



인공지능윤리교육의 난점과 과제

박대호¹

《 요 약 》

이 글은 인공지능윤리교육의 난점이 ‘교육은 시대의 변화에 따라야 한다’는 통념에 바탕을 두고 있다는 문제의식에서 출발한다. 그러한 통념은 시대의 변화에는 주목하지만 정작 교육의 의미와 가치를 소홀하게 한다는 문제를 안고 있으며, 이 문제는 인공지능윤리교육 전체를 그릇된 방향으로 이끌 가능성이 있다. 인공지능 시대의 교직이라는 주제는 그 가능성을 구체적인 형태로 보여준다. 인공지능 시대의 도래는 교사들에게 ‘지식의 전달’이 아니라 ‘인간적 상호작용’이라는 이름으로 새로운 역할을 모색해야 한다는 문제의식을 안겨 주었다. 그러나 이러한 이해는 지식의 전달이 곧 ‘정보의 전달’이라는 그릇된 전제 위에서만 성립될 수 있다. 교육이 시대의 변화를 따라야 한다는 생각은 초등 도덕과를 통한 인공지능윤리교육에도 영향을 미치고 있다. 도덕과에 새롭게 도입된 인공지능윤리교육 관련 내용 요소는 특정한 가치·덕목으로 초점화되어 있는 기존의 체계와 상이한 성격을 갖고 있으며, 이 문제는 학교 현장에 혼란과 어려움을 불러일으킬 가능성이 크다. 이러한 사정은 오늘날의 우리에게 ‘인공지능이 먼저인가, 교육이 먼저인가’라는 근본적인 질문을 제기하게 한다. 결국, 인공지능윤리교육의 성공적인 안착을 위해서는 교육과 교사, 도덕과교육, 수업과 같은 기초에 해당하는 개념들의 의미와 가치에 우선적인 관심을 기울일 필요가 있다.

주제어 : 인공지능윤리교육, 도덕과교육, 시대 변화와 교육, 범주오류, 교사의 역할

1. 청주교육대학교 조교수, dhpark@cje.ac.kr

I. 서론

‘변화의 시대’는 오늘날을 규정하기에 더없이 적절한 표현이다. 인류의 역사상에 변화라는 말이 거론되지 않은 시대는 없었겠지만, 이렇게 급격한 속도로 일어나는, 심지어 일일이 열거하기 어려울 만큼 다양하고 거대한 변화를 마주한 시대는 지금이 거의 유일하다고 보아야 할지도 모른다. 제목으로 내어 건 ‘인공지능’은 교육 분야가 마주하게 된 일체의 변화, 특히 현재 진행 중인 변화를 주도하고 이끄는 핵심 가운데 하나이다. 인공지능으로 대표되는 변화의 흐름은 이제 새로운 교육내용과 새로운 교육방법을 요구하는 것을 넘어서서 교육을 담당하는 교사의 새로운 역할까지 요구하는 목소리로 표출되고 있다. 그 목소리는 한 마디로 ‘교육은 시대의 변화에 발맞추어야 한다’라고 요약해도 무리가 없다.

위의 요약은 결코 오늘날의 교육이 나아가야 할 방향을 그릇되게 표현하는 것이 아니다. 그것은 교육이 시대의 요구를 수용하고 변화해야 한다는 점을 정확하게 표현하고 있다. 그러나 그 문장은 일정한 오해의 여지를 안고 있다. 그 문장은 해석의 방향에 따라 두 가지 다른 뜻으로 이해될 가능성이 있기 때문이다. 먼저, 가장 가능성이 큰 해석의 방향으로서, 그 문장은 현재의 교육이 시대의 요구에 부응하지 못하는 면이 있고, 그리하여 그것은 개선을 넘어 ‘개혁’의 대상이 되어야 한다는 의미로 이해될 수 있다. 최신의 것, 현대적인 것을 가르치는 교육, 즉 ‘새 시대의 새 교육’은 이러한 이해의 방향을 요약하는 표현이 될 만하다. ‘새롭다’라는 말에는 차이와 이질성에 대한 강조라는 의미요소가 불박여 있으며, 따라서 새로운 교육이라고 말할 때 그 교육은 기존의 교육을 대체하는 대안적 교육이라는 의미로 이해될 가능성이 크다. 기존의 교육으로는 인공지능 시대의 도래로 새롭게 제기되는 문제에 대응하기 어려운 만큼 현재에 맞게 교육을 바꿔야 한다는 생각이 여기에 반영되어 있다고 말해도 좋을 것이다.

2020년부터 본격적인 연구주제가 되어 온 ‘인공지능윤리교육’ 분야의 연구 동향은 이러한 이해의 방향을 얼마간 예시하고 있다. 시대의 변화에 발맞춘 인공지능윤리교육의 필요성을 주장하는 연구(변순용, 2020; 정창우, 이해진, 2022; 박형빈, 2024), 인공지능 시대에 새롭게 요구되는 윤리 원칙과 가이드라인의 적용방안을 탐색한 연구(권누리, 2024; 박형빈, 2024), 나아가 2022 개정 도덕과 교육과정과의 관련을 염두에 두고 인공지능윤리교육을 탐색한 연구(김하민, 2021; 홍현주, 2021; 박보람, 2025), 인공지능윤리교육의 실질적 방안을 제시한 연구(신현화, 신태수, 신현우, 2022) 등을 포함한 여러 연구의 기저에는 현재의 교육이 시대의 흐름에 민첩하게 반응해야 한다는 공통된 문제의식이 깔려있다. 이러한 문제의식은 그 자체로 타당성을 가지고 있으며, 따라서 여기에 이의를 제기하는 것은 무의미한 일이라고 보아야 한다. 그러나 교육이

시대의 변화에 발맞추어야 한다는 말이 언제나 ‘새 시대의 새 교육’이라는 식으로 해석되어야 하는가에 관해서는 깊이 생각해 볼 필요가 있다. 아닌 게 아니라, 이제는 교육이 도저히 거스를 수 없는 시대의 흐름에 몸을 맡길 수밖에 없다고 체념하면서-또 그 흐름을 따라가지 못할 때 뒤쳐질 수밖에 없다는 식으로 시급성을 특별히 강조하면서-교사의 역할이 장차 ‘학습 설계자’로 전환될 것이라는 주장이 공공연하게 제기되고 있다.¹⁾

이 글은 인공지능윤리교육 관련 연구가 ‘교육’ 자체에 관한 관심을 소홀히 하였다라는 문제의식에서 비롯된 것이다. 실지로 ‘인공지능 시대의 도래’라는 시대적 상황은 우리에게 교육의 안쪽이 아니라 바깥쪽을 바라보도록 부추기고 있다. 교육의 바깥쪽에 서서 교육의 안쪽을 바라볼 때, 교육은 뜯어고쳐야 할 대상으로 간주될 수밖에 없다는 점은 애써 강조할 필요가 없다. 그러나 문제는 ‘변화의 시대’라는 말과 거기에 입각한 주장이 특정 시대, 특정 장소를 막론하고 제기되어왔지만, 교육 그 자체는 변화할 수 없다는 데 있다. 달리 말해, 교육은 시대의 변화에 따라 ‘다른 양상’으로 나타나지만, 그러한 변화에도 불구하고 불변하는 본질로 말미암아 ‘동일한 교육’으로 남아 있다. 교육의 본질과 양상을 구분하는 일은-교육의 양상을 본질로 착각하는 일과 같은 본말전도가 교육에 재앙을 불러온다고 말할 수 있을 정도로-사소한 문제가 아니다. 그것은 교육이라는 것이 시작된 이래로 항구적인 문제로 간주되어 왔으며, 이 문제는 이 글이 드러내고자 하는 또 하나의 해석의 방향과도 관련을 맺고 있다. ‘현대 과학적 교육학의 아버지’라는 위상을 널리 인정받은 헤르바르트(J. F. Herbart)는 「일반교육학」 서론에서 당시의 교육을 ‘시대의 노리개’에 비유하면서, ‘사소한 것에 매달리는 것’은 교육을 타락시킨다고 말하였다(Herbart, 1982, 92). 적어도 그의 견해는 교육 그 자체에 관한 성찰 없이 수용되는 견해가 교육의 왜곡을 불러올 가능성이 있다는 점을 시사한다. 이 시사는 오늘날에도 여전히 유효하고 타당하다. 실지로 ‘디지털이 먼저인가, 교육이 먼저인가?’라는 비에스타(G. Biesta)의 질문은 오늘날의 우리가 어디로 향하고 있는가를 되돌아보게 한다(Biesta, 2020). 이하에서 살펴보겠지만, 여기에서 언급한 왜곡의 가능성은 인공지능윤리교육을 다루는 ‘도덕과’도 예외가 될 수 없다.

