

인공지능 윤리에 대한 대학원생의 교육 요구 분석*

Educational Needs of Graduate Students Regarding AI Ethics

1차원고접수: 2024.01.24. 수정원고접수: 2024.02.27. 최종 게재결정: 2024.03.18.

정수정¹ · 백새봄² · 임다미³¹서울대학교 경력개발센터, BK조교수²서울대학교 경력개발센터, 연구원³국립공주대학교 교육학과, 부교수Jeong, Soojeong¹Baek, Saebome²Im, Tami³^{1,2}SNU Career Development Center, Seoul National University³Dept. of Education, Kongju National University

국문초록 본 연구의 목적은 인공지능 윤리에 대한 대학원생의 교육 요구를 분석하고, 이러한 교육 요구가 인공지능 활용 빈도에 따라 차이가 있는지 살펴보는 것이다. 연구참여자는 서울 소재의 대규모 국립 종합대학교에 재학 중인 593명의 대학원생이었다. 자기보고식 설문조사를 통해 인공지능 윤리기준 관련 22문항에 대한 학생들의 현재수준과 중요수준을 측정하였으며, 수집된 데이터는 Borich 요구도 분석과 The Locus for Focus 모델을 활용하여 분석하였다. 연구 결과, 대학원생을 위한 인공지능 윤리 교육에서 가장 우선으로 다루어야 할 내용은 안전성, 책임성, 투명성, 인권보장, 침해금지로 확인되었다. 또한, 책임성은 인공지능 활용 빈도와 관계없이 일관되게 교육 요구 최우선순위로 도출되었다. 반면, 인공지능 활용 빈도가 낮은 집단일수록 프라이버시보호와 투명성이, 인공지능 활용 빈도가 높은 집단일수록 안정성에 대한 교육 요구가 높은 것으로 나타났다. 이러한 연구 결과는 대학원생 대상 인공지능 윤리 교육 프로그램 개발을 위한 기초 자료로 활용되어 학생들의 인공지능 윤리의식과 역량 함양에 이바지할 것으로 기대한다.

주요어 인공지능 윤리, 교육 요구, 인공지능 활용 빈도, 대학원생

Abstract The purpose of this study is to analyze the educational needs of graduate students regarding artificial intelligence (AI) ethics and to examine whether these educational needs differ according to the frequency of AI use. Participants were 593 graduate students enrolled in a large university in Seoul, South Korea. We used a self-reported survey to measure students' levels of current understanding and their perceived importance of 22 items related to AI ethical standards. We analyzed the data using the Borich needs assessment model as well as The Locus for Focus Model. Our results identified safety, accountability, transparency, human rights protection, and non-infringement as the top priorities for AI ethics education for graduate students. Accountability emerged consistently and was identified as the top priority for educational needs, regardless of the frequency of AI use. On the other hand, privacy and transparency were found to be more important for those who use AI less frequently, and safety was found to be more important for those who use AI more frequently. Our findings will provide valuable insights for designing and developing AI ethics educational programs for graduate students, ultimately helping them in enhancing their ethical awareness and competence in relation to AI.

Keywords AI ethics, educational needs, frequency of AI use, graduate students

I. 서론

1. 연구의 필요성 및 목적

인공지능 기술의 비약적 발전은 일상생활과 사회 전반에 걸쳐 막대한 변화를 추동하고 있다(이재호, 정소윤, 강정석, 2019). 이는 교육 영역에서도 예외가 아니다. 유아부터 고등교육에 이르기까지 인공지능이 광범

위하게 도입 및 활용되면서 교육과정과 방법의 패러다임이 크게 전환되고 있다(김수정, 서혜민, 이보람, 2023; Adams, Pente, Lemermeier, & Rockwell, 2023). 인공지능 시대의 인재 양성을 위한 융합 교육과정이 대대적으로 도입되고 있으며(이은경, 2020), 교수학습 현장에 인공지능을 활용함으로써 학습자 맞춤형 교육의 실현 가능성이 한 층 더 높아지고 있다(Zawacki-Richter,

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A3A2A01090926)

Corresponding to 임다미 (Im, Tami)

국립공주대학교 교육학과, 부교수 (Dept. of Education, Kongju National University) E-mail. tamiim@kongju.ac.kr

Copyright © 2024 The Korean Association of Human Development. All rights reserved.

Marin, Bond, & Gouverneur, 2019). 또한, ChatGPT와 같은 생성형 인공지능의 등장 및 발달과 함께 초등학생부터 대학원생까지 개인 학습 도구로 인공지능을 활용하는 것이 일반화되는 추세이다(송주영, 전용주, 2023).

하지만 현재 대학에서의 인공지능 활용에 관한 연구는 초기 단계로, 학부 수업에서 인공지능 번역 플랫폼, 생성형 인공지능 등을 활용한 사례 연구(오선경, 2023; 정남숙, 2021)가 주를 이루고 있고, 대학원생을 대상으로 한 연구는 인공지능 융합교육 교육과정 개발 사례 연구(김지윤, 김승현, 김귀훈, 2022)와 대학원생의 AI 활용에 대한 인식 분석 연구(정수정, 이명성, 김유겸, 임다미, 2023) 정도로 부족한 실정이다.

한편, 대학은 교육뿐만 아니라 연구 기능도 아울러 갖고 있어 인공지능 도입 및 활용 범위가 더 넓다. 최근 들어 인공지능 관련 연구가 분야를 막론하고 급증하고 있으며(황서이, 김문기, 2019), 연구 수행의 과정에도 인공지능 기술과 도구가 활발히 활용됨으로써 연구자들의 연구 효율성 및 생산성 향상에 일조하고 있다(Chubb, Cowling, & Reed, 2022). 인공지능 활용을 통한 효율성, 생산성 향상과 같은 장점을 유지하면서 인공지능을 올바르게 활용하기 위해서는 인공지능 활용으로 인해 발생할 수 있는 표절을 포함한 학문적 책무성, 프라이버시 침해 등과 같은 윤리적인 이슈들에 대한 인식과 이를 고려한 실천이 반드시 병행되어야 한다. 이런 점에서, 인공지능은 모든 학습자에게 중요하겠지만 특히 학습자이면서 연구자 또한 예비 교수자로서의 정체성을 가진 대학원생에게 인공지능을 제대로 이해하고 올바르게 활용하는 역량을 함양시키는 것은 당연하면서도 꼭 필요한 일이다(정수정 등, 2023).

인공지능의 도입과 활용은 개인 맞춤형 지원, 효율성 및 생산성 향상 등 긍정적인 측면이 많지만, 표절 및 저작권 침해, 책임 불분명, 차별과 불공정, 사생활 침해 등과 관련된 부정적 윤리적 이슈 및 사회적 문제를 유발하기도 한다(Bouhouita-Guermech, Gogognon, & Belisle-Pipon, 2023). 인공지능 챗봇 ‘이루다’의 사회적 약자에 대한 혐오 발언과 개인정보 침해, 미국 공립학교 교사 고용 과정에서 활용된 비밀 알고리즘 평가 등

이 대표적인 예이다(교육부, 2022). 최근에는 위조와 변조, 표절, 부당한 저자 표기 등 생성형 인공지능을 활용하여 학술논문을 작성할 때 생기는 다양한 윤리적인 이슈에 대한 논의도 진행되고 있다(손화철, 2023; 이주연, 2023; Hosseini, Resnik, & Holmes, 2023). 인공지능 기술이 빠르게 진보하고 널리 확산함에 따라 이러한 윤리적 및 사회적 문제의 발생 빈도와 심각성도 점차 증가하고 있는 현실이다. 따라서 인공지능을 개발하고 활용하는 당사자들이 인공지능과 관련된 올바른 윤리의식을 갖추고 적절한 윤리적 의사결정을 할 수 있는 능력을 기르는 것이 중요하다.

인공지능 윤리(AI ethics)는 ‘인공지능 연구, 개발, 적용, 폐기 등과 관련된 인공지능 기술의 전체 단계에서 인공지능을 둘러싼 규범적 문제’를 의미한다(허유선, 이연희, 심지원, 2020). 이와 비슷하게, 이우권(2020)은 인공지능 윤리를 ‘인공지능 관련 이해관계자들이 준수해야 할 보편적 사회 규범 및 관련 기술’이라고 정의한 바 있다. 김은경과 이영준(2023)에 따르면, 초기 인공지능 윤리 논의는 인공지능 자체의 책무에 초점을 맞췄으나, 점차 인공지능을 개발하고 활용하는 인간의 책무를 강조하는 방향으로 변화하고 있다. 인공지능과 관련된 여러 의사결정과 행위에 있어 기계에 불과한 인공지능에 책임을 전가하지 말고 인간이 어떤 방식으로든 책임을 부담해야 한다는 사회적 공감대가 형성되고 있는 것이다(이상욱, 이호영, 2021).

이에 세계 주요 국가 및 기관을 중심으로 바람직한 인공지능 개발 및 활용 방향을 제시하기 위한 윤리기준과 가이드라인이 마련되고 있다(Adams et al., 2023; Ashok, Madan, Joha, & Sivarajah, 2022; Fung & Etienne, 2023; Gal, 2019; Hagendorff, 2020). 김도연과 고영화(2022)에 따르면, 미국, 영국 등 서구 국가에서 최근 발표된 인공지능 윤리원칙 및 지침 관련 80여 편의 문헌에서 공통적으로 언급하고 있는 요소는 투명성, 공정성, 안전성, 책임성, 프라이버시보호, 편익성, 자율성, 신뢰성, 지속가능성, 존엄성, 연대성의 11가지이며, 이중 가장 빈번하게 언급된 4가지 원칙은 공정성, 책임성, 투명성, 안전성인 것으로 확인되었다. 세계경제포

럼, 유네스코, 유니세프 등의 인공지능 윤리 가이드라인을 비교한 Adams 등의 연구(2023) 결과, 이들 가이드라인이 투명성, 정의와 공정성, 비악의성, 책임성, 프라이버시, 공익성, 자유 및 자율성과 같은 주요 윤리원칙들을 공통적으로 포함하고 있었다. 이렇듯 세계 여러 국가와 기관의 인공지능 윤리원칙이 특정한 윤리 요소들을 보편적으로 공유하면서도 동시에 특수적이고 개별적인 요소들을 포함하고 있다. 그럼에도 이들 윤리원칙이 일관되게 견지하는 입장은 인공지능 윤리가 ‘인간중심적’이어야 한다는 점이다. 즉, 인공지능은 다양한 방식으로 인간의 복지 증진에 이바지해야 하며, 인공지능 개발 및 활용과 관련하여 발생하는 문제에 대한 책임은 인공지능이 아닌 인간에게 있다는 것이다(이상욱, 이호영, 2021).

