



예비교사의 AI 윤리 인식 차이 분석: 전공-성별-연령-AI 활용 경험을 중심으로

Unkyoung Maeng (Ajou University)
Ho Kyoung Ko (Ajou University)

Received: 22 February 2026
Revised: 16 March 2026
Accepted: 31 March 2026

Maeng, Unkyoung & Ko, Ho Kyoung. (2026). Exploring pre-service teachers' perceptions of AI ethics: Differences by major, gender, and AI usage experience. *Modern English Education*, 27, 160-178.

Keywords

AI ethics, pre-service teachers, awareness, human-centeredness, fairness, transparency, safety, accountability
AI 윤리, 예비교사, 인식, 인간중심성, 공정성, 투명성, 안전성, 책임성

Unkyoung Maeng (First author)
Professor
Graduate School of Education/English Education
Ajou University
my5329@ajou.ac.kr

Ho Kyoung Ko
(Corresponding author)
Professor
Graduate School of Education
Ajou University
Email: kohoh@ajou.ac.kr

*본 연구는 2023학년도
아주대학교 일반연구비에
의하여 연구되었음.

Abstract

The present study examines pre-service teachers' awareness of AI ethics in educational contexts, focusing on differences by major, gender, age, and prior experience using AI. A total of 283 pre-service teachers from universities and graduate schools of education in the Seoul-Gyeonggi area participated in online and paper-based surveys. Based on major international and national AI ethics frameworks, the survey measured five factors—human-centeredness, fairness, transparency, safety, and accountability—using 24 items. Descriptive statistics, repeated-measures ANOVA, independent-samples t-tests, and one-way ANOVAs were conducted. Overall, AI ethics awareness was moderate ($M = 3.18$, $SD = 0.55$), with the highest ratings for human-centeredness ($M = 3.81$) and the lowest for safety ($M = 2.81$). Male respondents reported significantly higher awareness than females in fairness, transparency, safety, and accountability, whereas no gender difference was found in human-centeredness. Age differences were most evident in fairness and transparency, with participants in their twenties scoring higher than older groups. Language-education majors demonstrated higher awareness across all factors than natural-science education and non-teaching majors. No significant differences were found based on experience with AI use or related coursework. These findings suggest the need for explicit and structured AI ethics and safety education in teacher training programs.

서론

인공지능(AI) 기술의 발전은 교육을 포함한 사회 전반에 걸쳐 혁신을 일으키고 있다. AI가 특정 과업을 수행하는 좁은 인공지능(Artificial Narrow Intelligence, ANI) 단계를 넘어, 인간이 수행할 수 있는 대부분의 인지 과제를 이해하고 학습하며 다양한 맥락에서 인간 수준의 과업을 수행할 수 있는 범용 인공지능(Artificial General Intelligence, AGI) 단계로 확장될 가능성이 있다는 점이 학문적으로 논의되고 있다(Raman et al., 2025; Saghiri et al., 2022). AGI는 추론, 학습, 문제 해결, 적응 능력을 포괄하는 인간 수준의 지능을 의미하며, 이러한 범용성은 기존 AI와 질적으로 다른 사회적·윤리적 함의를 지닌다(Raman et al., 2025; Saghiri et al., 2022). 더 나아가 일부 연구자들은 AGI가 일정 임계점을 넘어 인간의 최고 인지 능력을 거의 모든 영역에서 능가하는 초지능(Artificial Superintelligence, ASI)으로 발전할 가능성도 이론적으로 제기하고 있다. ASI에 도래할 경우 인간의 통제 가능성, 가치, 사회 제도의 안정성에 중대한 문제가 발생할 수 있음을 경고하고 있다(Gulchenko, 2024; Peters, 2025). 이러한 AGI 및 ASI 논의는 아직 실현 단계에 이르지 않았으나, AI 기술이 점차 자율적 판단과 복합적 의사결정을 수행하는 방향으로 발전하고 있다는 점에서 교육과 사회는 장기적 관점에서 윤리적, 제도적 대비를 할 필요가 있다(Raman et al., 2025). 특히 미래 세대의 학습과 가치 형성에 중요한 영향을 미칠 교사는 이러한 기술 발전의 추이와 이와 동반될 잠재적 윤리 문제를 비판적으로 이해할 필요가 있다.