이상의 문제의식을 바탕으로 하여, 이 글은 초등 도덕과를 통한 인공지능윤리교육의 난점이 ‘시대의 변화와 교육의 관계’에 관한 특정한 이해에서 비롯된다는 가설에서 출발한다. 이 가설의 타당성을 확인하기 위해서는 먼저 ‘교육은 시대의 변화를 따라야 한다’라는 통념에 가까운 명제가 어떻게 교육에 관한 그릇된 이해로 연결되는가를 검토할 필요가 있다. 이를 위해, 2장에서는 ‘인공지능 시대의 교직’이라는 주제를 분석의 대상으로 삼는다. 이 주제는 ‘인공지능윤리교육의

1) 교사의 역할이 장차 학습 설계자로 전환될 것이라는 주장을 내세우는 다수의 연구들을 일일이 제시하기는 어렵겠지만, 적어도 2018년 발간된 OECD 보고서(Teachers as Designers of Learning Environments: The Importance of Innovative Pedagogies)는 이러한 경향을 요약적으로 보여준다고 말해도 좋을 것이다.

난점과 과제’라는 본 연구의 주제에서 벗어나는 것처럼 보이지만, 다른 어떤 주제보다 이 글을 떠받치고 있는 아이디어-‘즉, 시대의 변화와 교육의 관계’-를 구체적인 형태로 드러내 줄 것이다. 이어지는 3장에서는 초등 도덕과를 중심으로 하여 시대의 변화에 대한 지나친 관심이 인공지능윤리교육에 어려움을 불러일으키는 근본적인 원인이라는 점을 살펴본다. 마지막으로 결론에서는 본문의 논의를 간략하게 정리하고, 앞서 언급한 그 원인과의 관련 속에서 인공지능윤리교육의 향후 과제를 제안하면서 글을 마치고자 한다.

II. 인공지능 시대의 교직

기계가 인간의 전유물로만 여겨지던 다양한 일을 대신하게 되리라는 기대와 그것에 따른 우려는 어제오늘의 것이 아니다. 실지로 인공지능으로 대표되는 현대 과학기술은 그 기대와 우려가 더 이상 상상의 영역에 머물러 있지 않다는 점을 보여주고 있다. 기계가 반복적인 단순 노동을 대체한 것은 이미 역사적으로 확인된 사실이요, 이제 기술은 오로지 인간에게 허락된 것이라고 생각했던 다양한 분야의 일을 대체하고 있다. ChatGPT로 대표되는 생성형 인공지능이 우리에게 확인시켜주고 있듯이, 기술은 글쓰기와 그림 그리기 같은 예술까지 그 활동 영역을 확장하고 있다. 인공지능은 이미 우리의 삶에 자연스럽게 녹아들어 있으며, 인간의 삶의 방식을 근본적인 수준에서 변화시키고 있다. 인공지능 시대의 도래는 인간으로 하여금 자신의 삶, 나아가 자기 자신을 되돌아보는 일을 ‘강제’하는 수준에 이르렀다고 말해도 무리가 없을 정도이다.

교육계에 몸담고 있는 사람이라면, 이상의 간단한 논의를 바탕으로 하여 다음과 같은 질문을 떠올렸을지도 모른다: 과연 인공지능은 교사를 대체할 수 있는가? 개인적인 경험에 비추어 보면, 이 질문은 각자의 행편과 입장에 맞추어 대답되는 경향이 있다. 예컨대, 에듀테크 분야에 종사하고 있는 사람은 어떤 대답을 하겠는가? 사실상 그들은 이미 그것의 가능성을 인정하고 그것의 실현을 자신의 목적으로 삼고 있다. 그들에게 인공지능으로 교사를 대체한다는 발상은 기이한 것도 아니요, 새로운 것도 아니다. 그 대체는 그들이 장차 성취해야 할 궁극적 목표인 것이다. 그러나 그 일이 당장 실현될 것 같지는 않다. 실지로 그런 사람들 가운데 하나인 카네기 러닝(Carnegie Learning)의 창업자 스티브 리터(S. Ritter)는 2021년 8월 미국 샌디에이고에서 개최된 에듀테크 컨퍼런스 ‘ASU+GSV 서밋’에서 이 주제와 관련된 놀라운 이야기를 털어놓은 바 있다.²⁾ 그의 논의는 인공지능과 교사의 역할에 관한 중요한 시사를-그것도 ‘에듀테크 개발

2) 이후 등장하는 스티브 리터의 발언에 관한 논의는 다음의 영상을 참고한 것임을 밝힌다.

자'의 입장에서-제공해 준다는 점에서 주목할 만한 가치가 있다.

그는 자신의 회사에서 개발한 수학 학습 도구(MATHia)를 학생들에게 제공해 주고 어떤 시점에 그것의 사용을 중단하고 교사에게 도움을 요청하는지를 관찰하였다. 학생들이 MATHia의 사용을 중단하는 것은 자신들이 개발한 학습 도구에 모종의 결함이 있다는 뜻이기 때문에, 그들에게 이 문제는 대단히 중요한 것이었다. 그는 이 관찰을 통하여 세 가지 시사점³⁾을 얻었다고 말하지만, 이 글의 맥락에서 중요한 것은 첫 번째 시사점이다. 그가 얻은 이 첫 번째 시사점을 약간의 수사를 곁들여 설명하면 다음과 같다. 실험 결과 그가 발견한 것은 참으로 기이한 것이었다. 학생들이 교사에게 가서 질문한 내용은 '누구나 예상할 수 있는 패턴의 수학 문제'였기 때문이다. 즉, 학생들은 인공지능이 문제를 풀지 못하거나, 설명을 어렵게 한다는 이유로 교사를 찾아간 것이 아니었다. 심지어 인공지능은 그들에게 부여된 과제가 성공적으로 해결되었다는 점을 알리고 있었다. 그럼에도 불구하고, 학생들은 MATHia의 사용을 중단하고 교사를 찾아갔다. 그들이 교사를 찾아간 이유는 무엇인가? 앞서 '놀라운' 대답이라는 표현을 쓴 것은 바로 그 이유 때문이다. 학생들은 단지 교사가 고개를 끄덕여 자신을 인정해 주기를 바라고 있었다. 실지로 리터는 학생과 대면한 교사가 '응, 그래, 잘 하고 있구나'와 같은 인정과 격려의 성격을 띠는 형태의 발언을 했다고 말한다. 이야기의 맥락으로 짐작하면, 그 발언을 들은 학생은 만족스러운 표정으로 자신의 자리로 돌아갔을 가능성이 크다. 리터는 이 발견을 증거로 하여 '인간적 상호작용'의 성격을 띠는 대화야말로 인공지능이 대체할 수 없는 영역이라고 고백한다.

리터의 고백은 곧장 우리에게 '모라벡의 역설'을 떠올리게 한다. 즉, '인간에게 어려운 것이 인공지능에게 쉽고, 인간에게 쉬운 것이 인공지능에게 어렵다'(Hard problems are easy and easy problems are hard)는 말이 이렇게 적절하게 들어맞는 경우가 있다면, 그것은 바로 교육이 될 수밖에 없는 것으로 보인다. 실험과 데이터에 기반한 리터의 설명에 나타나 있는 바와 같이, 인간에게는 너무나 자연스러운 '인간적 상호작용'이 인공지능에게는 너무나 어려운 일이다. 생각해 보라. 어떤 학생이 인간이 아닌 인공지능의 인정과 격려에 만족할 수 있겠는가? 그것으로 충분하다고 생각하는 학생은 적어도 인간이 사회적 존재라는 사실을 스스로 부정하는 것이다. 그렇다면 인공지능은 교사를 대체할 수 있는가? 인간은 인공지능이 도저히 해낼 수 없는 일을 할 수 있는 만큼, 거기에 우리는 고개를 가로저어야 할지도 모른다. 그러나 그렇지 않다. 리터의

https://www.youtube.com/watch?v=r-i51Bdb_zg/

3) 이 시사점은 다음과 같이 요약할 수 있다. 첫째, 인공지능은 인정과 격려가 포함된, 그리하여 '인간적 상호작용'의 성격을 띠는 대화를 대체할 수 없다. 둘째, 인공지능의 활용은 학습에 실질적 효과가 있다. 셋째, 교사 자원의 분배가 비효율적이다. 즉, 고작 1%의 학생—심지어 교사의 도움이 반드시 필요한 것이 아닌 학생—이 교사 자원, 구체적으로 말하여 그들의 시간 가운데 40%를 독차지한다.

발견과 고백은 다분히 ‘에듀테크 개발자’의 입장의 것이요, 그리하여 교사의 역할에 관한 심각하고 그릇된 오해를 불러일으킨다. 그리고 이 오해는 우리가 제대로 의식하지도 못한 사이에 문제 전체를 엉뚱한 방향으로 이끌고 간다.

