

- (가) 우선 간단한 퀴즈를 통해 책 내용을 다시 한번 상기하고 이해하게 한다. 예를 들면, 카드 앞면에 책 내용과 관련된 퀴즈와 쪽수를 적고 뒷면에 정답을 적은 카드를 각자 4개씩 준비한 후 학생들끼리 질문하여 맞으면 카드를 가져가게 하여 제한된 시간 내에 가장 많은 카드를 획득한 학생이 최종 우승자가 되는 퀴즈 활동을 할 수 있다.
- (나) 책의 내용을 줄거리 중심이 아닌 인물이나 사건 중심으로 정리할 수 있도록 한다. 예를 들면, 인공지능 담당 교사의 능력과 특성은 무엇인가? 공단비는 왜 인공지능 담당 교사를 좋아하게 되었나? 박한별은 왜 인공지능 담당 교사를 싫어했는가? 한민아 선생님 사건의 자초지종은 어떤 것인가? 등의 질문을 통해 인물이나 사건 중심으로 내용을 재정리하여 정확히 파악할 수 있도록 돕는다.
- (다) 자신만의 인공지능 로봇을 만든다면 어떤 로봇을 만들고 싶은지 말해 본다.
- (라) 자신만의 인공지능 로봇 또는 인공지능 담당 교사의 사용설명서를 작성해 본다.

〈표 5〉 인공지능 사용설명서(자료)

	달림 선생님 사용 설명서	
	1.	
	2.	
	3.	
	4.	
	5.	
	6.	
	7.	

(마) 이 책 82쪽에 교장 선생님이 던진 물음인 “과연 로봇에게도 윤리가 적용되는가?”라는 주제로 토론을 진행한다.

(바) 사람을 사람답게 만드는 것은 무엇인지 책에서 찾아 이야기한다. 예를 들면, 101쪽에서 교장 선생님은 ‘자기성찰’이라고 했고, 165쪽에서 작가는 ‘의인화, 감정이입, 애착’이라고 했다. 이밖에 책의 내용을 토대로 하여 과연 ‘인간성’이란 어떤 것인지 토론한 후 정리하고, 학생들 각자 자신만의 인간성 실천방안을 이야기하고 적는다.

이상에서 인간성의 추구에 대한 문학작품을 활용한 인공지능 윤리교육 방안을 살펴보았다. 인공지능 윤리교육은 인공지능 시대에 더욱 필요성이 요구되는 인간의 본질에 대한 인식을 제고하고, 나아가 학생들로 하여금 인간성을 함양할 수 있도록 돕는 방향으로 이루어져야 한다. 학생들이 상상하는 인공지능 교사의 장단점을 미리 알아보는 단순한 차원을 넘어서서, 인공지능 시대에 인간의 역할, 교사와 학생의 인격적 유대관계의 중요성, 더 궁극적으로는 인공지능이 대체할 수 없는 ‘인간성에 대한 탐구와 실천’으로 나아가야 한다.

IV. 결론

미래 교육은 개인의 이익과 성장을 위해 노력하는 학생이 아니라, 나와 사회의 건전한 성장을 위해 노력하는 학생을 길러내는 방향으로 전환되어야 한다. 현재 우리 사회는 잘못된 알고리즘 기반의 인공지능으로 인해 생겨난 분열되고 양극화된 사회의 문제점들이 생겨나고 있는데 역설적으로 우리는 그러한 문제 해결책을 인간이 아닌 인공지능에서 찾고 있지는 않은지 자문해 보아야 한다. 인공지능으로 생겨난 윤리적 문제의 최종적이며 궁극적 해결의 주체는 그 누구도 아닌 인간 자신이라는 점을 인식하고 이에 대한 깊은 성찰과 고민이 계속되어야 할 것이다.

인공지능 윤리의 문제는 그 해결을 위해 현재의 문제에 대한 반성이 아닌 미래 발생 가능한

문제에 대한 상상력이 요청된다. 인간과 인공지능의 공존을 전제로 한다면 인공지능으로 인해 생겨나는 다양한 문제들은 미래에 우리에게 다가올 것으로 그러한 문제 해결에 필요한 비판적 사고와 숙고 능력을 길러 줄 수 있어야 하며 그것이 바로 인공지능이 일상이 되어 살아가게 될 미래 세대들을 위해 도덕 및 윤리교육이 담당해야 할 부분이기도 하다.