AI 지능 수준의 확장은 실제 AI 기술 발전에도 확인되고 있는데, 최근 AI는 단순한 정보를 처리하거나 예측하는 수준을 넘어 이미지 생성, 음성 인식, 의미 추론 등 인간의 창작과 판단 영역에까지 확장되는 방향으로 빠르게 발전하고 있다(Brown et al., 2020; Devlin et al., 2019; Ramesh et al., 2022). 특히 생성형 AI는 고품질 텍스트, 이미지, 음성을 손쉽게 생성할 수 있게 됨에 따라 정보 생산의 진입 장벽을 크게 낮추었으며, 이는 교육 현장에 학습 지원이라는 긍정적 활용성을 제공하는 동시에 윤리적 문제도 함께 야기시키고 있다. 여러 선행연구에서 생성형 AI의 확산으로 인해 과제 대리 작성, 평가의 공정성 훼손, 학습자의 비판적 사고 약화, 딥 페이크 기술을 활용한 허위 정보 생산과 인격 침해와 같은 문제가 교육 및 사회의 문제로 부상하고 있음을 보고하고 있다(Chesney & Citron, 2019; Cotton et al., 2024; Jang et al., 2022; Oh et al., 2023). 이와 같은 문제는 AI 기술 자체의 발달에서 비롯된 것이지만 기술적 통제만으로는 문제를 해결하는 데 한계가 있어 교육을 통한 윤리적 판단 능력 함양이 필요하다는 논의로 이어지고 있다.

특히 학교는 학생들이 AI를 처음으로 체계적으로 접하고 활용하는 규범을 형성할 수 있는 중요한 공간으로, 학생들이 윤리적 판단에 기반하여 인간 중심으로 AI를 비판적으로 활용할 수 있는 디지털 역량을 함양하도록 교육해야 하는 책무가 있다. 이러한 측면에서 교사의 역할은 매우 중요하다(Song & Ko, 2023). 교사는 AI를 단순한 학습 도구로 소개하는 것을 넘어, 그 한계와 위험, 윤리적 사용 원칙을 함께 지도해야 한다. 따라서 교사의 AI 윤리 인식 수준은 교육현장의 AI 활용 방식과 직접적으로 연결되어 있다(Selwyn, 2019). 교육 분야에서 AI 윤리의 중요성을 인식하고 공정성, 책임성, 투명성, 개인정보 보호와 같은 핵심 윤리 원칙을 중심으로 여러 선행연구가 이루어져 왔으며 이를 통해 이론적 논의와 정책적 가이드라인이 제시되어 왔다(Fjeld et al., 2020; Jobin et al., 2019; UNESCO, 2021, 2023, 2024). 국내의 AI 윤리 관련 연구들도 AI 윤리 기준과 원칙 정립을 위한 이론 중심 연구, 학생의 윤리적 사고력과 실천 능력 향상을 위한 교육과정 및 프로그램 개발 연구, 교수학습 전략, 정책 방향 및 교사연수 체계에 관한 제안 중심으로 이루어져 왔다(Kim & Park, 2023; C. Kim, 2024; Park & Yun, 2025).