이러한 국제사회의 동향에 발맞춰 우리나라 과학기술정보통신부와 정보통신정책연구원에서는 지난 2020년 ‘사람이 중심이 되는 인공지능 윤리기준’을 제정하고 발표한 바 있다(과학기술정보통신부, 2020). 이 기준은 인공지능 기술의 바람직한 개발과 활용을 위해 모든 사회구성원이 모든 분야에서 준수해야 할 3가지 기본 원칙으로 인간의 존엄성, 사회의 공공선, 기술의 합목적성을, 10대 핵심 요건으로 인권보장, 프라이버시 보호, 다양성존중, 침해금지, 공공성, 연대성, 데이터관리, 책임성, 안전성, 투명성을 제시하였다.

인공지능 윤리기준 및 가이드라인의 확립과 더불어, 일반 사용자의 인공지능 윤리 인식 및 역량을 측정하거나 탐색하기 위한 연구가 시도되고 있다(고영화, 임춘성, 2021; 김도연, 고영화, 2022; 배진아, 이정훈, 홍미선, 조정원, 2022b; 송애리, 고정아, 2023; 유수진, 장윤재, 2021; 이해령, 2023). 김도연과 고영화(2022)는 다양한 인공지능 윤리 가이드라인이 존재한다고 해도 인공지능 제품 및 서비스 사용자들이 적절한 윤리적 인식과 판단 능력을 갖추지 못할 경우, 인공지능 관련 윤리 문제가 근본적으로 해결되기 어렵다는 점을 강조한다. 이에 저자들은 인공지능 사용자의 인공지능 기술 윤리성을 정확히 측정하는 도구를 개발하고자 하였다. 전국 10 ~ 70대 남녀를 대상으로 실시한 자기보고식

설문을 기반으로 개발된 그들의 척도는 투명성, 안전성, 공정성, 책임성의 4개 인공지능 윤리 세부 요인을 측정할 수 있는 총 16개의 문항으로 구성되었다. 이와 유사한 방식으로 배진아 등(2022b)은 중·고등학생을 위한 인공지능 윤리역량 평가 척도 4개 요소(인간존중 및 책임, 공공성 및 다양성, 개인정보보호, 데이터관리 및 침해금지)에 대한 25개 문항을 개발하였다.

한편, 고영화와 임춘성(2021)은 전국 만 17세 이상 남녀 260명의 데이터를 분석한 그들의 연구를 통해 인공지능 기술과 서비스의 효과에 대한 일반 사용자들의 인식은 긍정적인 편이나, 이에 수반되는 부정적 측면 즉 윤리적 인식은 아직 매우 부족한 상태임을 보고하였다. 인공지능 활용 경험이 있는 전국 성인 남녀 1,000명을 대상으로 인공지능 윤리 인식 조사를 수행한 배영임과 김유나의 연구(2022) 결과에서도 일반 사용자들은 인공지능 제품과 서비스의 편리성 및 유용성에 대해 대체로 만족하며 긍정적인 인식을 드러냈다. 더욱이, 이 연구의 참여자들은 인공지능 윤리 문제의 심각성에 대해 매우 민감하게 반응하는 경향을 나타냈다. 특히 개인정보 침해, 책임 소재, 오남용과 관련된 문제들을 상당히 중요하게 여기고 있었다. 또한, 이 연구에서는 정부가 제시한 인공지능 윤리기준 3대 기본 원칙 및 10대 핵심 요건에 대해 참여자들이 인식하는 중요도 우선순위를 조사하였는데, 인간 존엄성 원칙 및 인권보장, 프라이버시보호, 안전성 등의 요건이 특히 중요한 것으로 나타났다.

한편, 인공지능 윤리 인식 및 태도를 결정짓는 요인(예: 인구·사회학적 특성, 인공지능 이해 및 활용 경험 등)을 탐색하는 연구도 수행되고 있다(김귀식, 신영준, 2021a; 오새내, 김효정, 2023; Farhi et al., 2023; Hartwig, Ikkatai, Takanashi, & Yokoyama, 2023). 예를 들어, Hartwig 등(2023)은 일본과 미국 일반인 각각 1,108명과 1,063명을 대상으로 인공지능 윤리 태도를 분석하였다. 연구 결과, 국가, 연령, 인공지능에 대한 이해 순으로 인공지능 윤리에 대한 태도를 예측하는 데 중요한 인구·사회학적 요인인 것으로 나타났고, 교육 수준, 성별, 결혼 여부 등은 상대적 중요성이 낮은 요인

이었다. 초·중·고등학생 577명을 대상으로 한 김귀식과 신영준의 연구(2021b)에서는 인공지능 윤리의식에 대한 성차를 밝혀냈는데, 예컨대 여학생은 남학생보다 인공지능에 대한 신뢰도가 낮고 차별 금지에 대한 우려가 큰 것으로 나타났다.

연령, 성별 등과 같은 인구·사회학적 특성과 더불어 인공지능 기술 및 도구에 대한 이해나 경험 역시 사용자의 인공지능 윤리 인식 및 역량과 관련되는 것으로 보인다. 예를 들어, 아랍에미리트(Unted Arab Emirates) 대학생 338명을 대상으로 ChatGPT에 대한 학생들의 관점, 우려 및 윤리의식을 탐색한 Farhi 등의 연구(2023)에 따르면, ChatGPT의 특성과 활용도에 대한 이해가 높은 학생일수록 ChatGPT 관련 윤리의식이 높은 것으로 나타났다. 오세내와 김효정(2023)은 국내 대학생 274명을 대상으로 ChatGPT 사용과 과제 표절 인식의 관계를 분석하였는데, 그 결과 ChatGPT 사용 경험 유무에 따라 ChatGPT 지속적 사용의도와 표절 정당화 인식의 상관관계에 차이가 있음을 확인하였다. ChatGPT 사용 경험이 없는 집단은 지속적 사용의도와 책임부존재형 정당화만 유의미한 상관관계를 보인 반면, ChatGPT 사용 경험이 있는 집단은 지속적 사용의도와 책임부존재형 정당화, 윤리상대주의 정당화, 형평주의 정당화에서 상관관계를 보이는 차이를 나타냈다.

인공지능 윤리에 대한 관심과 중요성이 증가함에 따라, 인공지능 윤리의식을 제고하고 교육하는 데 대한 필요성 인식이 커지는 것과 더불어 이를 위한 실천적 노력도 점차 확대되고 있다. 이에 인공지능 윤리 교육을 위한 교육과정 및 프로그램 개발 연구도 증가하는 추세에 있으나, 초등학생과 초등교사를 대상으로 한 교육 프로그램이 주를 이룬다(김하민, 2022; 배진아, 이정훈, 조정원, 2022a; 송주영, 전용주, 2023; 이지형, 변순용, 2023; 장연주, 최승윤, 조현기, 김현철, 2022; 최선유, 김귀훈, 2023; Lin et al, 2023). 예를 들어, 정다이와 박형빈(2023)은 초등학교 2022 개정 교육과정 중 도덕과 교육과정 내용 체계를 인공지능 윤리 교육의 측면에서 분석하여 자율주행 자동차, 사이보그, 뇌, 교육의 4가지 주제를 중심으로 한 인공지능 윤리프로그램을 개

발하였고 장연주 등(2022)은 신뢰성, 편향성, 악용가능성, 책임성을 주제로 초·중등학생을 위한 인공지능 윤리 교육 프로그램을 개발한 바 있다. 대학생 및 성인을 대상으로 한 연구도 소수지만 찾아볼 수 있는데, 예컨대 손혜숙(2022)은 인공지능 윤리기준 중 책임성, 투명성, 공정성, 통제성을 공통 요소로 도출한 후 실제 대학교양 수업에서 활용할 수 있는 인공지능 윤리 수업을 소설에 접목하여 설계 및 적용한 사례 연구를 수행하였다. 김은경과 이영준(2023)은 정부의 인공지능 윤리기준 3대 기본 원칙(인간 존엄성, 사회 공공선, 기술 합목적성)과 10대 핵심 요건이 반영된 인공지능 윤리 딜레마 9개를 인공지능 윤리 교육 자료로 개발하였다. 또한, 전수진(2023)은 CAI모델(Concept-Activity-Impression)을 적용하여 인공지능 윤리 교육 프로그램을 개발하고 효과성을 분석한 바 있다.

선행 연구에서는 공통적으로 인공지능의 바람직한 개발과 활용을 위해 다양한 이해관계자들을 위한 인공지능 윤리 교육이 필요하고, 이를 통해 인공지능이 인간의 삶에 미치는 부정적 영향과 잠재적인 위험을 감소시키면서 인공지능의 긍정적인 효과를 증대시키는 방법을 함양시킬 필요가 있음을 제안한다(Bogina, Hartman, Kuflik, & Shulner-Tal, 2021; Borenstein & Howard, 2021). 이에 상술한 것과 같이 다양한 인공지능 윤리 교육 프로그램이 개발되고 있으나 대학원생을 위한 프로그램은 찾기 힘들다(송선영, 2022). 하지만 대학원생은 학생, 연구자, 예비 교수자 등 다양한 정체성을 가지고 연구와 교육을 비롯한 많은 영역에서 인공지능을 개발 또는 활용하는 주체라는 점에서 인공지능 윤리와 관련된 적절한 교육이 필수적이다.