리터의 설명에 귀 기울인 교사 가운데 몇몇은 고개를 끄덕이면서 ‘아직 우리가 할 수 있는 일이 남아 있다’라는 투의 확신을 갖게 되었을지도 모른다. 그러나 그러한 확신은 도리어 교사가 스스로 교직의 의미와 가치를 격하시키는 역설적인 결과를 가져온다. 인정과 격려가 교사의 역할이라면 그것은 누구나 할 수 있는 일로 되기 때문이다. 정서적 돌봄의 측면에서 약간의 훈련을 받아야 할 필요가 있을지도 모르지만, ‘응 그래, 잘 하고 있구나’라는 말과 함께 인정과 격려를 해주는 일은 그다지 전문적인 일로 보이지 않는 것이다. 그리고 지식의 전달이라는 과업을 인공 지능에게 맡겨두고, 인정과 격려의 역할을 담당하면서 거기에 만족하는 교사를 상상하기는 쉽지 않다. 엮힌 데 덮친 격으로, 에듀테크 개발자들은 교사를 인공지능으로 대체하겠다는 원대한 포부를 품고 있는 듯하다. 과연 그들이 ‘인간적 상호작용’이라고 말했던 그 일을 인간의 영역으로 계속 남겨두려고 할지는 의문이다.

그렇다면 다시, 인공지능은 교사를 대체할 수 있는가? 리터의 발언을 그대로 받아들일 때 오늘날의 교사들은 대체되지 않을 이유가 없다. 그러나 ‘교육의 자동화’라고 명명해도 좋을 이러한 대체는 자연스러운 것이 아니다. 그 대체에 교사의 본래적 역할에 관하여 깊이 성찰하지 않는 교사 자신이 적극적으로 힘을 실어주고 있기 때문이다. 아닌 게 아니라, ‘지식의 전달’이 아니라 ‘인간적 상호작용’이라는 이름으로 교사가 해야 할 일을 찾아야 한다는 생각의 밑바닥에는 전자를 중심으로 하는 교사를 후자를 중심으로 하는 교사로 바꾸어야 한다는 발상-이른바, 교사의 자기부정적 발상-이 붙박여 있다. 물론, 오늘날의 ‘모든’ 교사가 인공지능에 의해 대체되리라는 주장은 대단히 의심스러운 것이다. 그러나 교사 가운데 ‘일부’가 대체되리라는 생각은 이미 오래 전부터 학자들 사이에 일정한 공감대를 형성하고 있었다고 말해도 무리가 아니다. 예컨대, 브루너(J. Bruner)는 장차 기계가 교사를 대체하지 않을 것이라고 단언하면서도 “실상 수업에서 귀찮은 부분을 기계에 맡길 수 있으면 있을수록 그만큼 유능한 교사가 더 필요하게 될 가능성이 있다”(Bruner, 2006, 365)라고 말하였다. 이 말이 어떤 교사에게는 희망적인 메시지로 들릴지 모른다. 그러나 그 말을 ‘희망적인 메시지’로 받아들이는 사람은 어디까지나 ‘유능한 교사’라고 보아야 하며, 그 반대편의 ‘형편없는 교사’에게는 그 말이 ‘절망적인 메시지’로 들리지 않으리라는 보장이 없다. 그 말의 핵심은 아마도 중간언어와 지식의 구조라는 브루너의 유명한 구분을 활용하여 다음과 같이 요약해도 좋을 것이다. 즉, 지식의 구조를 전달하지 못하는 교사, 달리 말해 중간언어의 전달에 매몰되어 있는 교사는 대체될 수밖에 없는 처지에 놓여 있다. 실지로

이 문제에 관한 한, 가장 체계적이지도 최신의 논의를 하고 있는 셀윈(N. Selwyn)은 자신의 책 「로봇은 교사를 대체할 것인가?」에서 ‘모든 인간이 좋은 교사인 것은 아니며, 그리하여 인간 교사를 맹목적으로 옹호하지 말아야 한다’라는 견해를 피력한 바 있다(Selwyn, 2022, 35).

그렇다면, 이제 교사는 결국 인공지능에 의하여 대체되고 마는 것인가? 그렇지 않다. 이 글이 강조하고자 하는 바는 우리가 그 질문을 특정한 방식으로 이해하고 있다는 데 있다. 즉, 리더의 사례가 보여주듯이, ‘인공지능이 교사를 대체할 수 있는가’ 하는 질문은 마치 ‘지식의 전달은 인공지능에게 맡기고 그 이외에 교사가 해야 할 일은 무엇인가’ 하는 식으로 이해되는 경향이 있다. 그러나 ‘지식의 전달’을 빼고 나면 교사는 무슨 일을 해야 하는가? 셀윈은 교육과 관련된 기술에는 교육의 의미에 관한 암묵적인 가정이 담겨 있다는 점을 예리하게 지적한다(Selwyn, 2022, 130). 그의 지적을 고려하면, 지식의 전달을 인공지능에게 맡기겠다는 식의 이해는 지식의 전달이 곧 ‘정보의 전달’이라는 그릇된 전제 위에서만 성립될 수 있다. 그러나 지식의 전달은 마치 손에 물건을 쥐어주듯이 이루어지는 것이 아니며, 교사가 지식을 전달하는 이유는 수많은 정보 수준의 지식을 정확하게 되풀이할 수 있는 학생을 기르기 위해서가 아니다. 화이트헤드(A. N. Whitehead)가 교육에서 가장 주의해야 할 문제는 ‘무기력한 관념’(inert ideas)이라고 말한 이유가 바로 여기에 있다(Whitehead, 1967, 1). 교사가 지식을 전달하는 이유는 그러한 지식을 스스로 찾아내고 조직화하고 응용할 수 있는 ‘능력’을 기르기 위해서이기 때문이다. 라일(G. Ryle)이 지적한 바와 같이, 심지어 정보를 전달받는 일과 능력을 배우는 일 사이에는, 전자가 ‘한순간’에 일어나는 일인데 반하여, 후자는 ‘점진적’, ‘연속적’으로 이루어질 수밖에 없는 일이라는 현격한 차이가 놓여 있다(Ryle, 1994, 74). 예컨대, ‘물은 100°C에서 끓는다’라는 정보는 그것을 전달하는 순간 피전달자에게 그대로 수용되지만, 그러한 정보를 스스로 찾아내고 조직화하고 응용할 수 있는 능력은 장기간에 걸친 교육을 통해 습득될 수밖에 없다. 이 구분에 의하면, 교육을 지엽적이고 단편적인 지식, 즉 정보의 전달로 간주하는 것은 교육이 ‘한순간’에 일어나는 일이라고 말하는 것과 꼭 같이 불합리하다.

여기에 더하여, 지식을 가르치고 배우는 일과 인간적 상호작용이 별개의 일이라는 통념 또한 가르치는 일에 관한 오해에서 비롯된 것이라고 보아야 한다. 이 점을 살펴보는 데는 교육에 관한 오우크쇼트(M. Oakeshott)의 견해가 도움을 제공한다. 그는 “우리 가운데 누구도 태어날 때부터 인간인 사람은 없으며, 우리는 누구나 학습을 통하여 인간이 된다”고 말하였다(Oakeshott, 2001, 6). 그가 보기에, 학교에서 학생에게 무엇인가를 가르친다면 그것은 삶을 살아가는데 유용한 정보를 전달하기 위한 것이 아니다. 가르치는 일을 정보 전달의 수단으로 이해하는 것은 인간이 거주하는 세계가 물리적인 사물로 이루어진 것이 아니라, 이해의 대상으로서의 의미로 이루어

어져 있다는 점을 간과한 불행한 결과이다. 한 아이가 태어나는 세상은 사물의 세계이기도 하지만, 엄연히 의미의 세계이기도 하다. 거기에는 인간이 인간으로서 살아가기 위해 반드시 요구되는 감정, 정조, 상상, 신념, 이해, 절차, 관습, 관례, 도덕적 규범과 원리 등등이 들어있다. 지식의 전달과 그것을 통한 학습은 새로 태어난 세대를 장차 그들이 거주하게 될 세계, 인간이라면 마땅히 거주해야 할 바로 그 의미의 세계에 입문시키기 위한 것이다. ‘인간다운 삶’이라는 것은 인류 공동의 업적, 한마디로 딜타이(W. Dilthey)가 말한 ‘정신세계’에 입문한 사람의 삶을 가리킨다고 말할 수 있으며, 그곳으로의 입문은 지식을 가르치고 배우는 일에 의하여 가능한 것이 된다. 오우크쇼트의 관점에서 보면, 지식을 가르치고 배우는 일은 인간다운 삶을 가능하게 하는 지극히 인간적인 상호작용이다. 이 점을 인정할 수 있다면, ‘인공지능은 교사를 대체할 수 있는가’ 하는 질문은 ‘지식의 전달이라는 인간 고유의 과업을 인공지능이 대신해도 좋은가’ 하는 질문으로 이해되어야 한다.