혹독한 환경 속에서 살아남은 사람은 신체적으로 탁월한 능력을 지닌 인간이 아니라 절망적 상황에서도 끊임없이 호기심을 바탕으로 지적 추구를 해 온 사람이라는 사실을 우리는 인류의 역사를 통해 쉽게 알 수 있다. 도구를 만들어 사용하는 존재인 인간의 도구 중에서 인간 사회에 가장 강력한 영향을 끼친 것 중의 하나가 디지털 도구이다. 디지털 도구의 활용으로 인해 인간은 편리함과 풍요를 누리게 되었지만, 그 이면에는 일찍이 없던 ‘격차 사회’로의 도래가 성큼 다가왔다. 기술 시스템을 설계, 운영하는 사람과 집단에게는 최고의 권력과 풍요를 안겨주지만, 기술 변화가 가져온 새로운 사회의 질서를 파악하지 못하는 사람들은 그 기회를 발견하기도 활용하기도 어렵게 되었다(구본권, 2020: 20-21). 그러므로 인터넷과 컴퓨터에 기반한 인공지능을 설계하고 운영하여 활용하는 사람들의 역할이 중요하며 특히 그들의 가치관과 세계관에 대한 교육이 무엇보다 필요한 시점이 되고 있다. 이들이 어떤 목적으로 어떤 것을 선정하여 그것을 어떻게 활용할 것인가를 결정하는 중요한 역할을 하고 있기 때문이다. 이러한 맥락에서 본 연구에서는 인공지능의 목적과 내용 그리고 교육 방법 등에 대한 시론적 차원의 시도를 해 보았으며 추후 이러한 연구가 계속되기를 기대해 본다.

※ 논문 투고일: 2022. 1. 10. ※ 논문 수정일: 2022. 2. 16. ※ 게재 확정일 : 2022. 2. 21.

〈참고문헌〉

- 과학기술정보통신부(2020). **사람이 중심이 되는 인공지능(AI) 윤리기준(안)**.
보도자료(2020.11.27)
- 구본권(2020). **로봇시대 인간의 일**. 서울: 어크로스.
- 김효은(2019). **인공지능과 윤리**. 서울: 커뮤니케이션북스.
- 변순용(2020a). AI 윤리교육의 필요성에 대한 연구. **한국초등교육**, 31(3), 153-164.
- _____ (2020b). AI 시민성 교육에 대한 시론. **초등도덕교육**, 67, 427-445.
- 변영계, 김광휘(2002). **협동학습의 이론과 실제**. 서울: 학지사.
- 이경화(2021). **담임 선생님은 AI**. 경기: 창비.
- 이상욱(2021). **유네스코 인공지능(AI) 윤리 권고 해설서 인공지능 윤리 이해하기**. 한국유네스코 위원회.
- 이원태(2017). **인공지능 윤리담론의 지형 왜 알고리즘 윤리인가**. 디지털사회통합정책연구센터.
- 이지성(2020). **에이트**, 서울: 차이정원.
- 정지은 율김(2021). **공감은 지능이다**. 경기: 푸른 숲.
- 한국정보화진흥원(2018). **지능정보사회의 윤리 가이드라인**. 정보문화포럼.
- 한국정보화진흥원(2019). **인공지능 윤리 가이드라인-일본과 EU사례를 중심으로**. 지능정보사회 법제도 이슈리포트.
- 한찬희(2020). 초등 도덕과를 통한 인공지능윤리교육 방안 모색. **초등도덕교육**, 69, 229-253.
- EU(2019). *Ethics Guidelines for Trustworthy AI*. HLEG ON AI.
- 인공지능, 두산백과. https://www.doopedia.co.kr/doopedia/master/master.do?_method=view&MAS_IDX=101013000752124 (검색일자, 2022.01.02.)
- <http://www.aitimes.kr/news/articleView.html?idxno=23359>
- http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html
- <https://riverxchange.com/2015/05/03/riverxchange-video-projects-showcase-21st-century-skills>
- Habermas, J.(1987). *Theorie des kommunikativen Handelns*; 장춘익 역(2006). **의사소통행위이론**
1. 경기: 나남.
- _____ (1987). *Theorie des kommunikativen Handelns*; 장춘익 역(2006). **의사소통행위이론**
2. 경기: 나남.
- Zaki, Jamil.(2019). *The War for Kindness*; 정지은 역(2021). **공감은 지능이다**. 경기: 심심.

〈Abstract〉

Ethics Education in Relation to Artificial Intelligence as Part of Moral Education in Elementary School

Shin, Hyunhwa¹, Shin, Taesoo², Shin, Hyunwoo³

Undeniably, artificial intelligence (AI) has become a part of modern life, because its use offers convenience and benefits. However, it faces various problems with several issues that require solutions. Children are no exception to this reality.

Therefore, this study intended to theoretically explore the values and norms specific to children who are facing the age of AI. We conducted an investigation of the actual situation of ethics in relation to AI and the negative and positive effects and dilemmas associated with such ethics. Based on the results, the study proposed several methods for effectively teaching AI ethics and provided suggestions for the future direction for AI ethics as part of moral education in elementary school.

Therefore, this thesis consists of the theoretical exploration of ethics in relation to AI and presents the goals and contents of and methods for teaching for AI ethics.

Key words : artificial intelligence, artificial intelligence ethics, digital literacy, cooperative learning, metaverse, rational communication, humanity.

1. Teacher, Seogok Elementary School, sh2@knue.ac.kr (Lead Author)

2. Associate Professor, Hannam University, penpipe@hnu.kr (Co-Author)

3. Professor, Cheongju National University of Education, shinnew@cje.ac.kr (Corresponding Author)