또한, 인공지능 윤리 교육에 관한 대부분의 선행 연구들이 실제 교육 대상자에 대한 교육 요구도를 체계적으로 분석하지 않았다는 제한점을 가진다. 인공지능 윤리기준 및 가이드라인이 매우 다양한 윤리개념과 쟁점을 내포한다는 점을 감안할 때, 효과적인 윤리 교육을 제공하기 위해서는 학습자의 현재 인공지능 윤리기준에 대한 이해 수준을 파악하고 그들에게 우선적으로 필요한 교육 내용이 무엇인지 살펴볼 필요가 있다. 특

히 학습자의 인식에 기반한 교육 요구를 조사하는 것이 매우 중요한데, 학습자가 스스로 필요성을 인식하고 교육에 참여할 때 더 높은 몰입과 효과를 기대할 수 있기 때문이다(신주연, 김진선, 정윤경, 2020). 이에 본 연구에서는 정부가 제시한 인공지능 윤리기준에 대한 대학원생의 현재 이해 수준(이하 ‘현재수준’) 및 그들이 인식하는 윤리기준의 중요수준을 분석하고, 이를 바탕으로 인공지능 윤리에 대한 대학원생의 교육 요구도를 분석하고자 한다.

2. 연구문제

- 연구문제 1. 인공지능 윤리에 대해 대학원생이 인식하는 교육 요구는 어떠한가?
- 연구문제 2. 인공지능 활용 빈도에 따라 인공지능 윤리에 대한 대학원생의 교육 요구에 차이가 나타나는가?

II. 연구방법

1. 연구참여자

본 연구를 위해 서울 소재 대규모 국립 종합대학교에 재학 중인 대학원생(수료생, 휴학생 포함)을 대상으로 2023년 12월 14 ~ 18일에 구글 폼을 통해 온라인 설문 조사를 실시하였다. 연구 참여에 동의한 619명으로부터 설문 결과를 수집하였고, 이 중 중복응답 및 불성실 응답 26건을 제외하고 총 593명의 설문 응답을 최종 분석에 사용하였다. 연구참여자의 인구통계학적 특성은 표 1과 같이 성별 분포는 여성 61.4%, 남성 38.4%, 미공개 0.2%이고, 학위과정은 석사과정 47.7%, 박사과정 29.3%, 석박통합과정 23%이었으며 전공계열은 이공계열 46.7%, 인문사회계열 37.3%, 의약학계열 14.8%, 예체능계열 1.2% 순이었다.

표 1. 연구참여자의 인구통계학적 특성 (N = 593)

항목	구분	빈도(명)	비율(%)
성별	남성	228	38.4
	여성	364	61.4
	미공개	1	0.2
소속	일반대학원	479	80.78
	전문대학원	114	19.22
학위과정	석사과정	283	47.7
	박사과정	174	29.3
	석박통합과정	136	23.0
전공계열	인문사회계열	221	37.3
	이공계열	277	46.7
	의약학계열	88	14.8
	예체능계열	7	1.2
계		593	100

2. 연구도구

1) 인공지능에 대한 이해 및 활용 경험

인공지능 윤리 관련 질문에 대한 응답의 신뢰성을 확보하기 위한 목적으로 연구참여자의 인공지능 이해 수준과 활용 경험을 측정하였다. 우선 인공지능에 대한 학문적 정의를 제공하고, 인공지능 및 관련 기술에 대해 얼마나 잘 알고 있다고 생각하는지 7점 Likert 척도 (1 = 잘 모른다, 7 = 매우 잘 안다)로 평정하게 하였다. 이어서 다양한 인공지능 제품, 도구 및 서비스의 12가지 사례(예: 자율주행 자동차, 인공지능 챗봇, 음성인식 인공지능 비서, 인공지능 번역 등)를 제공하고 자신이 활용해본 것을 모두 표시하도록 하였다. 마지막으로 전반적으로 얼마나 자주 인공지능 제품, 도구 또는 서비스를 활용하고 있는지를 파악하여 인공지능 활용 빈도 수준을 조사하였다. 활용 빈도는 전혀 사용하지 않음, 주 1 ~ 2일, 주 3 ~ 4일, 주 5 ~ 6일, 매일의 다섯 개 선택지로 측정하였으며, 참여자 집단별 분석을 위해 인공지능을 주 2회 이하로 사용하는 경우를 인공지능 활용 빈도 ‘저수준’으로 구분하였으며, 인공지능을 주 3회 ~ 4회 정도 사용하는 경우를 ‘중수준’으로, 인공지능을 주 5회 이상 활용하는 경우를 ‘고수준’으로 구분하였다.

2) 인공지능 윤리 인식

대학원생이 인공지능 윤리에 대해 얼마나 관심이 있고 그 중요성 및 심각성에 대해 어느 정도로 인지하고 있는지 파악하기 위해 관련 5개 문항(예: ‘인공지능 관련 윤리적 이슈는 얼마나 중요하다고 생각하십니까?’, ‘인공지능 관련 윤리적 이슈는 현재 얼마나 심각하다고 생각하십니까?’)을 사용하였다. 문항을 제시하기 전 인공지능 챗봇 이루다 논쟁, 영국 대입시험 알고리즘 차별 논란, 아마존의 데이터 편향성 논란 등 인공지능 윤리와 관련된 사회적 이슈를 사례로 제시하였다. 모든 문항은 7점 Likert 척도(1 = 전혀 모른다, 7 = 매우 잘 안다)로 평정하게 하였다.

3) 인공지능 윤리기준에 대한 현재수준 및 중요 수준

선행연구에서 살펴본 바와 같이 국내외에서 발표된 인공지능 윤리기준들은 매우 다양하면서도 상호 유사한 특성을 공유한다(김도연, 고영화, 2022; 이상욱, 이호영, 2021). 본 연구에서는 국내 인공지능 윤리기준 중 가장 대표적이면서 앞으로 인공지능 연구 및 개발과 활용 현장에서 매우 빈번하게 사용될 것으로 예상되는 정부의 인공지능 윤리기준을 측정도구로 선택하였다. 인공지능 윤리기준에 대한 대학원생의 현재수준(현재 이해하고 있는 정도) 및 중요수준을 측정하기 위해 정부에서 제시한 ‘사람이 중심이 되는 인공지능 윤리기준(과학기술정보통신부, 2020)’의 10대 핵심 요건 총 22개 항목을 사용하였다.

우선, 해당 항목에 대한 대학원생의 현재수준을 측정하기 위해 각 항목의 의미 및 관련 윤리적 이슈와 사례를 현재 어느 정도 알고 있는지 7점 Likert 척도(1 = 전혀 모른다, 7 = 매우 잘 안다)로 평가하게 하였다. 다음으로 인공지능 윤리 항목의 중요수준을 평가하기 위해 각 항목별로 응답자가 본인의 삶과 우리 사회에 해당 인공지능 윤리 항목이 어느 정도 중요하다고 여기는지 7점 Likert 척도(1 = 전혀 중요하지 않다, 7 = 매우 중요하다)로 표시하도록 하였다. 인공지능 윤리기준에 대

한 현재 및 중요수준을 측정하기 위해 본 연구에 사용한 문항별 내용과 신뢰도(α)는 표 2에서 보는 바와 같이 .69 ~ .89로 신뢰로운 수준으로 나타났다.

3. 자료 분석

본 연구의 목적인 대학원생의 인공지능 윤리 교육의 요구도를 분석하기 위해 Borich 요구도 분석과 The Locus for Focus 모델을 활용하였으며, 모든 분석에는 SPSS Statistics 26을 사용하였다. 본 연구의 자료에 결측치는 존재하지 않아 별도의 결측치 처리 방법을 사용하지 않았다. 구체적인 분석 절차와 방법은 다음과 같다. 우선, 설문도구로 사용한 문항 중, 인공지능 이해와 활용 경험, 인공지능 윤리의 중요성과 심각성 인식에 대한 인식, 인공지능 윤리 교육에 대한 연구참여자 전체의 현재수준 및 중요수준, 그리고 인공지능 활용 빈도에 따른 인공지능 윤리 교육의 현재수준 및 중요수준 대한 기술 통계 분석을 실시하였다. 다음으로, 인공지능 윤리기준 10대 핵심 요건 별로 대학원생의 현재수준과 중요수준에 유의미한 차이가 있는지 분석하기 위해 대응표본 t -검정을 실시하였다. 이어서, 현재수준과 중요수준의 차이를 바탕으로 교육 요구의 우선순위를 파악하기 위해 Borich 요구도 분석을 실시하였다. Borich 요구도 분석은 현재수준과 중요수준의 차이를 고려하여 각 항목의 중요수준 평균값을 가중치로 부여하고 이를 총 사례 수로 나누어 요구수준을 파악하는 방식으로 (Borich, 1980), 계산 방식은 그림 1과 같다. Borich 요구도의 이 같은 계산 방식은 단순히 두 수준 간의 차이값을 기준으로 우선순위를 도출하는 IPA(Importance-Performance Analysis)보다 더 정교한 방법으로 여겨진다(조대연, 2009).