앞서 연구자는 ‘인공지능 시대의 도래’라는 시대적 상황은 우리에게 교육의 안쪽이 아니라 바깥쪽을 바라보도록 부추기고 있다고 말하였다. 지금까지의 고찰은 이 문제가 교육에 관한 오해와 조바심을 동시에 불러일으킨다는 점을 보여준다. 인공지능 시대가 도래했고, 이제는 인공지능이 ‘가르치는 일’을 대신할 수 있다고 생각하는 순간 우리는 교육의 안쪽에 해당하는 ‘가르치는 일’이 무엇인지를 깊이 생각해 볼 겨를이 없다. 심지어 교육의 바깥쪽에 일방적으로 주목할 때, 우리는 인공지능이 대체할 수 없는 ‘인간적 상호작용’이 무엇인가를 찾는 일이야말로 중요하고 시급한 문제라고 생각하게 된다. 그러나 ‘가르치지 않는 교사’라는 말은 넌센스이다. 실지로 가르치는 일, 즉 ‘수업’을 배제한 상태에서 교사에 관하여 논의한다는 것은 도저히 상상하기 어렵다.⁴⁾ 이러한 사정에 비추어 보면, ‘교육은 시대의 변화에 발맞추어야 한다’라는 명제가 우리를 엉뚱한 방향으로 이끌어 간다는 주장은 전혀 허황된 것이 아니다.

그렇다면 인공지능 시대의 도래는 교직에 아무런 영향을 끼치지 않는다고 보아야 하는가? 연구자가 보기에, 적어도 그 영향은 앞에서와는 다른 방식으로 이해되어야 한다. 즉, 인공지능 시대의 도래는 교직을 위협하는 것이 아니라, ‘교사의 본래적 역할은 무엇인가’라는 질문을 던지고 있다는 것이다(박대호, 2022, 101). 인공지능 시대의 도래는 그것 없이는 인간다운 삶이 불가능하면서도, 정작 시간을 들여 애써 생각해 보지 않았던 ‘가르치는 일’에 관하여 성찰을 강요한

4) 수업이 교사의 전문성의 핵심을 이룬다는 점에 대해서는 이홍우, 유한구, 장성모(2003)의 책 『교육과정이론』 제9장 ‘교사의 수업 전문성’(403-424)를 참고하기 바람. 이 글은 지금까지의 논의의 핵심을 ‘수업 전문성’이라는 주제를 중심으로 다루는 만큼, 여기에 관심이 있는 독자(특히, 교사)라면 읽어볼 만한 가치가 있다. 이 책에 적혀있는 다음의 구절은 본 연구에도 시사하는 바가 크다: “교사는 자신의 수업을 통하여 교과가 싸늘하게 죽어 있는 정보의 형태가 아닌 살아 움직이는 지식으로 되도록 하는 바로 그 일을 한다(408).” 여기에 언급된 ‘살아 움직이는 지식’은 앞서 언급한 ‘능력’과 다른 것이 아니다.

다는 생각이 들 정도이다. ‘인공지능이 교직을 위협한다’는 식의 발상이 가진 거의 유일한 이점이 바로 여기에 있다. 실지로 기계가 교사를 대체하지 않을 것이라는 주장을 통하여 브루너 자신이 강조하려고 했던 바는 기계의 등장으로 인하여 일부 교사가 대체되리라는 주장에 있지 않다. 그는 가르치는 기계의 등장이 우리로 하여금 교육다운 교육, 그의 용어로 ‘지식의 구조’를 전달하는 진정한 교육을 지향하고 추구하도록 한다는 점을 강조했던 것이다.

비에스타는 ‘디지털 교육’에 지나치게 주목하고 열광하는 모습에 우려를 표하면서, 우리가 반성 없이 사용하는 교육과 수업에 관한 전제와 개념을 비판적으로 분석하는 일이야말로 이 시대의 중요한 과제라고 말한 바 있다(Biesta, 2020, 6). 그가 보기에, ‘디지털 교육’이라는 말의 중요성은 그 말 자체에 놓여 있는 것이 아니다. 차라리 그것은 교육과 수업의 본래적 의미를 생각해 보게 하는 ‘계기’가 된다는 점, 바로 그 점에서 중요성을 갖는다. 그의 주장은 인공지능에도 그대로 적용될 수 있을 것이다. 즉, 인공지능과 같은 최신 과학기술은 교직을 위협하는 것이 아니라, 우리가 간과해 왔던 교직, 나아가 교육의 본질을 되묻고 있다는 것이다. 결국, 시대의 변화에 따른 교육의 변화는—새로운 교육으로 과거의 교육을 대체하는 것이 아니라—그 변화를 계기로 하여 교육의 본질에 조금씩 가까워지는 모습을 가리켜 부르는 것으로 이해될 수 있다. 이러한 이해의 방향을 도외시할 때 교육은, 마치 지식의 전달은 인공지능에게 맡겨두고 교사가 해야 할 그 이외의 일을 찾으려는 경향에서와 같이, 그릇된 방향으로 이끌릴 수밖에 없다. 이하에서는 이러한 아이디어를 가지고 인공지능윤리교육의 난점을 초등 도덕과를 중심으로 하여 본격적으로 살펴보겠다. 3장의 논의는 시대의 변화와 교육의 관계에 관한 특정한 이해가 도덕과 교육과정에도 깊숙이 영향을 미치고 있다는 점을 보여줄 것이다.

III. 인공지능윤리교육, 왜 어려운가?

이번 절의 주제는 시대의 변화와 교육의 관계에 관한 이상의 논의를 바탕으로 하여 인공지능윤리교육이 ‘왜 어려운가’를 탐색하는 것이다. 사실 교육과정이 개정되고, 거기에 맞추어 새로운 교과서가 등장할 때마다 교사들이 어려움을 겪는 것은 지극히 자연스러운 일이다. 교육과정을 뜻하는 커리큘럼(curriculum)이라는 영어 단어는 ‘달리다’라는 의미의 라틴어 단어 쿠레레(*currere*)에서 파생된 것이라고 알려져 있다. 교육과정의 개정은 어원상 ‘달리는 길’을 고치는 일이며, 이 점에서 이전과는 다른 길을 달려야 하는 대상은 얼마간 어려움을 겪을 수밖에 없다. 여러 번의 교육과정 개정을 경험한 교사들은 이 어려움이 ‘불가피한 것’이라는 점을 이미 인지하

고 있는 만큼, 그들이 그 불가피한 혼란과 어려움에 맞서 이의를 제기하는 모습을 상상하기는 쉽지 않다. 그러나 도덕과에 최초로 도입된 인공지능윤리교육에 관한 내용 요소는 예외라는 생각이 든다. 앞 절의 논의에 비추어 보면, 그것이 불려일으키는 난점은 ‘불가피한 것’이라기보다는 ‘불필요한 것’이 아닌가 하는 생각을 떨쳐버릴 수가 없다.

불가피한 혼란과 어려움이 아니라, ‘불필요한’ 혼란과 어려움이라는 말에 관해서는 약간의 설명이 보태질 필요가 있다. 도덕과를 담당하는 현장 교사들이 그 내용에 관하여 어떤 느낌을 받겠는가를 상상해 보는 일은 그러한 설명을 하기 위한 좋은 출발점이 된다. 이를테면, 도덕과에 전문성을 갖춘 한 교사가 있다고 가정해 보자. 즉, 그 교사는 자신의 경험과 노력을 통해 확립한 전문성 덕분에 교육과정이 개정되더라도 거기에 담긴 내용 요소만 보고도 어떻게 가르쳐야 하겠다는 일정한 판단을 내릴 수 있고, 그 판단에 따라 대략적인 수업의 흐름을 상상할 수 있다. 교육과정 개정에 의하여 도덕과의 특징 자체가 변경되는 것은 아니기 때문에, 그 교사에게 교육과정 개정에 따라붙는 어려움은 그다지 심각한 것이 아니며, 오히려 그 어려움은 보다 좋은 도덕과 수업으로의 발전을 위해서 반드시 겪어야 하는 어려움이라는 점에서 개인적으로 바랄 만한 것이기도 하다. 그러나 그러한 수준의 교사에게조차도 ‘인공지능윤리교육’에 해당하는 부분은 그가 지금까지 쌓아 온 도덕과에 관한 안목이 통하지 않을 가능성이 크다. 그 주제는 우리 교육과정에서 새롭게 시도된 것이기 때문에 도덕과에 전문성을 갖춘 교사조차도 이 내용 요소에 관해서는 속수무책일 수밖에 없다는 것이다. 다만, 연구자가 보기에 ‘속수무책’이라는 표현이 성립되는 진짜 이유는 다른 데 있다. 그 이유를 확인하기 위해서는 먼저 2022 개정 도덕과 교육과정의 내용 체계표를 살펴보는 것이 도움이 된다.