이와 같이 기존 연구들은 AI 윤리 원칙 제시나 교육 프로그램 개발에 상대적으로 집중해 왔으며, 미래 교육 현장을 담당할 예비교사들이 생성형 AI 기술 발전과 더불어 발생하는 윤리적 문제를 어떻게 인식하고 해석하는지에 대한 실증적 연구는 미흡한 상태이다(M-S. Kang, 2024; Song & Ko, 2023). 일부 예비교사를 대상으로 한 AI 관련 인식 연구를 살펴보면, 예비교사들이 AI 교육이나 AI 기술에 대해 대체적으로 긍정적인 태도를 갖고 있으면서도 윤리적 문제에 대해서는 불안과 걱정을 하고 있는 것으로 나타났다(M-S. Kang, 2024). Selwyn(2019)은 이러한 교사 인식이 AI를 기술적 도구로만 이해하는 수준에 머무를 경우 교육의 윤리적 문제가 간과될 수 있음을 비판적으로 제시하였다. 예비 도덕 교사를 대상으로 AI 윤리 인식 유형을 탐색한 연구에서는, AI의 공정성, 책임성, 통제 가능성에 대한 인식이 예비교사 집단 내에서도 상이하게 나타났다고 보고하면서 성공적인 AI 윤리 교육이 실현되기 위해서는 교사 양성 과정에서부터 인식의 균형을 맞추는 노력이 필요하다고 제안하고 있다(Song & Ko, 2023). 또한 Pei 외 2인(2025) 연구에 따르면 예비교사들은 AI 기술 메커니즘의 이해 수준이 상대적으로 낮으며 이러한 이해 격차가 AI 윤리 인식, 교육적 활용 가능성에 영향을 미친다는 것이다. 이와 같은 AI 윤리에 대한 예비교사들의 인식 연구에서는 예비교사의 AI 활용 태도나 일반적 윤리 인식을 중심으로 분석이 이루어졌으며,

실제 교육현장에서 활용되는 AI 시스템을 전제로 한 윤리적 판단, 즉 평가의 공정성, 알고리즘 편향, 학생 개인정보 보호, AI 의사결정의 투명성과 같은 구체적인 윤리 쟁점을 종합적으로 분석한 실증 연구는 제한적인 수준에 머물러 있다. 또한 예비교사의 전공 배경에 따른 인식 차이를 충분히 고려하지 못한 한계도 존재한다. 이에 본 연구에서는 예비교사들이 미래 교실에서 직접 마주하게 될 AI 기반 교육 시스템을 상정하고, AI 윤리의 핵심 원칙이 교육 맥락에서 전공 계열에 따라 어떻게 인식되는지를 설문을 통해 실증적으로 탐색하고자 한다. 이를 통해 향후 학교 현장에서 실질적으로 적용될 수 있는 AI 윤리 교육의 내용과 방향을 설계하고 교사 교육 과정을 개선하는데 중요한 기초자료를 제공하고자 한다. 이를 위한 구체적인 연구질문은 다음과 같다.

첫째, 예비교사들이 실제 활용하고 있는 AI 도구 및 플랫폼의 유형과 수강한 AI 관련 강좌의 유형은 무엇이며 이러한 활용 양상은 성별, 연령, 전공별로 어떠한 차이를 보이는가?

둘째, 예비교사들의 AI 윤리 인식 수준은 어떠하며, 개인적·교육적 배경 변인(성별, 연령, 전공, AI 사용 경험)에 따라 차이를 보이는가?

이론적 배경 및 선행연구

AI 윤리의 핵심요소

AI 기술의 급격한 발전으로 AI가 자율적인 판단과 예측을 하는 기능을 수행함에 따라 제조, 의료, 교통, 환경, 교육 등 다양한 영역으로 AI 활용이 확산되고 있다. 동시에 기술의 오남용, 알고리즘에 의한 차별, 데이터 편향성, 프라이버시 침해 등과 같은 윤리 문제가 부각되면서 기술적 효율성과 더불어 사회적 책임성과 윤리성이 중요한 교육의 의제로 부상하였다(Oh et al., 2023; Raman et al., 2025). AI 윤리는 사회에서 AI를 책임감 있게 사용할 수 있는 기준(framework)과 지침(guideline)을 마련하고 기술, 철학, 법적 측면에서 윤리적 연구와 실천을 종합적으로 안내하는 체계이다(Winfield & Jirotko, 2018). 국내외에서 AI 설계, 개발, 배포, 사용되는 전 과정에서 발생할 수 있는 윤리적 문제를 규범적으로 고찰하고 인간의 존엄성, 사회 정의, 공공의 이익을 지키기 위한 기준을 마련하는 노력을 지속적으로 수행하고 있다(Hoe et al., 2022; Jun, 2023; OECD, 2019, 2023; UNESCO, 2019, 2023).