마지막으로, Borich 요구도 분석에서 도출된 우선순위 정보를 어느 순위까지 일차적으로 고려해야 하는지 결정하기 위해 The Locus for Focus 모델을 활용하였다. The Locus for Focus 모델은 좌표평면을 통해 교육 우선순위를 시각적으로 확인할 수 있게 한다. 구체적으로, 좌표평면의 x 축 중앙값은 중요수준의 평균값을, y

측의 중앙값은 중요수준과 현재수준 간 차이의 평균값을 표시한다. 이때, 제1사분면(HH)에 속한 항목들은 중요수준과 현재수준 차이가 평균보다 높고 중요수준의 값도 평균보다 높은 항목들로, 가장 높은 수준의 요구가 있다고 판단할 수 있다. 제2사분면(LH)과 제4사

분면(HL)에 속한 경우, 중요수준은 낮지만 중요수준과 현재수준 차이가 크거나, 반대로 중요수준은 높지만 중요수준과 현재수준 차이가 낮은 경우로 둘 중 어느 사분면이 요구도가 더 높은지는 연구 맥락과 연구자의 해석에 의존한다. 제3사분면(LL)은 중요수준도 낮고 중

표 2. 인공지능 윤리기준 문항 및 신뢰도

구분	내용	신뢰도(α)	
		현재 수준	중요 수준
A. 인권보장	1. 인공지능의 개발과 활용은 모든 인간에게 동등하게 부여된 권리를 존중하고, 다양한 민주적 가치와 국제 인권법 등에 명시된 권리를 보장하여야 한다.	.79	.69
	2. 인공지능의 개발과 활용은 인간의 권리와 자유를 침해해서는 안 된다.		
B. 프라이버시 보호	3. 인공지능을 개발하고 활용하는 전 과정에서 개인의 프라이버시를 보호해야 한다.	.78	.77
	4. 인공지능 전 생애주기에 걸쳐 개인정보의 오용을 최소화하도록 노력해야 한다.		
C. 다양성존중	5. 인공지능 개발 및 활용 전 단계에서 사용자의 다양성과 대표성을 반영해야 하며, 성별·연령·장애·인종·종교·국가 등 개인 특성에 따른 편향과 차별을 최소화하고, 상용화된 인공지능은 모든 사람에게 공정하게 적용되어야 한다.	.82	.76
	6. 사회적 약자 및 취약 계층의 인공지능 기술 및 서비스에 대한 접근성을 보장하고, 인공지능이 주는 혜택은 특정 집단이 아닌 모든 사람에게 골고루 분배되도록 노력해야 한다.		
D. 침해금지	7. 인공지능을 인간에게 직·간접적인 해를 입히는 목적으로 활용해서는 안 된다.	.81	.78
	8. 인공지능이 야기할 수 있는 위험과 부정적 결과에 대응 방안을 마련하도록 노력해야 한다.		
E. 공공성	9. 인공지능은 개인적 행복 추구 뿐만 아니라 사회적 공공성 증진과 인류의 공동 이익을 위해 활용해야 한다.	.89	.84
	10. 인공지능은 긍정적 사회변화를 이끄는 방향으로 활용되어야 한다.		
	11. 인공지능의 순기능을 극대화하고 역기능을 최소화하기 위한 교육을 다방면으로 시행하여야 한다.		
F. 연대성	12. 다양한 집단 간의 관계 연대성을 유지하고, 미래세대를 충분히 배려하여 인공지능을 활용해야 한다.	.87	.82
	13. 인공지능 전 주기에 걸쳐 다양한 주체들의 공정한 참여 기회를 보장하여야 한다.		
	14. 윤리적 인공지능의 개발 및 활용에 국제사회가 협력하도록 노력해야 한다.		
G. 데이터관리	15. 개인정보 등 각각의 데이터를 그 목적에 부합하도록 활용하고, 목적 외 용도를 활용하지 않아야 한다.	.84	.72
	16. 데이터수집과 활용의 전 과정에서 데이터 편향성이 최소화되도록 데이터 품질과 위험을 관리해야 한다.		
H. 책임성	17. 인공지능 개발 및 활용과정에서 책임주체를 설정함으로써 발생할 수 있는 피해를 최소화하도록 노력해야 한다.	.85	.81
	18. 인공지능 설계 및 개발자, 서비스 제공자, 사용자 간의 책임소재를 명확히 해야 한다.		
I. 안전성	19. 인공지능 개발 및 활용 전 과정에 걸쳐 잠재적 위험을 방지하고 안전을 보장할 수 있도록 노력해야 한다.	.86	.73
	20. 인공지능 활용 과정에서 명백한 오류 또는 침해가 발생할 때 사용자가 그 작동을 제어할 수 있는 기능을 갖추도록 노력해야 한다.		
J. 투명성	21. 사회적 신뢰 형성을 위해 인공지능의 투명성과 설명 가능성을 높이고, 타 원칙과의 상충관계를 고려하여 활용 상황에 적합한 수준의 투명성과 설명 가능성을 높이려는 노력을 기울여야 한다.	.85	.74
	22. 인공지능기반 제품이나 서비스를 제공할 때 인공지능의 활용 내용과 활용 과정에서 발생할 수 있는 위험 등의 유의사항을 사전에 고지해야 한다.		

$$Borich \text{ 요구도} = \frac{\sum (\text{중요수준} - \text{현재수준}) \times \text{중요수준 평균}}{\text{전체 사례 수}}$$

그림 1. Borich 요구도 계산 방식

요수준과 현재수준 차이가 낮은 경우로 전체 요구도는 가장 낮다고 볼 수 있다. 본 연구에서는 Borich 요구도 분석에서 상위 순위로 도출된 항목 중 The Locus for Focus의 제 1사분면에 속하는 항목을 가장 교육 요구가 높은 항목으로 선정하였고, The Locus for Focus 모델의 제 1사분면 또는 Borich 요구도 상위 순위 중 어느 하나에만 요구도가 나타나면 해당 항목을 차순위 항목으로 결정하였다(조대연, 2009).

III. 연구결과

1. 인공지능 윤리에 대한 대학원생의 교육 요구

1) 인공지능 이해와 활용 경험

인공지능 윤리 교육 요구도를 해석하는 데 필요한 연구참여자의 기본 정보를 파악하고 응답의 신뢰성을 확보하기 위해 그들의 인공지능 이해 수준과 활용 경험을 조사하였다. 본 연구에 참여한 대학원생의 인공지능 및 관련 기술에 대한 이해는 보통 수준인 것으로 나타났다($M = 3.97$, $SD = 1.47$). 인공지능 제품, 도구 및 서비스 12가지 사례에 대한 활용 경험을 살펴본 결과, 모든 참여자가 이 중 최소 1가지 이상을 활용해 본 경험이 있었다. 이 중 응답 비중은 인공지능 챗봇(예: ChatGPT, 이루다 등)이 89.88%로 가장 높았으며, 인공지능 번역 서비스(예: 파파고 등)가 77.40%, 콘텐츠 추천 기능(예: 넷플릭스, 유튜브 등)이 71.67%, 음성인식 인공지능 비서(예: 오케이구글, 시리 등)가 68.63%로 그 뒤를 이었다. 또한, 인공지능 활용 빈도가 고수준(주 4회 이상)에 속하는 대학원생이 전체 참여자의 46.2%로 가장 큰 비중을 차지했으며, 저수준(주 2회 이하, 28%), 중수준(주 3회, 25.8%) 순으로 나타났다.

2) 인공지능 윤리의 중요성과 심각성에 대한 인식

인공지능 윤리에 대한 전반적 인식을 살펴본 결과, 대학원생의 인공지능 윤리에 대한 관심은 보통을 상회

하는 수준이었으나($M = 4.61$, $SD = 1.47$) 인공지능 윤리에 대한 이해 정도는 보통 수준에 그쳤으며($M = 4.02$, $SD = 1.42$), 이 차이는 통계적으로도 유의하였다($t = 10.32$, $p < .001$). 한편, 대학원생이 인지하는 인공지능 윤리의 중요성($M = 5.96$, $SD = 1.14$)과 심각성(현재: $M = 4.81$, $SD = 1.29$, 미래: $M = 5.83$, $SD = 1.23$)은 대체적으로 높은 편이었다. 특히, 학생들은 인공지능 윤리 문제가 지금보다 앞으로 더 심각한 문제가 될 것이라고 지각하고 있었다($t = 21.08$, $p < .001$).

3) 인공지능 윤리에 대한 교육 요구도

인공지능 윤리기준 10대 핵심요건에 대한 대학원생의 교육 요구도를 파악하기 위해 대학원생이 보유한 인공지능 윤리 인식의 현재수준과 중요수준의 차이를 분석하고 교육 우선순위를 도출하였으며, 그 결과를 표 3에 제시하였다. 전반적으로, 윤리기준 10대 요건에 대한 대학원생의 현재수준은 보통을 웃도는 수준($M = 4.48 \sim 5.36$)인 반면, 대학원생이 인식하는 윤리기준의 중요수준은 그보다 더 높은 편이었다($M = 5.65 \sim 6.3$). 구체적으로, 현재수준이 가장 높은 항목은 침해금지($M = 5.36$, $SD = 1.38$)로 나타났고, 인권보장($M = 4.48$, $SD = 1.41$)은 가장 낮게 나타났다. 중요수준은 침해금지($M = 6.30$, $SD = 0.85$)가 가장 높았고, 연대성($M = 5.65$, $SD = 1.10$)이 가장 낮았다. 다시 말해, 본 연구에 참여한 대학원생들은 인공지능 윤리기준 중 침해금지에 대한 현재 인식 수준이 가장 높고 인권보장에 대한 인식 수준이 가장 낮게 나타났으며, 동시에 침해금지에 대한 중요성을 가장 높게 인식하고 있었고, 연대성에 대한 중요성 인식은 가장 낮은 것으로 나타났다.

t-검정 결과, 윤리기준 모든 항목에서 현재수준과 중요수준의 차이가 통계적으로 유의한 것을 확인하였다. Borich 요구도 값은 인권보장(8.34)이 가장 높았고, 다음으로 안전성(6.56), 책임성(6.41), 투명성(6.23), 침해금지(5.93)의 순으로 높게 나타났다. 반면, 연대성(5.24), 데이터관리(4.86), 공공성(4.84)의 Borich 요구도 값은 상대적으로 낮은 편이었다.