〈표 1〉 2022 개정 도덕과 교육과정 내용 체계 일부

범주	학년군
	5-6학년군
지식·이해	<ul style="list-style-type: none"> 타인을 왜 도와야 하며, 어떻게 도울 수 있을까? 서로의 다름을 존중해야 하는 이유는 무엇일까? 인공지능 로봇과 친구가 될 수 있을까?
과정·기능	<ul style="list-style-type: none"> 타인의 상황을 주의 깊게 관찰하고 다양한 도움 방안 탐색하기 편견 사례를 찾고 수정 방안 제안하기 인공지능 로봇과 관계 맺을 때 필요한 윤리적 원칙 점검하기
가치·태도	<ul style="list-style-type: none"> 타인을 위하는 자세 다양성을 존중하는 태도 인공지능 로봇과의 바른 관계 형성 의지 함양

제시된 표는 도덕과의 내용 체계표 가운데 5-6학년군의 ‘타인과의 관계’ 영역을 가져온 것이다. 일단 밑줄이 그어진 세 번째 항목을 제외하고 보면, 첫 번째 항목과 두 번째 항목은 도덕과에 관심을 가져온 사람들에게 이미 익숙한 내용이라고 말할 수 있다. 이 점을 염두에 두고, 앞서 언급한 교사에게 첫 번째 항목과 두 번째 항목의 가치·덕목이 무엇인가를 묻는다면 그는 무엇이라고 대답하겠는가? 어디까지나 예상에 지나지 않지만, 대체로 첫 번째 항목은 ‘배려’ 혹은 ‘봉사’, 두 번째 항목은 ‘다양성 존중’과 관련된 표현이 답변으로 제시될 것이라고 말한다고 해서 특별히 문제가 될 것 같지 않다. 실제로 교육과정에 그 부분의 성취기준은 각각 “[6도02-01] 봉사의 의미와 중요성을 이해하고, 타인이 처한 상황과 환경에 대한 주의 깊은 관심을 바탕으로 봉사를 실천한다”, “[6도02-02] 편견이 발생하는 이유를 탐색하여 해결 방안을 살펴보고, 다양성 존중을 바탕으로 다른 사람과 올바른 관계를 맺기 위한 실천 방안을 탐구한다”(교육부, 2022, 12)로 되어 있다. 여기에서 문제는 세 번째 항목이다. 앞의 두 항목에서와 마찬가지로 그에게 동일한 질문—세 번째 항목의 가치·덕목은 무엇인가—을 던진다면 그는 무엇이라고 대답하겠는가? 이 질문에 관하여 그가 모종의 대답을 내어놓을 가능성까지 배제할 필요는 없겠지만, 앞의 항목에서와는 달리 이 질문에 대답하기란 쉬운 일이 아니라는 점은 분명하다. 어쨌든, 이 항목의 성취기준은 “[6도02-03] 인간과 인공지능 로봇 간의 다양한 관계를 파악하고 도덕에 기반을 둔 관계 형성의 필요성을 탐구한다”(교육부, 2022, 12)로 되어 있다. 우리는 이 성취기준에 의하여 한 가지 중요한 사실을 확인하게 된다. 2022 개정 도덕과 교육과정의 내용 체계표에는 가치·덕목을 명확하게 말할 수 없는 내용 요소가 있다는 것이다. 요컨대, 기존 도덕과의 체제에 익숙한 교사가 이 지점에서 원하는 것은 인간과 인공지능 로봇 간의 도덕에 기반을 둔 관계를 형성해야 한다는 공허한 말이 아니라 그 관계를 부르는 이름, 즉 가치·덕목이지만, 위의 세 번째 성취기준에는 바로 그 핵심이 결여되어 있다. 이러한 사정은 그것을 가르쳐야 할 교사, 특히 도덕과에 전문성을 지닌 교사마저도 인공지능윤리교육의 방향을 설정하는 데 실질적인 어려움을 겪을 수밖에 없다는 점을 보여준다.

이상의 문제는 다른 각도에서 한 번 더 생각해 볼 필요가 있다. 여기에서 다른 각도라는 것은, 대답을 기대하기 어려워 보이는 질문을 반복적으로 던지는 일은 그만두고, 위에서 제기한 문제가 무엇인지를 분명히 해야 한다는 점을 뜻한다. 라일이 「마음의 개념」에서 언급한 ‘범주오류’(category mistake)는 그 문제를 포착하는 데 도움을 제공한다. 그는 범주오류의 사례로서 옥스퍼드 대학을 방문한 어떤 사람이 ‘나는 대학생, 교수, 사무직원은 보았지만, 이들이 머물고 활동하는 옥스퍼드 대학은 보지 못했다’고 항변하는 경우를 든다(Ryle, 1994, 20). 그가 이 사례를 통해 말하고자 했던 바는, 옥스퍼드 대학의 경우와 유사하게 ‘마음’을 그것이 발휘하는 여러

가지 능력과 별도로 존재하는 실체로 간주하는 것, 또는 마음을 그 능력과 동일한 형태로 존재한다고 간주하는 것도 오류라는 점이었다. 물론, 이 글의 맥락에서 보다 중요한 것은 마음에 관한 그의 견해가 아니라, 범주오류 그 자체이다.

옥스퍼드 대학의 예보다는 훨씬 실화에 가까운, 그렇기 때문에 훨씬 더 강력한 예로서 다음과 같은 경우를 들 수 있다. 한국의 어떤 어른이 한 아이에게 ‘주차장에서 출발 시간을 기다리는 버스에 사람이 많이 타고 있는가를 보고 오라’고 시켰더니 그 아이가 갔다 와서 ‘군인만 잔뜩 타고 있고 사람은 없더라’고 말했다는 것이다. ‘군인’과 대등한 범주는 ‘주부’나 ‘직장인’일지언정 ‘사람’은 아니다. 우리가 일상 사용하는 ‘국가’라는 용어로도 ‘범주오류’를 예시하는 것은 충분히 가능하다. ‘국가’를 보여 달라는 어떤 사람에게 국가 안에서 일어나는 일들이나 사람들의 행동을 보여 주면서 ‘이것이 국가이다’라고 말했을 때, 그 사람은 ‘아니, 이런 사건들, 이런 사람들의 행동 말고 “국가”를 보여 달라’고 항변할 수 있을 것이다(이홍우, 2024, 30).

이상의 범주오류에 관한 설명은 앞서 제기한 질문—인공지능윤리교육에 관한 세 번째 항목의 가치·덕목을 말하는 것이 왜 어려운가—에 한 가지 가능한 대답을 시사한다. 버스에 사람이 많이 타고 있는가를 보고 오라고 했더니 ‘군인만 잔뜩 타고 있고 사람은 없더라’고 말한 것과 마찬가지로 도덕과를 담당할 교사들이 기대했던 것은 ‘가치·덕목’이지만, 도덕과 교육과정에는 그것과 다른 범주의 어떤 것이 제시되어 있다는 것이다. 어떤 독자는 이러한 짐작에 맞서 다음과 같이 반론할지도 모른다. 내용 체계표 ‘지식·이해’ 범주에 들어 있는 세 가지 항목은 ‘타인을 왜 도와야 하며, 어떻게 도울 수 있을까?’, ‘서로의 다름을 존중해야 하는 이유는 무엇일까?’, ‘인공지능 로봇과 친구가 될 수 있을까?’이다. 여기에 언급된 항목들의 범주가 다르다는 것이 무슨 뜻인가? 세 항목 모두 ‘화두형’이라는 접근 방식을 공유하고 있는 만큼, 삼자는 모두 동일한 범주에 속한다고 보는 것이 타당한 것 아닌가? 실지로 교육과정 문서에는 ‘교육과정 설계의 개요’ 부분에서 이렇게 밝히고 있다.

지식·이해 범주는 학생의 도덕 발달 수준에 부합하는 도덕적 지식과 실천의 연계 과정을 촉진하는 데 주안점을 두었다. 그에 따라 도덕 수업의 과정은 정해진 답을 제시하기보다는 보다 바람직한 삶을 향한 각자의 답을 찾아가는 과정에 초점을 두어야 한다는 점을 감안하여 질문형으로 제시하고자 한다. 이런 진술 방식의 선택은 이미 우리 전통 속에서 ‘화두(話頭)’라는 개념으로 정착하여 일상화된 방식이기도 하고, 도덕과가 포함하고 있는 철학교육과 메타적 차원의 종교교육을 포용하기 위한 선택이기도 하다(교육부, 2022, 4).