UNESCO는 AI 윤리를 인간의 존엄성과 인권을 보호하기 위한 규범적 틀로 규정하며, 인간중심성(human agency), 공정성(equity), 포용성(inclusion)을 핵심 가치로 제시하였다(UNESCO, 2021). 인간중심성(human agency)이란 교육 현장에서 AI가 교사의 의사결정을 보조하는 도구로 활용되어야 하며, 학습자의 주체적인 학습 능력을 저해해서는 안 된다는 점을 의미한다. 따라서 AI 기술이 인간(교사와 학습자)의 주도성과 자율적 선택권을 침해하지 않고, 교육적 목적을 달성하기 위한 능동적 수단으로 기능해야 함을 의미한다. 두 번째, 공정성(equity)은 단순한 기회의 균등을 넘어, AI 알고리즘이 특정 학습자 집단에게 편향된 결과를 제공하지 않도록 보장하는 것이다. 개별 학습자의 배경(전공, 성별, 연령 등)에 관계없이 각자의 필요에 맞는 맞춤형 교육 자원이 배분될 수 있도록 알고리즘적 정의를 실현하는 데 정당성이 있다. 세 번째, 포용성(inclusion)은 AI 교육 도구가 신체적, 학습적 취약계층을 포함한 모든 학습자에게 접근 가능해야 함을 의미하는 것으로 디지털 격차가 교육 격차로 이어지지 않도록 보편적 학습 설계(UDL)의 관점에서 AI 기술이 통합적 교육 환경을 조성하는 데 기여해야 함을 주장한다. 이러한 가치는 이후 AI 교육 가이드라인에서도 반복적으로 강조되며, 교육 맥락에서 학습자의 주체성 보장과 형평성, 포용적 접근의 원칙으로 확장되었다(UNESCO, 2023). 또한 이러한 가치는 이후 OECD, EU, 한국 및 세계 여러 나라의 AI 윤리 기준 및 AI 시민성 관련 연구를 통해 보다 구체적인 실천 원칙과 교육적 요소로 재구성되면서 AI 기술이 인간과 사회에 미치는 영향에 대한 윤리적 성찰을 공통적으로 강조하고 있다. 구체적으로 UNESCO(2021)는 AI 윤리를 인간의 존엄성과 인권을 보호하기 위한 규범적 틀로 규정하며, 차별과 배제 방지, 투명한 의사결정, 책임 있는 거버넌스를 핵심 원칙으로 제시하였다. OECD(2019)는 신뢰할 수 있는 AI를 구현하기 위한 조건으로 AI 개발과 활용에 있어서 인간 중심 가치(human-centered values), 공정성(fairness), 투명성(transparency), 안전성(safety), 책임성(accountability) 등을 핵심 원칙으로 제안하였다. EU는 기술적, 윤리적, 법적 신뢰성을 모두 갖춘 AI를 구축하기 위해 인간의 대리성과 감시 가능성, 기술적 견고성, 프라이버시 보호, 책임성과 설명 가능성과 같은 원칙을 강조하였다(EU, 2019).