대학원생의 인공지능 윤리 교육에 관한 요구도를 The Locus for Focus 모델을 사용하여 분석한 우선순위는 그림 2와 같다. 제1사분면(HH)에 위치한 항목은 안전성, 책임성, 투명성이었고, 제2사분면(LH)에 위치한 항목은 인권보장이 유일하였다. 그리고 제3사분면(LL)에 위치한 다양성존중, 연대성, 공공성을 제외한 나머지 3개의 항목이 제4사분면(HL)에 위치하는 것으로 확인되었다.

Borich 요구도에 의한 우선순위와 The Locus for Focus 모델을 비교하여 도출한 대학원생의 인공지능

윤리에 관한 교육 요구도 우선순위는 표 4와 같다. 즉 Borich 요구도와 The Locus for Focus 모델에서 공통적으로 높은 우선순위로 나타난 항목은 안전성, 책임성, 투명성이라 볼 수 있으며, The Locus for Focus 모델에서는 제 1사분면에 위치하지 않았으나, Borich 요구도 공식에서는 높은 우선순위로 나타난 인권보장과 침해금지의 항목은 차순위로 요구도가 높은 항목으로 볼 수 있다.

표 3. 인공지능 윤리 교육에 대한 Borich 요구도 분석 결과

항목	현재수준		중요수준		차이(중요-현재)			Borich 요구도	우선 순위
	M	SD	M	SD	M	SD	t		
A. 인권보장	4.48	1.42	5.90	0.98	1.41	1.45	23.78***	8.34	1
I. 안전성	5.03	1.47	6.11	0.90	1.08	1.52	17.21***	6.56	2
H. 책임성	4.98	1.51	6.04	0.96	1.06	1.48	17.52***	6.41	3
J. 투명성	4.99	1.48	6.03	1.48	1.03	1.51	16.65***	6.23	4
D. 침해금지	5.36	1.38	6.30	0.85	0.94	1.41	16.19***	5.93	5
B. 프라이버시보호	5.24	1.36	6.18	0.91	0.94	1.45	15.90***	5.84	6
C. 다양성 존중	4.82	1.52	5.79	1.05	0.97	1.53	15.36***	5.59	7
F. 연대성	4.73	1.48	5.65	1.10	0.93	1.44	15.64***	5.24	8
G. 데이터관리	5.32	1.38	6.12	0.93	0.80	1.38	14.04***	4.86	9
E. 공공성	5.07	1.42	5.89	0.98	0.82	1.43	14.00***	4.84	10

*** $p < .001$.



그림 2. The Locus for Focus 분석 결과 (대학원생 전체)

2. 인공지능 활용 빈도에 따른 인공지능 윤리 교육 요구 차이

인공지능 활용 빈도에 따라 대학생의 인공지능 윤리에 대한 교육 요구도에 차이가 존재하는지 살펴보았다. 인공지능 활용 빈도 수준(저·중·고)별 인공지능 윤리기준에 대한 Borich 요구도 1 ~ 5순위에 해당하는 항목을 표 5에서 제시하였다. 인공지능 활용 빈도가 저 수준(주 2회 이하)에 속한 집단의 경우 인권보장(8.39)의 Borich 요구도 값이 가장 높았고, 다음으로는 책임성(6.46), 프라이버시보호(6.28), 침해금지(6.26), 투명성(6.26)이 뒤를 이었다. 인공지능 활용 빈도 중수준(주 3회) 집단의 Borich 요구도 값은 인권보장(7.81), 투명성(6.59), 안전성(6.36), 프라이버시보호(5.74), 책임성(5.65) 순으로 높게 나타났다. 마지막으로, 인공지능 활

용 빈도 고수준(주 4회 이상) 집단의 Borich 요구도 우선순위는 인권보장(8.63), 안전성(7.13), 책임성(6.82), 침해금지(6.22), 투명성(6.04)의 순이었다. 이상 Borich 요구도 분석 결과를 종합하면, 인공지능 활용 빈도에 상관없이 교육 요구도가 가장 높은 항목은 인권보장이었고, 책임성, 투명성도 높은 요구도를 나타냈다. 한편, 인공지능 활용 빈도가 고수준으로 높아지면서 프라이버시보호의 교육 요구도가 드러나지 않는 반면, 안전성의 교육 요구도는 높은 우선 순위로 나타나는 경향을 보였다.

대학원생의 인공지능 활용 빈도에 따른 The Locus for Focus 모델을 분석한 결과는 그림 3과 같다. 인공지능 활용 빈도 저수준 집단의 경우 The Locus for Focus 모델의 제1사분면(HH)에 속한 항목이 총 3개로 책임성, 프라이버시보호, 투명성이 그것이었다. 인공지능 활용 빈도 중수준 집단의 제1사분면(HH)에는 투명성,

표 4. 대학원생의 인공지능 윤리에 관한 교육 요구도 우선순위

항목	Borich 요구도	The Locus for Focus	최종 교육요구도
A.인권보장	1	LH	차순위
I.안전성	2	HH	최우선순위
H.책임성	3	HH	최우선순위
J.투명성	4	HH	최우선순위
D.침해금지	5	HL	차순위

표 5. 인공지능 활용빈도에 따른 인공지능 윤리에 관한 Borich 교육 요구도 분석 결과

활용 빈도	항목	현재수준		중요수준		차이(중요-현재)			Borich 요구도	우선 순위
		M	SD	M	SD	M	SD	t		
저	A. 인권보장	4.26	1.41	5.72	1.06	1.47	1.55	12.15***	8.39	1
	H. 책임성	4.90	1.58	5.98	0.92	1.08	1.54	9.00***	6.46	2
	B. 프라이버시보호	4.95	1.44	6.00	1.01	1.05	1.54	8.77***	6.28	3
	D. 침해금지	5.15	1.46	6.17	0.89	1.02	1.54	8.49***	6.26	4
	J. 투명성	4.88	1.53	5.93	0.91	1.05	1.64	8.23***	6.26	5
중	A. 인권보장	4.54	1.47	5.87	1.04	1.33	1.36	12.14***	7.81	1
	J. 투명성	4.92	1.37	6.02	0.97	1.09	1.36	9.94***	6.59	2
	I. 안전성	5.09	1.37	6.12	0.88	1.04	1.32	9.73***	6.36	3
	B. 프라이버시보호	5.31	1.25	6.23	0.85	0.92	1.27	8.97***	5.74	4
	H. 책임성	5.07	1.35	6.01	1.00	0.94	1.25	9.33***	5.65	5
고	A. 인권보장	4.58	1.38	6.02	0.87	1.43	1.44	16.49***	8.63	1
	I. 안전성	5.07	1.52	6.22	0.88	1.15	1.56	12.19***	7.13	2
	H. 책임성	4.98	1.56	6.10	0.96	1.12	1.55	11.94***	6.82	3
	D. 침해금지	5.43	1.38	6.40	0.77	0.97	1.41	11.44***	6.22	4
	J. 투명성	5.10	1.52	6.10	0.92	0.99	1.51	10.84***	6.04	5

*** $p < .001$.

안전성, 책임성이 위치하였다. 마지막으로 인공지능 활용 빈도 고수준 집단은 안전성, 책임성만이 제1사분면(HH)에 나타났다.

인공지능 활용 빈도에 따른 대학원생 집단별로 Borich 요구도 분석 결과와 The Locus for Focus 모델을 통한 시각화 결과를 비교한 결과를 표 6에 제시하였다. 연구 결과, 인공지능 활용 빈도에 따라 대학원생의 인공지능

윤리 교육 요구도는 차이가 있는 것으로 나타났다. 구체적으로, 인공지능 활용 빈도 집단별 교육 요구 최우선순위 내용에 차이가 관찰되었다. 인공지능 활용 빈도가 가장 낮은 집단의 학생에게 가장 우선으로 다루어야 할 내용은 책임성, 프라이버시보호, 투명성으로 확인되었다. 이어서 인공지능 활용 빈도 중수준 집단에는 책임성, 투명성, 안전성이, 고수준 집단에는 책임성과

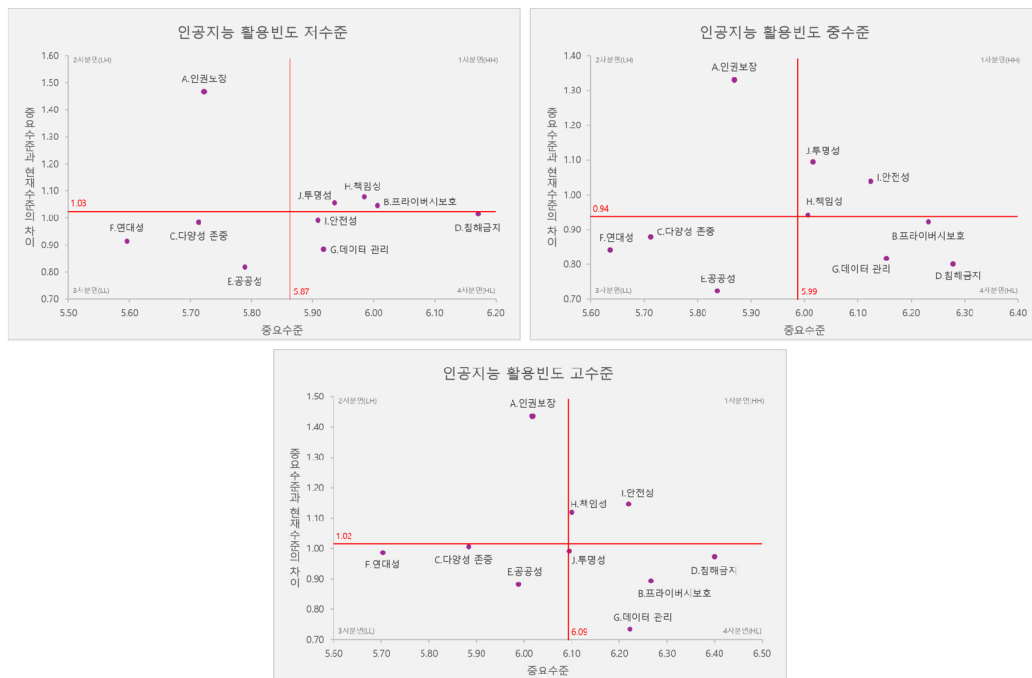


그림 3. The Locus for Focus 분석 결과 (인공지능 활용빈도 수준에 따라)