세 항목은 겉으로 보기에 모두 인용문의 설명을 따르고 있으며, 그런 만큼 동일한 범주라고 생각하는 것은 무리가 아니다. 그러나 문법적 형식이 동일하다는 사실을 근거로 하여 해당 항목들이 동일한 범주라고 생각하는 것은, 비트겐슈타인(L. Wittgenstein)의 표현을 빌리자면, ‘언어의 홀림’(bewitchment of language)에 걸려든 것이다. 애초에 라일이 말한 범주오류란 이러한 ‘언어의 홀림’의 한 양상으로 이해될 수 있다. ‘친구와 왜 사이 좋게 지내야 하는가’와 ‘인공지능 로봇과 친구가 될 수 있을까’는 동일한 문법적 형식을 사용하고 있으며 심지어 각각 ‘우정’이라는 동일한 가치·덕목을 떠올리게 하지만, 전자는 우정의 의미가 정해져 있지 않은 상태에서 그 의미를 탐색하도록 하는 질문이며, 후자는 이미 사람들 사이에 공유되고 있는 우정의 의미를 인공지능 로봇에 적용해도 좋은가를 생각해 보게 하는 질문이라는 점에서 양자는 서로 구분된다.⁵⁾ 요컨대, 전자가 ‘의미의 탐색’이라는 성격을 지닌다면, 후자는 ‘개념의 적용’이라는 성격을 지닌다. 양자를 면도칼로 가르듯 명확하게 나눌 수는 없다고 말하는 것이 정확하겠지만, 적어도 도덕과 수업이 전자에 가깝다는 점은 부정하기 어렵다. 애초에 후자를 중심으로 하는 것은 이미 정해진 개념이나 규범을 단순히 반복하는 것이요, 그리하여 ‘수업’보다는 ‘훈련’에 가깝다고 보아야 한다. 그렇다고 하여 도덕과 수업에서 후자에 해당하는 활동이 일어나지 않는 것은 아니다. 거기에서도 ‘개념의 적용’에 해당하는 활동은 이루어질 수 있고, 이루어져야 한다. 다만, 그 활동은—그것이 도덕과의 활동이 되기 위해서는—어디까지나 ‘의미의 탐색’의 안내를 받아 이루어져야 한다. 우정에 관한 의미의 탐색 없이, 학생이 이미 알고 있는 우정이라는 개념을 적용하고 활용하는 것은 애초에 수업이라는 개념과는 거리가 멀다. 거기에는 우정에 관하여 새롭게 배운 ‘의미’가 없기 때문이다. 이 구분에 의하면, ‘의미의 탐색’과 ‘개념의 적용’은 각각 ‘지식의 습득’과 ‘지식의 적용’이라는 말로 바꿔 불러도 무방하다.

2022 개정 도덕과 교육과정의 내용 체계표를 통해서도 확인할 수 있는 바와 같이, 거기에 들어 있는 질문은 대체로 도덕적 규범에 관한 질문이요, 그것도 특정한 가치·덕목으로 ‘초점화된 질문’이다.⁶⁾ 그러나 앞에서 살펴본 ‘인공지능윤리교육’ 관련 내용 요소는 그렇지 않다. 이러한

5) 여기에서, 후자 또한 ‘결국에는’ 우정의 의미에 관한 질문으로 연결된다는 주장이 제기될 수도 있을 것이며, 이 주장은 분명 타당한 면이 있다. 그러나 여기에서 문제삼는 것은 도덕과의 ‘내용 요소’로 적합한가 하는 문제이다. 도덕과가 지향하고 추구하는 핵심적인 가치·덕목을 담고 있어야 할 내용 체계표에 일종의 ‘길목’에 해당하는 질문을 담는 것이 타당한지에 관해서는 별도로 논의할 문제라고 판단된다. 여기에 더하여, ‘인공지능과 친구가 될 수 있을까’라는 내용 요소는 여러 가치·덕목의 ‘길목’이 될 수 있으며, 따라서 ‘우정’ 이외의 다른 가치·덕목과도 연결될 수 있다는 또 다른 문제를 불러일으킨다.

6) 이 점에 관해서는 2015 개정 도덕과 교육과정의 내용 체계표가 보다 더 분명한 형태로 보여준다고 말할 수 있다. 거기에는 2022 개정 도덕과 교육과정에서와는 달리 각각의 내용 요소가 지향하는 가치·덕목이 명시적으로 기술되어 있다. 예컨대, 5-6학년군 타인과의 관계 영역에 내용 요소는 ‘우리는 남을 왜 도와야 할까? (봉사)’와 같은 형태로 기술되어 있다(교육부, 2015, 6).

사정은 새로 도입된 내용 요소의 범주가 기존의 것과는 성격상 상이하다는 점을 보여준다. 문제는 여기에 그치지 않는다. 예컨대, 봉사, 다양성 존중에 관한 탐구는 도덕과와 자연스럽게 연결되는 데 비하여, 인공지능 로봇과 친구가 될 수 있는가에 관한 탐구는 그에 관한 탐구의 과정에서 애써 관련 가치·덕목을 거론하고 연결하지 않는 이상 도덕과와는 이질적인 것으로 보일 수 있다. 표면상 전자가 ‘가치·덕목 중심’의 내용인 데 비하여, 후자는 ‘사례 중심’이기 때문이다. 이런 이유로 후자의 경우 그것과 짝을 이루는 특정 가치·덕목을 쉽게 상상하기 어렵다는 문제, 거기에 우리가 도덕과를 통해 가르치고 배워왔던 거의 모든 가치·덕목이 그 후보에 들어갈 수 있다는 문제가 따라 나온다. 애써 하나의 가치·덕목과 연결하지 않는 이상 그것은 성실, 정의, 배려, 책임, 존중, 정직 등 도덕과에서 다루는 일체의 가치·덕목과 관련짓는 것이 얼마든지 가능하다는 것이다.⁷⁾ 이 가능성은 인공지능윤리교육이라는 현대적 관심사가 ‘인공지능 로봇과 친구가 될 수 있는가’라는 내용 요소로 정리되어 도덕과교육에 도입될 수 있었던 이유가 ‘관련성’이라는 말로 요약될 수 있다는 점을 보여준다. 그러나 지금까지의 논의가 보여주듯이, 관련성을 갖고 있다는 이유로 도덕과교육의 내용이 될 수 있다고 주장하는 것은 무리가 있다. 관련성은 교육내용 선정의 필요조건일 수는 있지만 ‘충분조건’은 아니기 때문이다. 결국, 초등 도덕과를 통한 인공지능윤리교육의 난점은 새로운 내용의 도입이라는 말로 온전히 설명될 수 없다고 보아야 한다. 기존의 것과는 상이한 범주의 내용 요소를 도입했다는 것, 달리 말해 도덕과에 그것의 본래 목소리와는 이질적인 어떤 것이 침범했다는 것, 바로 그것이 근본적인 이유라고 말해야 한다는 것이다.

‘인공지능 로봇과 친구가 될 수 있을까?’라는 내용 요소가 급변하는 시대적 상황에 부응하려는 시도라는 점에 대해서는 그 나름의 평가를 받아야 한다. 그러나 그러한 평가와는 별개로, 학자들에게 일정한 비판의 대상이 되어 왔다는 점, 보다 구체적으로 그 비판의 근거에 대해서도 귀를 기울일 필요가 있다. 그 비판 가운데 인공지능에 관한 지나친 ‘의인화’에 대한 우려는 주목할 만하다(김하민, 2023; 이정렬, 2024; 권재은, 양해성, 2024). 따지고 보면, ‘인공지능 로봇과 친구가 될 수 있을까’라는 고민은 인공지능 기술을 사람처럼 표현하는 과정에서 발생한 혼란의 산물이라는 것이다. “인간은 인간이고 인공지능 로봇은 로봇”(이정렬, 2024, 95)이라는 주장은 인공지능 로봇과 친구가 될 수 있는가의 여부가 아니라 ‘의인화 현상’ 자체가 문제일 수도 있다

7) 최근 한 연구는 도덕과 공통교육과정에 반영된 인공지능윤리교육이 ‘인간과 인공지능 로봇과의 관계’로 수업되고 있다는 점(그 연구의 표현으로, ‘영역 제한형 모델’)을 지적하고, 그 대안으로서 ‘영역 분산형 모델’을 제안한 바 있다(윤준식, 2025). 이 제안에 들어 있는 인공지능윤리교육이 모든 영역에 분산될 수 있다는 발상은 이미 인공지능윤리교육이라고 특정할 수 있는 내용이 도덕과에서 다루는 일체의 가치·덕목과 관련지을 수 있다는 점을 보여준다고 말해도 좋을 것이다.

는 점을 시사한다. 즉, 의인화라는 문제를 걷어낼 때, ‘그것과 친구가 되어야 하는가’라는 질문은 무의미한 것이 되고 만다. 이러한 시사는 ‘인간-인간’의 관계가 ‘인간-인공지능’의 관계와 엄밀하게 구분될 수 있고, 또 구분되어야 한다는 점을 보여준다. 실지로 임상수는 인공지능 기술을 사용자와 대비되는 등가적 존재이자 행위자로 간주하는 것은 오류라고 지적한다. 그에 의하면, 사용자에 대비되는 등가적 존재는 인공지능 기술이 아니라, 인공지능 기술을 개발하고 운용하는 IT기업과 정부 당국이다(임상수, 2023, 61). 인공지능에 관한 ‘의인화 현상’에 관한 지적이나, ‘인간-인공지능’의 관계가 아니라 ‘인간-인간’의 관계에 주목해야 한다는 주장은 한 가지 동일한 방향을 가리키고 있다. 인공지능에 의해 제기되는 다양한 문제들은 사실 인공지능 그 자체의 문제라기보다는 그것을 개발하고 사용하는 인간의 문제라는 것이다. 한마디로, 인공지능윤리는 결국 ‘인간의 윤리’이다(심지원, 이은재, 김문정, 2022).