국내연구와 정책 문헌에서도 이와 유사한 윤리적 요소들을 제시하고 있다. Ministry of Science and ICT (2020)의

‘사람이 중심이 되는 인공지능(AI) 윤리기준’과 Ministry of Education(2022)의 ‘교육분야 인공지능 윤리원칙’에서 인간 중심성, 공정성, 안전성, 책임성을 핵심 가치로 설정하고, 교육 맥락에서는 학습자의 권리보호와 윤리적 판단 역량을 강조하고 있다. 또한 AI 시민성 및 AI 윤리 교육관련 선행연구에서는 이러한 윤리 원칙을 시민의 태도와 역량 차원에서 재해석하고 있다. Bak(2023)은 UNESCO, EU와 한국의 AI 윤리기준을 분석하여 AI 시민성 교육의 핵심 가치를 주체성, 안전성, 성찰성, 책임성, 공정성, 공공성, 협력성, 포용성으로 도출하였다.

국내외 AI 윤리 연구를 종합하여 보면 AI 윤리는 단일한 규범이 아니라 여러 핵심 요소가 상호 유기적으로 결합된 개념임을 알 수 있다. 표현 방식과 적용 맥락에 차이가 있으나 인간중심성, 공정성, 투명성, 안전성, 책임성은 AI 윤리를 구성하는 공통된 핵심 요소로 수렴될 수 있다. 먼저 인간중심성은 AI가 인간의 존엄성과 자율성을 침해하지 않고 인간의 의사결정과 성장을 보조해야 한다는 원칙으로, 모든 AI 윤리 논의의 출발점이라고 볼 수 있다. 둘째, 공정성은 데이터 편향이나 알고리즘 설계로 인해 특정 집단이 차별받지 않아야 한다는 윤리적 요구로써 사회적 불평등의 재생산을 방지하는 핵심적인 역할을 한다. 셋째, 투명성은 AI 시스템의 작동 원리, 데이터 사용 방식, 의사결정과정을 명확하게 공개하고 설명 가능하도록 하며, 사용자가 작동 원리를 이해하고 신뢰할 수 있어야 한다는 원칙으로 이용자가 AI의 판단을 비판적으로 검토할 수 있는 전체 조건이 된다. 넷째, 안전성은 AI 기술이 물리적, 심리적, 사회적 위험을 최소화하고 신뢰할 수 있는 방식으로 작동해야 함을 시사한다. 마지막으로, 책임성은 AI의 개발, 운영, 활용 과정에서 발생하는 결과에 대한 명확한 책임 주체를 설정하고 윤리적 책무를 이행해야 한다는 원칙으로, AI 거버넌스의 핵심요소로 작용한다.

AI 윤리교육의 쟁점

AI 기술 발달은 교육 현장에도 혁신적인 변화를 불러 일으키고 있다. AI 도입으로 교수·학습 방식의 혁신 및 교사 업무 경감 및 효율성을 높이는 동시에 새로운 윤리적 쟁점도 야기하고 있다. 특히 초·중·고 및 대학교육의 실제 수업, 평가, 교육과정 운영과정에서 발생하는 문제로 인해 추상적인 윤리 개념을 현실화하는 계기가 되고 있다. 쟁점은 사회적 영향, 데이터 윤리, 알고리즘 윤리로 분류할 수 있으며, 각 영역은 학생, 교사, 교육과정 연구자 등 교육 구성원 모두의 책임과 연결되어 있다.