표 6. 인공지능 활용 빈도에 따른 인공지능 윤리에 관한 교육 요구도 우선순위

인공지능 활용빈도	항목	Borich 요구도	The Locus for Focus	최종 교육요구도
저	A. 인권보장	1	LH	차순위
	H. 책임성	2	HH	최우선순위
	B. 프라이버시보호	3	HH	최우선순위
	D. 침해금지	4	HL	차순위
	J. 투명성	5	HH	최우선순위
중	A. 인권보장	1	LH	차순위
	J. 투명성	2	HH	최우선순위
	I. 안전성	3	HH	최우선순위
	B. 프라이버시보호	4	HL	차순위
	H. 책임성	5	HH	최우선순위
고	A. 인권보장	1	LH	차순위
	I. 안전성	2	HH	최우선순위
	H. 책임성	3	HH	최우선순위
	D. 침해금지	4	HL	차순위
	J. 투명성	5	HL	차순위

안전성이 가장 우선으로 교육에 제공될 필요가 있는 항목으로 분석되었다. 즉 책임성은 인공지능 활용 빈도와 관계없이 일관되게 최우선순위의 교육 요구도를 나타내었으며, 투명성은 저수준과 중수준 집단에서, 안전성은 중수준과 고수준 집단에서 교육 요구 최우선순위로 도출되었다. 또한, 인공지능 활용 빈도에 따라 교육 요구 정도에 변화가 생기는 윤리기준이 관찰되었는데, 예컨대 프라이버시보호와 안전성의 경우이다. 프라이버시보호는 인공지능 활용 빈도 저수준, 중수준 집단에서는 각각 최우선순위, 차순위로 도출되었다가 인공지능 활용 빈도 고수준 집단에서는 우선순위 1~5순위에서 제외되었다. 반대로 안전성은 인공지능 활용 빈도 저수준 집단에서는 교육 우선순위 1~5순위에 도출되지 않았지만 중수준, 고수준 집단에서는 최우선순위로 선정되었다.

IV. 논의 및 결론

본 연구는 Borich 요구도 분석과 The Locus for Focus 모델을 활용하여 인공지능 윤리에 대한 대학원생의 교육 요구도를 파악하고자 하였다. 이와 더불어 학생들의 인공지능 활용 빈도에 따라 인공지능 윤리 교육 요구도에 차이가 있는지 알아보았다. 연구 결과를 요약하면, 전반적으로 학생들은 인공지능 윤리 문제를 상당히 중요하게 인식하며 이 문제가 앞으로 더욱 심각해질 것으로 전망하고 있었으나, 스스로 인지하는 인공지능 윤리에 대한 이해 정도는 그만큼 높지 않았다. 인공지능 윤리기준 10대 핵심 요건을 바탕으로 살펴본 대학원생의 교육 요구도 분석 결과, 대학원생을 위한 인공지능 윤리 교육에서 우선적으로 다루어야 할 내용은 안전성, 책임성, 투명성, 인권보장, 침해금지 등으로 확인되었다. 한편, 이 같은 교육 요구의 우선순위는 대학원생의 인공지능 활용 빈도에 따라 다소 차이가 있었다.

본 연구는 교육 요구 도출에 있어 직접적인 교육 수혜자의 관점과 특성을 반영하였다는 점과 Borich 분석

및 The Locus for Focus 모델을 동시에 활용하여 보다 정밀한 방법으로 요구도를 분석하였다는 데에 그 의미가 있다. 본 연구의 결과는 대학원생 대상 인공지능 윤리 교육 프로그램 개발을 위한 유용한 기초 자료로 활용되어 학생들의 인공지능 윤리의식과 역량을 고취하는데 일조할 것으로 기대한다. 주요 연구 결과를 정리하고 그에 따른 함의와 시사점을 논의해보면 다음과 같다.

첫째, 본 연구에 참여한 전체 대학원생 및 인공지능 활용 빈도에 따라 구분된 하위 집단(저·중·고) 모두에서 인공지능 윤리에 대한 교육 요구 최우선순위로 도출된 내용은 책임성인 것으로 나타났다. 책임성은 다양한 인공지능 윤리원칙 및 가이드라인에서 가장 빈번하게 언급되는 핵심 요소 중 하나라는 점에서(김도연, 고영화, 2022), 대학원생이 스스로 그 교육의 필요성을 지각하고 있다는 점은 매우 고무적인 일이다. 인공지능 책임성 원칙은 인공지능 개발 및 활용 과정 중 발생할 수 있는 피해를 최소화하기 위해 책임 주체를 설정하고, 이해관계자(예: 개발자, 서비스 제공자, 사용자 등) 간의 책임소재를 분명히 해야 한다는 것이 그 골자이다(과학기술정보통신부, 2020). 책임성과 관련된 인공지능 윤리적 이슈의 대표적 사례로 자율주행차량의 교통사고와 관련된 책임소재 문제를 들 수 있다. 자율주행차량 교통사고 발생 시 그 책임의 주체가 차량을 제작하고 판매한 공급자에게 있는지, 차량의 사용한 운전자에게 있는지, 아니면 자율주행시스템 관리자에게 있는지는 개인이 쉽게 판단하기 어려운 문제이며 충분한 사회적 합의가 필요한 사안이기도 하다(김민호, 박현지, 2020; 임이정, 송재인, 황기연, 2019). 대학원생과 가장 관련이 있을 법한 인공지능 책임성 이슈로는 생성형 인공지능을 활용한 논문 작성과 관련된 문제가 있다. 최근 국제 저명 학술지 네이처와 사이언스는 ChatGPT가 사람 저자와는 달리 연구 수행 과정과 그 결과에 대한 도덕적, 사회적 책임을 지지 않는다는 점을 지적하며 ChatGPT를 논문 저자로 인정하지 않는 가이드라인을 제시한 바 있다(한겨레, 2023). ChatGPT 논문 저자 인정 여부의 문제는 미시적 차원으로는 연구 윤리(research ethics) 측면에서 판단할 수 있겠지만 거시적으로는 연구자의 사

회적 책임의 관점에서 논의할 수 있다. 이렇듯 인공지능 책임성과 관련된 대부분의 윤리 문제가 복잡하고 난해한 성격을 띠므로 관련 주제와 쟁점들을 충분히 이해하고 이를 바탕으로 바람직한 윤리적 판단과 의사결정 능력을 함양할 수 있도록 교육 내용을 구성하는 것이 필요하다.

둘째, 책임성과 더불어 안전성, 투명성 역시 본 연구의 참여자 전체에 대한 교육 요구 최우선순위로 도출된 내용이었다. 이는 세계 각국에서 발표된 80여편의 인공지능 윤리 관련 문헌에서 공통적으로 언급하고 있는 네 가지 윤리원칙(공정성, 책임성, 투명성, 안전성)과 궤를 같이하는 결과이다. 대학원생을 위한 인공지능 윤리 교육 제공 시 여러 인공지능 윤리기준 중에서도 가장 본질적이고 핵심적인 요소들을 우선적으로 다룰 필요가 있다는 점을 시사한다. 인공지능 안전성은 인공지능 개발 및 활용의 전 과정에서 잠재적 위험을 방지할 수 있도록 노력하며, 예기치 않은 오류나 침해발생 시 이에 대처할 수 있는 장치를 갖추도록 해야 한다는 원칙이다(과학기술정보통신부, 2020). 자율주행차량, 인공지능 로봇 등 인공지능 제품이나 서비스가 기계적 결함이나 오작동으로 인해 사람에게 신체적, 정신적 해를 가하게 되는 경우가 안전성의 원칙이 준수되지 않았을 시 발생할 수 있는 사례이다. 이 외에도 인공지능 알고리즘으로 인해 개인정보가 유출되거나 개인 프라이버시가 침해되는 등의 문제도 안전성과 연관된다(김명주, 2017). 한편, 인공지능 투명성은 인공지능 제품이나 알고리즘의 작동 원리와 성능 등의 정보 공개와 인공지능 활용 내용 및 과정에서 발생할 수 있는 위험에 대한 사전 고지를 강조하는 윤리 원칙이다(과학기술정보통신부, 2020). 인공지능을 직접 개발하는 일 또는 연구를 수행하는 경우를 제외하고, 일반적인 사용자로서의 대학원생에게 안전성과 투명성의 교육 요구가 높다는 것은 인공지능 보급 확산과 활용 증대로 야기될 수 있는 예기치 못한 피해에 대한 학생들의 막연한 두려움이 반영된 결과일 수 있다. 인공지능 안전성과 투명성의 원칙을 충족하기 개발자뿐만 아니라 사용자 측면에서 숙지하고 실천해야 할 구체적 전략에 대한 교

육이 필요할 것으로 보인다.