이상의 분석은 인공지능윤리교육에 제기되는 비판들이 ‘범주오류’라는 동일한 문제로 요약될 수 있다는 점을 보여준다. 지나친 의인화에 관한 우려나 ‘인간-인공지능’의 관계가 아니라 ‘인간-인간’의 관계에 주목해야 한다는 주장은 다 같이 인간에게 정상적으로 적용될 수 있는 개념을 인공지능에 부당하게 적용하는 데서 비롯된 문제이기 때문이다. 그러나 아직 한 가지 문제가 더 남아 있다. 인공지능윤리교육 분야의 혼란을 초래하는 범주오류는 어디에서 온 것인가? 지금까지의 논의에 비추어 보면, 그것은 ‘변화의 시대’라거나 ‘시대의 요구’라는 목소리가 자아내는 ‘조바심’이라고 말할 수 있다. 그 조바심은 오늘날의 현격한 변화에 대응하기 위하여 시급히 모종의 조치를 강구해야 한다는 생각, 교사가 새 시대의 새 역할을 다해야 하며, 교육에는 기존의 것과는 다른 새로운 내용이 도입되어야 한다는 생각에 잘 나타나 있다. 결국, 교육 바깥의 요란한 소음에 모든 이목을 집중시키는 그러한 조바심은 정작 교육 그 자체의 의미와 가치를 간과하도록 하며, 인공지능윤리교육의 난점 또한 여기에서 비롯된다고 말해도 좋을 것이다.

물론, 사정이 이상과 같다고 하여 이 글이 인공지능윤리교육이 불가능하다거나 불필요하다거나 하는 주장으로 연결되는 것은 아니다. 시대의 요구라는 목소리가 불려일으키는 조바심이 그 자체로 심각한 문제라는 주장과 교육이 시대적 요구를 수용하고 변화해야 한다는 주장은 얼마든지 양립할 수 있기 때문이다. 그러나 후자의 주장은 교육이 시대의 변화에 끌려간다는 식의 일방적인 관계로 이해될 여지가 있고, 이러한 이해 방식은 모든 문제를 엉뚱한 방향으로 이끌고 간다는 점에 대해서는 앞으로도 깊은 고민이 필요하다. 그도 그럴 것이, 왜 시대의 변화와 교육의 관계를 시대의 변화가 먼저 있고, 교육이 그 뒤에 따라와야 한다는 식으로만 생각해야 하는가? 이러한 식의 생각이 널리 호응을 얻고 있기는 하지만, 반대로 교육이 시대의 변화를 주도하고 이끈다는 생각, 또는 교육은 시대의 변화를 넘어선다는 생각은 그릇된 것인가?⁸⁾ 지금까지 살펴

보았듯이 전자에 해당하는 발상이 도덕과의 인공지능윤리교육에 어려움을 불러일으키는 근본적인 문제라면, 이제 우리는 후자에 해당하는 발상에 조금 더 귀를 기울일 필요가 있다.

결과적으로, 도덕과가 이상의 문제에서 벗어나기 위해서는 앞서 언급한 조바심에서 벗어나 인공지능윤리교육이 다루려는 도덕적 지식은 기존의 윤리학, 나아가 도덕과교육에서 가르치고 배워왔던 것과 상이한 도덕적 지식이 아니라는 점을 분명히 해야 한다. 실지로 윤리학은 장구한 변화의 흐름 위에서도 그 변화에만 매몰되지 않고, 오히려 그 변화가 나타나는 현실의 이리저리한 도덕적 측면을 파악하는 데 요구되는 개념들을 정련하고 체계화해 왔다. 다시 말하여, 우리는 윤리학적 개념을 배우지 않고는 현실의 도덕적 측면을 파악하는 것, 즉 모종의 사태를 도덕적 관점으로 바라보는 것이 불가능하다(이홍우, 1982, 125). 도덕과가 윤리학적 개념, 즉 가치·덕목을 내용으로 삼는 것은 그것의 전달을 통하여 학생들이 사태를 도덕적 관점으로 바라보도록 하기 위해서이다. 이 점에서 보면, 시대의 변화가 도덕과에서 다루어야 할 새로운 내용을 만들어 내는 것이라고 보는 것은 지나친 과장이다. 차라리 시대의 변화는 교육을 통하여 습득한 도덕적 지식을 동원하여 파악되어야 할 사태이며, 다시 그 사태는 기존의 도덕적 지식을 그 본래의 의미에 가깝게 정련하도록 하는 외적 계기로 이해되어야 한다. 결국, 시대의 변화와 교육의 관계를 올바르게 바라보도록 노력하는 일은 앞으로 도덕과를 통해 이루어질 인공지능윤리교육의 중요한 과제라고 말할 수 있을 것이다.

IV. 결론

비에스타는 ‘디지털이 우선인가, 교육이 우선인가?’라는 질문을 다루면서 오늘날이 전례를 찾아보기 힘든 시기라는 주장은 한편으로는 타당하지만, 다른 한편으로 보면 꼭 그렇지 않다고 말한다. 교사는 인간의 변화하는 삶, 변화하는 시대를 배경으로 하여 가르치는 일을 계속해 왔으며, 그것은 새삼스러운 일이 아니기 때문이다. 실지로 교사에 의하여 이루어지는 수업의 본질은 익숙한 것과 낯선 것, 전형적인 것과 전례 없는 것이 뒤섞인 상황에서 의미 있는 교육을 만들어 나가는 데 있다(Biesta, 2020, 6). 비에스타의 견해에 애써 주목하지 않더라도, 교사는 오로지 학생을 현재에 안주하도록 하는 일을 하는 것이 아니며, 교육은 처음부터 시대에 맞는 인간을 기르기 위하여 고안된 것이 아니다. 교육이라는 영위 전체를 두고 볼 때 시대에 꼭 맞는 인간은

8) 적어도 이러한 생각이 완전히 새로운 것은 아니라고 보아야 한다. ‘변화하는 시대와 자유교육의 이념’이라는 제목의 박병철(2024)의 연구에 의하면, 변화하는 시대는 전통적인 의미에서의 교육을 배척하기보다는 오히려 요청한다.

어디까지나 불완전하고 왜곡된 인간이라고 보아야 한다. 교육이 지향하고 추구하는 이상적인 인간은 시대의 변화를 넘어선다. 교육의 영역에서 교육의 목적을 다룰 때 언제나 등장하는 ‘자율성’이라는 개념은 교육이 시대의 변화를 넘어서는 인간의 형성을 겨냥하고 있다는 점을 여실하게 보여주고 있다.

지금까지 이 글은 시대의 변화에 관한 지나친 관심이 교사와 도덕과교육의 본래적 역할을 도외시하게 만드는 문제를 불러일으킨다는 점, 나아가 그 문제가 인공지능윤리교육의 근본적인 어려움을 불러일으킨다는 점을 논의하였다. 지금까지의 논의에 의하면, 새 시대에 맞추어 교사와 도덕과가 각각 새로운 역할과 내용을 찾아나서야 한다는 생각은 다 같이 시대의 변화와 교육의 관계를 ‘시대의 변화→교육’이라는 식으로 파악한 결과라고 말할 수 있다. 물론 그러한 사고방식이 모조리 그릇된 의견을 개진하고 있다고 말하는 것은 문제를 지나치게 단순화하는 것이다. 건축 기술의 발달이 수렵 사회를 정주형태의 사회로 변화시켰다는 인류학적 사실처럼, 시대의 변화가 삶에 미치는 영향을 부정할 수는 없다. 다만 일종의 ‘기술결정론’과 같은 사고방식이 간과하는 것은 최신 기술 또한 교육의 산물, 그것도 그 산물의 일부라는 데 있다. 기술은 세계를 이해하려는 인간의 구체적인 노력의 산물이며, 그 노력의 밑바닥에는 언제나 교육이 놓여 있다. 여기에 비추어 보면, 기술은 교육을 전제조건으로 한다고 보아야 하며, 기술이 교육에 선행하는 것은 아니다. 시대의 흐름에 부응하여 교사에게 새로운 역할이 부여되어야 한다는 생각이나, 도덕과에 새로운 내용이—그것도 도덕과와 이질적인 형태로—들어와야 한다는 생각은 바로 이 점을 도외시한 결과라고 말해야 한다.

여기에서 간단하게나마 ‘사소한 것에 매달리는 것은 교육을 타락시킨다’라는 헤르바르트的主張과 ‘교육이 우선’이라는 비에스타의 주장⁹⁾이 어떤 의미를 가지는가를 지금까지의 논의의 맥락에서 생각해 볼 필요가 있다. 적어도 그들이 보기에, 교육의 안쪽에 의한 규제를 벗어나 시대의 변화라는 명목 아래 교육 바깥의 관심사를 다루는 것은 ‘교육의 일탈’이라고 이름 붙일 수 있으며, 이 일탈은 교육을 진정한 의미에서의 교육과 아무런 관련을 맺지 못한 상태로 만들 가능성이 있다. 달리 말해, 최신의 것, 현대적인 것을 다루어야 한다는 주장에 입각하여 단기적이고 즉각적인 것에 무조건적 관심을 국한시키는 일은 교육의 정당한 관심사를 멀리 밀어낸다. 물론, 이러한 일이 한순간에 일어나는 것이라고 말할 수는 없다. 교육의 일탈이 가진 위험성은 그것이 교육 전체를 눈 깜짝할 사이에 본래의 모습을 형해화하는 데 있는 것이 아니다. 그 위험성은 교육 바깥의 관심사가 발산하는 매력을 동력으로 하여 계속해서 교육을 밑바닥에서부터

9) “교육이 우선이라는 그의 주장은 교육의 실천 과정에서 ‘방법’(how)을 결정하려고 할 때 반드시 ‘목적’(what for)을 기준으로 삼아야 한다는 말을 뜻한다”(Biesta, 2020, 5).

잠식하는 데 있다고 보아야 한다. 그리고 그 위험성은 인공지능윤리교육을 다루는 도덕과교육에도 그대로 적용될 수 있다. 도덕과교육의 본래적 성격에 의해 규제되지 않는 최신의 것과 현대적인 것이 도덕과교육을 뒤흔들어 놓을 가능성은 언제나 열려있다.

결과적으로 볼 때, 장차 초등학교 교실에서 도덕과를 통해 이루어질 인공지능윤리교육은 상당한 수준의 혼란과 어려움을 겪을 수밖에 없을 것으로 보인다. 그러나 시대의 변화가 우리에게 생각해 볼 문제를 제기하며, 또 사태의 본질에 더 가까이 다가갈 수 있게 하는 계기가 된다는 이 글의 주장은 도덕과에 제기된 문제 자체에도 적용되어야 한다. 이 글의 주장에 의하면, 도덕과를 통해 이루어질 인공지능윤리교육의 난점은 도덕과의 의미와 가치에 관한 성찰을 강요한다. 그리고 기술이 강요한 성찰이 우리를 도덕과교육의 본질로 밀어붙일 때, 그 성찰의 과정에서 우리는 인공지능윤리교육에 제기되었던 난점이 그다지 심각한 문제가 아니었다는 점을 깨닫게 될지도 모른다. 또한, 그 과정에서 우리에게 닥친 진짜 문제는 인공지능윤리와 같은 최신의 내용이기보다는 도덕과교육, 나아가 ‘교육’ 그 자체에 관한 이해와 관심의 결여라는 점이 밝혀질지도 모른다. 역설적으로 보일지 모르겠지만, 인공지능윤리교육의 성공적인 안착을 위해서는 시대의 요구에 귀 기울이는 것 이상으로 교육, 교사, 도덕과교육, 수업과 같은 기초에 해당하는 용어들의 의미와 가치에 보다 깊은 관심을 기울일 필요가 있다.

※ 논문 투고일: 2025. 09. 26. ※ 논문 수정일: 2025. 11. 17. ※ 게재 확정일 : 2025. 12. 10.

〈참고문헌〉

- 교육부(2015). **도덕과 교육과정**. 세종: 교육부.
- 교육부(2022). **도덕과 교육과정**. 세종: 교육부.
- 권누리(2024). 초등 도덕과에서의 인공지능윤리교육 방안 연구-인공적 도덕 행위자(AMA) 논의를 중심으로-. **초등도덕교육**, **87**, 87-108.
- 권재은, 양혜성(2024). 인공지능 윤리교육 방안 연구-인공지능에 대한 의인화를 중심으로-. **윤리교육연구**, **74**, 87-118.
- 김하민(2021). 초등 도덕과 미래형 교육과정 개정 방향 탐색: 인공지능 윤리 교육을 중심으로. **초등도덕교육**, **73**, 1-27.
- 김하민(2023). 2022 개정 초등 도덕 교육과정에 대한 비판적 고찰-인간과 인공지능 로봇 관계 내용 서술을 중심으로-. **도덕윤리과교육**, **79**, 29-56.
- 박대호(2022). 도덕과를 통한 인공지능윤리교육의 방향. **도덕교육연구**, **34**(3), 83-104.
- 박병철(2024). 변화하는 시대와 자유교육의 이념. **교양교육연구**, **18**(5), 113-123.
- 박보람(2025). 2022 도덕과 교육과정에서의 ‘인공지능 윤리’ 교육 -초·중·고 계열 연계성과 윤리적 기반 탐구-. **윤리연구**, **148**, 37-61.
- 박형빈(2024). 규범적 AI윤리 가이드라인과 초등 AI윤리교육 적용 방안 모색. **한국초등교육**, **35**(2), 79-95.
- 변순용(2020). AI 윤리 교육의 필요성에 대한 연구. **한국초등교육**, **31**(3), 153-164.
- 신현화, 신태수, 신현우(2022). 초등학교 도덕과 수업에서 인공지능 윤리교육 방안 탐색 연구. **학교와 수업 연구**, **7**(1), 75-96.
- 심지원, 이은재, 김문정(2022). 인간의 윤리로서 인공지능윤리-인공지능윤리의 가치와 자리-. **철학·사상·문화**, **38**, 46-64.
- 윤준식(2025). 현대 기술윤리와 인공지능윤리교육: 도덕과 공통 교육과정 인공지능윤리교육의 영역 분산형 모델 제안. **2025 한국도덕윤리과교육학회 연차학술대회 자료집**, 373-396.
- 이정렬(2024). 도덕과 교육과정 내용 요소 “인공지능 로봇과 친구가 될 수 있을까”에 대한 비판적 탐색. **초등도덕교육**, **88**, 79-105.
- 이홍우(1982). 도덕교육의 내용으로서의 윤리학. **도덕교육연구**, **1**, 103-128.
- 이홍우, 유한구, 장성모(2003). **교육과정이론**. 파주: 교육과학사.
- 이홍우(2024). **메타프락시스**. 파주: 교육과학사.
- 임상수(2023). 소비자 윤리 관점에서 본 2022 개정 도덕과 교육과정과 인공지능 윤리. **초등도덕교육**, **85**, 51-74.
- 정창우, 이혜진(2022). 도덕과에서 AI윤리교육의 필요성과 과제. *The SNU Journal of Education Research*, **31**(1), 55-82.
- 홍현주(2021). 인공지능 윤리교육의 초등도덕 교육과정 적용 방안. **초등도덕교육**, **75**, 183-206.
- Biesta, G.(2020). Digital first or education first?: Why we shouldn't let a virus undermine our educational artistry. *PESA Agora*. <https://pesaagora.com/columns/digital-first-or-education-first-why-we-shouldnt-let-a-virus-undermine-our-educational-artistry/>
- Bruner, J. S.(2006). **교육의 과정**. [The Process of Education] 이홍우 역. 파주: 교육과학사. (원저출판년도 1960년)
- Herbart, J. F.(1982). *The Science of Education: Its General Principles Deduced from its Aim*.

- M. H. Felkin & E. Felkin(Trans.). London: Swan Sonnenschein & Co. (원저출판년도 1806년)
- Oakeshott, M.(2001). A Place of Learning. T. Fuller(Ed.). *The Voice of Liberal Learning: Michael Oakeshott on Education*, New Haven & London: Yale Univ. Press.
- OECD.(2018). Teachers as Designers of Learning Environments: The Importance of Innovative Pedagogies. https://www.oecd.org/en/publications/teachers-as-designers-of-learning-environments_9789264085374-en.html
- Ryle, G.(1994). **마음의 개념**. [The Concept of Mind] 이한우 역. 서울: 문예출판사. (원저출판년도 1949년)
- Selwyn, N.(2022). **로봇은 교사를 대체할 것인가?** [Should Robots Replace Teachers?] 정바울, 박다빈, 박민혜, 정소영 역. 서울: 에듀니티. (원저출판년도 2019년)
- Whitehead, A. N.(1967). *The Aims of Education and Other Essays*. Toronto: Collier-Macmillan Canada Ltd.

〈Abstract〉

Challenges and Tasks in AI Ethics Education

Park, Daeho¹

This paper begins with the critical awareness that the difficulties surrounding AI ethics education (AIEE) stem from a belief that “education must follow the changes of the times.” Such a belief draws attention to external changes, but it neglects the very meaning and value of education. This, in turn, risks leading AIEE as a whole in a misguided direction. The theme of “the teaching profession in the age of artificial intelligence” offers a concrete example of that risk. The advent of the AI era has prompted educators to seek new roles, not in terms of mere “knowledge transmission,” but in terms of “human interaction.” However, this understanding can only be sustained on the flawed premise that knowledge transmission is equivalent to the delivery of information. The notion that education must adapt to the changes of the times also influences the structure and content of AIEE within the moral education curriculum. The newly introduced content elements related to AI ethics in the elementary moral education curriculum differ in nature from the existing framework, which has been focused on specific values and virtues. This divergence is likely to cause confusion and challenges in actual school settings. Such circumstances compel us to ask a fundamental question: “AI first, or education first?” For the successful implementation of AIEE, it is essential to give due attention to the meaning and value of the very concepts of education, teachers, Moral Education, and instruction.

Keywords : AI ethics education, Moral Education, education in changing times, category mistake, role of the teacher

1. Assistant Professor, Cheongju National University of Education, dhpark@cje.ac.kr