셋째, 대학원생의 인공지능 윤리 교육 요구도에서 최우선순위로 나타난 책임성, 안전성, 투명성에 이어 교육 요구 차순위로 확인된 것은 인권보장, 침해금지인 것으로 분석되었다. 특히, 인권보장의 경우, 참여자의 현재수준이 가장 낮은 것으로 나타난 항목이었다. 인공지능 개발과 활용이 다양한 민주적 가치를 보장하여야 하고 인간의 권리와 자유를 침해하면 안 된다는 것이 핵심인데, 그 개념이 추상적이고 포괄적인 면이 있어 참여자들이 정확한 의미와 구체적 사례를 알고 있다고 여기기 쉽지 않았을 것으로 짐작된다. 지적재산권 문제, 개인 프라이버시 문제 등 매우 다양한 이슈들이 인권보장 윤리원칙과 관련되는데, 예컨대 인공지능 음성 인식 비서(예: 오케이구글, 시리)의 경우, 사용자의 음성 데이터가 수집되는 과정에서 지적재산 및 개인정보 유출, 프라이버시 침해 문제 등이 발생할 수 있다(오하경, 김태성, 2017; 윤진숙, 2022). 또한, 대부분의 인공지능 비서가 여성의 목소리를 기본값으로 낸다는 점에서 비서는 여성이라는 고정관념을 심화시켜 여성 인권 침해의 여지도 존재한다(윤진숙, 2022). 인공지능 음성 인식 비서는 본 연구의 참여자의 약 70%가 사용해본 적이 있다고 답변했을 정도로 일상에서 흔히 접하는 인공지능 서비스인 만큼 관련 윤리적 이슈를 충분히 인지하고 피해를 효과적으로 예방 및 대처할 수 있도록 돕는 적절한 교육이 필요하다. 또한, 음성인식 비서를 포함한 다양한 사례를 통해 인권보장 및 침해금지와 관련된 쟁점들을 효과적으로 이해하고 생활 속에서 실천할 수 있는 역량을 함양시키는 것이 중요하다.

넷째, 인공지능 활용 빈도 집단별 교육 요구 최우선순위 내용에 차이가 관찰되었는데, 구체적으로 인공지능 활용 빈도 저수준 집단은 책임성, 프라이버시보호, 투명성이, 중수준 집단은 책임성, 투명성, 안전성이, 고수준 집단은 책임성과 안전성이 최우선순위로 도출된 내용이었다. 인공지능 제품, 도구 또는 서비스를 자주 활용할수록, 또는 어떤 유형의 인공지능을 주로 활용하느냐에 따라 인공지능과 관련된 특정 윤리적 이슈에 더 민감해질 가능성이 있다. 본 연구의 참여자들의 경

우 인공지능을 더 자주 활용할수록 안전에 대한 민감성이 더 증가하는 것으로 보이며, 따라서 관련 교육의 필요성을 더욱 체감하는 것으로 짐작된다. 인공지능 활용 빈도가 높아질수록 프라이버시보호에 대한 교육 요구가 점차 줄어드는 점도 흥미로운 결과이다. 프라이버시보호 원칙의 경우는 인공지능을 자주 접하고 활용하는 것만으로도 그 개념을 쉽게 터득하고 관련 전략을 습득할 수 있을지도 모른다. 이상의 결과는 대학원생을 위한 인공지능 윤리 교육은 학습자 특성을 고려한 맞춤형으로 설계할 필요가 있음을 시사한다. 본 연구에서 살펴본 것처럼 평소에 얼마나 인공지능을 자주 활용하는지도 하나의 고려사항이 될 수 있고, 선행연구(김귀식, 신영준, 2021b; 오새내, 김효정, 2023; Farhi et al., 2023; Harwig et al., 2023)에서 언급되었던 성별, 연령, 인공지능에 대한 사전 지식 등을 중요한 학습자 특성으로 고려할 수 있다.

본 연구의 제한점 및 후속 연구를 위한 제언을 하자면 다음과 같다. 첫째, 본 연구의 결과는 서울의 한 종합대학에서 실시한 조사를 기반으로 도출되었다는 점에서 전체 대학원생의 일반적인 교육 요구도로 일반화하기에는 무리가 있겠다. 추후 연구는 여러 지역과 다양한 유형의 대학에 속한 학생들을 대상으로 수행하여 인공지능 윤리기준에 따른 대학원생의 교육 요구가 보편적인 경향을 나타내는지, 아니면 특정 지역이나 대학 유형에 국한된 특수한 요구가 있는지 분석할 필요가 있다.

둘째, 앞서 언급한 제한점과 유사한 맥락으로, 본 연구에서는 여러 학습자 특성 중 특히 인공지능 활용 빈도에 주목하여 교육 요구도 차이를 분석하는 데 그쳤다. 선행 연구에 따르면 인공지능 윤리 인식에는 연령, 성별과 같은 인구·사회학적 특성 역시 영향을 미칠 수 있다(김귀식, 신영준, 2021c; Hartwig et al., 2023). 이점에 착안하여 향후 연구는 연령, 성별, 전공 등 다양한 학습자 특성에 따른 인공지능 윤리 교육 요구에 대한 차이를 보다 면밀히 살펴볼 것을 제안한다.

셋째, 본 연구는 대학원생의 인공지능 윤리의식 함양을 위해 우선적으로 교육이 제공되어야 할 인공지능 윤리 영역과 개념이 무엇인지를 탐색하고 있다는 점에

서 ‘무엇을’ 가르칠 것인지에 대해 주로 천착할 뿐 ‘어떻게’ 가르칠 것인지에 대한 문제는 다루지 않았다. 김은경과 이영준(2023)은 윤리기준을 단순히 제시하고 이를 준수하도록 독려하는 방식으로는 사용자들의 윤리적 의사결정과 행동에 실질적 변화를 이끌어 내기 어렵다는 점을 지적한다. 후속 연구에서는 강의, 토론, 실습, 체험 등 다양한 형태의 교육 프로그램을 설계하고 그 효과성을 검증함으로써 학생들의 인공지능 윤리 역량을 보다 효과적으로 길러줄 수 있는 교수학습 방법이 무엇인지 깊이 탐구해야 할 것이다.

참고문헌

- 고영화, 임춘성 (2021). 인공지능 기술수용과 윤리성 인식의 이용의도에 미치는 영향. *디지털융복합연구*, 19(3), 217-225.
<https://doi.org/10.14400/JDC.2021.19.3.217>
- 과학기술정보통신부. (2020. 11. 27). 사람이 중심이 되는 「국가 인공지능 윤리기준」(안) 공개. <https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=112&pageIndex=&bbsSeqNo=94&nttSeqNo=3179630&searchOpt=ALL&searchTxt=에서> 2023년 12월 22일 인출
- 교육부. (2022. 8. 11). 교육분야 인공지능 윤리원칙 상세 설명자료. <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=W&s=moe&m=020402&opType=N&boardSeq=92297>에서 2023년 12월 20일 인출
- 김귀식, 신영준 (2021a). 인공지능 윤리의식에 관한 횡단적 연구 - 인공지능 활용 기후변화 교육에 대비하여-. *에너지기후변화교육*, 11(1), 27-36.
<https://doi.org/10.22368/ksecce.2021.11.1.27>
- 김귀식, 신영준 (2021b). 초·중·고등학생의 인공지능 윤리의식의 성차 분석. *과학교육연구지*, 45(1), 105-117.
<https://doi.org/10.21796/jse.2021.45.1.105>
- 김귀식, 신영준 (2021c). 인공지능 윤리의식 검사 도구 개발 연구. *인공지능연구 논문지*, 2(1), 1-19.
<https://doi.org/10.52618/aied.2021.2.1.1>
- 김도연, 고영화 (2022). 인공지능기술 윤리성 인식 척도개발 연구. *디지털융복합연구*, 20(1), 71-86.
<https://doi.org/10.14400/JDC.2022.20.1.071>
- 김명주 (2017). 인공지능 윤리의 필요성과 국내외 동향. *정보와통신*, 34(10), 45-54.
- 김민호, 박현지 (2020). 자율주행자동차의 윤리적 고찰과

- 법제정비 방안. 미국헌법연구, 31(1), 1-36.
- 김수정, 서혜민, 이보람 (2023). 어린이집 교실 내 유아와 인공지능 로봇 간의 상호작용 양상과 시간의 흐름에 따른 변화 분석. 인간발달연구, 30(4), 73-87.
<https://doi.org/10.15284/kjhd.2023.30.4.73>
- 김은경, 이영준 (2022). Moral machine을 활용한 인공지능 윤리교육이 초등학생의 인공지능에 대한 인식에 미치는 영향. 컴퓨터교육학회 논문지, 25(3), 1-7.
<https://doi.org/10.32431/kace.2022.25.3.001>
- 김은경, 이영준 (2023). 인공지능 윤리 역량 신장을 위한 인공지능 윤리 딜레마 개발. 컴퓨터교육학회 논문지, 26(5), 31-41.
<https://doi.org/10.32431/kace.2023.26.5.003>
- 김지윤, 김승현, 김귀훈 (2022). 인공지능 융합교육 관심도에 기반한 대학원 전공 필수 교과목 개발 사례연구. 학습자 중심교과교육연구, 22(2), 323-341.
<https://doi.org/10.22251/jlcci.2022.22.2.323>
- 김하민 (2022). 초등교사를 위한 인공지능 윤리 교육 프로그램 개발 연구. 초등도덕교육, 77, 261-283.
<http://doi.org/10.17282/ethics.2022..77.261>
- 배영임, 김유나 (2022). 인공지능의 윤리적 쟁점에 관한 탐색적 연구. 경기: 경기연구원 기본연구.
- 배진아, 이정훈, 조정원 (2022a). 실천력 강화를 위한 인공지능 윤리 교육 모델. 산업융합연구, 20(5), 83-92.
<https://doi.org/10.22678/JIC.2022.20.5.083>
- 배진아, 이정훈, 홍미선, 조정원 (2022b). 중등학생 대상 인공지능 윤리역량 평가척도 개발. 컴퓨터교육학회 논문지, 25(6), 103-117.
<https://doi.org/10.32431/kace.2022.25.6.008>
- 손혜숙 (2022). 인공지능 윤리 의식 함양을 위한 대학 교양교육 방안 연구. 교양학연구, 21, 33-61.
- 손화철 (2023). ChatGPT와 연구윤리. 지식경영연구, 24(3), 1-15.
<https://doi.org/10.15813/kmr.2023.24.3.001>
- 송선영 (2022). 학교 교육과정에서 융복합적 AI윤리교육의 과제 - 교육대학원 AI융합교육전공과 ‘AI 윤리’ 수업 사례를 중심으로. 윤리교육연구, 63, 51-77.
<http://doi.org/10.18850/JEES.2022.63.03>
- 송애리, 고정아 (2023). 인공지능 윤리에 대한 예비 도덕 교사의 인식 유형 연구-Q 방법론 접근. 도덕윤리과교육, (80), 105-136.
<https://doi.org/10.18338/kojmee.2023..80.105>
- 송주영, 전용주 (2023). 책임윤리 기반의 초등학생을 위한 인공지능 윤리교육 프로그램 개발 및 적용. 컴퓨터교육학회 논문지, 26(2), 29-39.
<https://doi.org/10.32431/kace.2023.26.2.004>
- 신주연, 김진선, 정윤경 (2020). 역량기반 교양교육에 대한 I 대학교 학부생의 교육요구도 분석. 교양교육연구, 14(3), 39-54.
<http://doi.org/10.46392/kjge.2020.14.3.39>
- 오선경 (2023). 대학 교양 글쓰기에서의 챗 GPT 활용 사례와 학습자 인식 연구. 교양교육연구, 17(3), 11-23.
<https://doi.org/10.46392/kjge.2023.17.3.11>
- 오새내, 김효정 (2023). 대학생의 ChatGPT 사용 의식과 과제 표절 인식의 관계. 한성어문학, 50, 253-281.
<http://doi.org/10.23114/hsemh.2023.50.253>
- 오하경, 김태성 (2017). 인공지능 비서 이용과 프라이버시 침해에 대한 사용자 인식 연구. 산업과경영, 30(2), 81-95.
- 유수진, 장운재 (2021). 텍스트 마이닝으로 살펴본 대학생들의 인공지능 윤리 인식 연구. 정보교육학회논문지, 25(6), 947-960.
<http://doi.org/10.14352/jkaie.2021.25.6.947>
- 윤진숙 (2022). AI의 지적재산권과 인권 보장-저작권을 중심으로-. 법학연구, 32(2), 483-504.
<https://doi.org/10.21717/ylr.32.2.13>
- 이상욱, 이호영 (2021). 인공지능(AI) 윤리와 법(I): AI 윤리의 쟁점과 거버넌스 연구: 유네스코한국위원회-한국법제연구원 공동연구. 서울: 유네스코위원회.
- 이우권 (2020). 인공지능의 윤리적 사용을 위한 정책대안에 관한 연구. 한국정책논집, 20(1), 1-15.
- 이은경 (2020). 국내외 초·중등학교 인공지능 교육과정 분석. 컴퓨터교육학회 논문지, 23(1), 37-44.
<https://doi.org/10.32431/kace.2020.23.1.003>
- 이재호, 정소윤, 강정석 (2019). 인공지능 기술의 행정분야 활용에 관한 탐색적 연구. 서울: 한국 행정연구원.
- 이주연 (2023). 저작권법과 연구윤리관점에서 본 학술논문과 인공지능의 저작성-최근 ChatGPT를 둘러싼 논의를 중심으로-. 경영법률, 33(4), 127-176.
- 이지형, 변순용 (2023). 인공지능 윤리교육 프로그램의 효과 검증-초등 도덕과 교육 적용을 중심으로. 초등도덕교육, 84, 285-309.
- 이혜령 (2023). 초등학생 인공지능 윤리 인식 척도 개발. 인공지능윤리연구, 2(1), 98-127.
- 임이정, 송재인, 황기연 (2019). 자율주행자동차 윤리 가이드라인 인식 조사 연구. 교통연구, 26(1), 47-62.
<https://doi.org/10.34143/jtr.2019.26.1.47>
- 장연주, 최승윤, 조현기, 김현철 (2022). 초·중등 학생을 위한 인공지능 윤리 교육 프로그램 개발 및 적용. 컴퓨터교육학회 논문지, 25(5), 1-13.
<https://doi.org/10.32431/kace.2022.25.5.001>
- 전수진 (2023). CAI 모델 기반 인공지능 윤리 교육 프로그램 개발 및 효과 분석. 컴퓨터교육학회 논문지, 26(1), 23-30.
<https://doi.org/10.32431/kace.2023.26.1.003>
- 정남숙 (2021). AI 번역기를 활용한 활동이 대학생 영어쓰기 능력과 정의적 태도에 미치는 영향. Multimedia-Assisted Language Learning, 24(1), 134-157.
<https://doi.org/10.15702/MALL.2021.24.1.134>
- 정다이, 박형빈 (2023). 초등학교 AI 윤리교육 프로그램의

- 개발 및 적용. *한국초등교육*, 34(3), 39-54.
- 정수정, 이명성, 김유겸, 임다미 (2023). 텍스트 마이닝을 통해 살펴본 대학원생의 AI 활용에 대한 인식. *교육문화연구*, 29(5), 153-178.
<https://doi.org/10.24159/joec.2023.29.5.153>
- 조대연 (2009). 설문조사를 통한 요구분석에서 우선순위결정 방안 탐색. *교육문제연구*, (35), 165-187.
- 최선유, 김귀훈 (2023). 노벨 엔지니어링을 활용한 초등학교 고학년 대상 인공지능 윤리교육 프로그램 개발. *학습자 중심교과교육연구*, 23(22), 573-585.
<https://doi.org/10.22251/jlcci.2023.23.22.573>
- 한겨레 (2023. 1. 29). “인공지능, 논문 저자 못 된다” 네이처·사이언스의 교통정리. https://www.hani.co.kr/arti/science/science_general/1077316.html에서 2023년 12월 22일 인출.
- 황서이, 김문기 (2019). 국내 인공지능분야 연구동향 분석-토픽모델링과 의미연결망분석을 중심으로. *디지털콘텐츠학회논문지*, 20(9), 1847-1855.
<https://doi.org/10.9728/dcs.2019.20.9.1847>
- 허유선, 이연희, 심지원 (2020). 왜 윤리인가: 현대 인공지능 윤리 논의의 조망, 그 특징과 한계. *인간·환경·미래*, (24), 165-209.
<https://doi.org/10.34162/hefins.2020..24.006>
- Adams, C., Pente, P., Lerner, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. *Computers and Education: Artificial Intelligence*, 4, 100131.
<https://doi.org/10.1016/j.caeai.2023.100131>
- Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, 62, 102433.
<https://doi.org/10.1016/j.ijinfomgt.2021.102433>
- Bogina, V., Hartman, A., Kuflik, T., & Shulner-Tal, A. (2021). Educating software and AI stakeholders about algorithmic fairness, accountability, transparency and ethics. *International Journal of Artificial Intelligence in Education*, 32, 808-833.
<https://doi.org/10.1007/s40593-021-00248-0>
- Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1, 61-65.
<https://doi.org/10.1007/s43681-020-00002-7>
- Borich, G. D. (1980). A needs assessment model for conducting follow-up studies. *Journal of Teacher Education*, 31(3), 39-42.
<https://doi.org/10.1177/002248718003100310>
- Bouhouita-Guermech, S., Gogognon, P., & Belisle-Pipon, J. C. (2023). Specific challenges posed by artificial intelligence in research ethics. *Frontiers in Artificial Intelligence*, 6, 1-17.
<https://doi.org/10.3389/frai.2023.1149082>
- Chubb, J., Cowling, P., & Reed, D. (2022). Speeding up to keep up: Exploring the use of AI in the research process. *AI & SOCIETY: Journal of Knowledge, Culture and Communication*, 37(4), 1439-1457.
<https://doi.org/10.1007/s00146-021-01259-0>
- Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Al-shami, S. A., & Slamene, R. (2023). Analyzing the students' views, concerns, and perceived ethics about chat GPT usage. *Computers and Education: Artificial Intelligence*, 5, 100180.
<https://doi.org/10.1016/j.caeai.2023.100180>
- Fung, P., & Etienne, H. (2023). Confucius, cyberpunk and Mr. Science: Comparing AI ethics principles between China and the EU. *AI and Ethics*, 3(2), 505-511.
<https://doi.org/10.1007/s43681-022-00180-6>
- Gal, D. (2019). Perspectives and approaches in AI ethics: East Asia. In *Oxford handbook of Ethics of artificial intelligence* (pp. 606-624). England: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780190067397.013.39>
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99-120.
<https://doi.org/10.1007/s11023-020-09517-8>
- Hartwig, T., Ikkatai, Y., Takanashi, N., & Yokoyama, H. M. (2023). Artificial intelligence ELSI score for science and technology: A comparison between Japan and the US. *AI & SOCIETY*, 38(4), 1609-1626.
<https://doi.org/10.1007/s00146-021-01323-9>
- Hosseini, M., Resnik, D. B., & Holmes, K. (2023). The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*, 19(4), 449-465.
<https://doi.org/10.1177/17470161231180449>
- Lin, X. F., Wang, Z., Zhou, W., Luo, G., Hwang, G. J., Zhou, Y., Wang, J., Hu, Q., Li, W., & Liang, Z. M. (2023). Technological support to foster students' artificial intelligence ethics: An augmented reality-based contextualized dilemma discussion approach. *Computers & Education*, 201, 104813.
<https://doi.org/10.1016/j.compedu.2023.104813>
- Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education - Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39.
<https://doi.org/10.1186/s41239-019-0171-0>