사회적 영향 영역에서의 주요한 쟁점은 AI 도구 사용으로 인한 학생의 주체성 약화와 기술 의존으로 인한 비판적 사고력 저하이다. ChatGPT와 같은 AI 도구가 대학교 학생들의 전략적 쓰기에 도움을 줄 수 있으나 비판적 사고 저하, 표절과 같은 윤리적 문제를 초래할 수 있음을 여러 연구에서 보고하고 있다(Awal, 2024; Y. Lee, 2023). 특히 학생들의 비판적 평가 능력이 부족할 때 학습과정에서 부정적인 영향을 미칠 수 있음을 강조하는 연구도 있다(Mogavi et al., 2024). 또한 AI에 과도하게 의존할 경우 학습자의 자율성, 비판적 사고력이 약화되어 학습에 부정적인 태도가 형성될 수도 있음을 보고하는 연구도 다수 있다(Maeng, 2025; Wang & Chen, 2025). 대학생들의 AI 과의존이 비판적 사고, 분석적 추론, 의사결정과 같은 핵심 인지 능력에 부정적인 영향을 줄 수 있으며, 태도에도 부정적인 영향을 줄 수 있음을 보고하고 있다(Zhai et al., 2024). 초등학생의 경우도 AI를 작문 활동에 활용하는 과정에서 학생들이 스스로 표현하거나 판단하는 경험이 축소될 우려가 제기되었으며(Kim & Park, 2023), 중·고등학교 현장에서도 AI 활용은 학습 효율성을 높일 수 있으나, 학습자의 자율성, 비판적 사고, 학습동기 및 책임감 저하와 같은 부정적인 영향을 미칠 수 있음을 여러 선행연구에서 제시하고 있다(Akgun & Greenhow, 2022; Gouseti et al., 2025; Huang et al., 2026; Lodhi & Lodhi, 2025). AI 시스템 정답 제시, 자동 피드백 제시, 의사결정 지원 등의 기능에 과도하게 학생들이 의존할 때 학생들이 스스로 사고하고 탐색하는 기회가 감소되고 그 결과 특히 초·중등 학생의 경우 AI를 무비판적으로 수용하는 경향이 강화될 수 있다(Akgun & Greenhow, 2022; Huang et al., 2026). 그 외에도 이러한 AI 학습 지원으로 인해 학생들이 학습 성취의 주체가 자신이 아닌 AI 시스템으로 인식할 위험이 있으며, 학습 실패 경험을 통한 성찰 기회가 감소되어 내적 동기와 학습 책임감이 저하될 수 있음을 보고하였다(Akgun & Greenhow, 2022; Gouseti et al., 2025).

데이터 윤리 측면에서 학생 데이터 수집 및 활용에 대한 인식 부족이 핵심 쟁점으로 나타났다. 여러 선행연구에서 초·중·고 교육에서 AI 활용은 학생 개인정보의 과도한 수집, 데이터 소유권과 통제권 불명확성, 데이터 기반 평가의 투명성 부족, 학습환경의 감시화와 같은 데이터 윤리 문제를 동반할 수 있음을 밝히고 있다(Akgun & Greenhow, 2022; Gouseti et al., 2025; D.H. Kim, 2025; Ma et al., 2025). Huang 외 3인(2026)은 초·중·고 학생 데이터가 교육 목적을 넘어 상업적, 기술적 개발에 재사용될 위험을 보고하고 있으며, 또 다른 연구에서는 AI 피드백 및 평가의 결과가 어떤 데이터가 어떻게 활용되어 그러한 결과를 제공하는지에 대한 학생과 교사의 이해가 부재한 경우가 많다고

보고하고 있다(Akgun & Greenhow, 2022; Gouseti et al., 2025). D. H. Kim (2025)은 중등 교사의 약 60%가 AI 기반 학습 시스템이 학생 데이터를 어떻게 수집하고 활용하는지를 정확히 모른다고 응답하였다고 보고하였다. 이처럼 개인정보 보호 및 알고리즘 투명성에 대한 교육계의 감수성 부족이 문제로 대두되고 있다(Bae et al., 2022). 그 외에도 안면 인식 기술이나 감성 AI를 사용하여 학생들의 표정, 지루함, 좌절감 등과 같은 자료를 수집하고 분석하는 것이 학생들의 내밀한 감정까지 감시하는 것이라는 우려를 낳을 수 있다고 보고하고 있는 연구(Huang et al., 2026)와 AI 시스템 사용시 과도한 양의 개인 정보가 노출되는 문제, 사용자가 인지하지 못하는 사이 언어, 인종적 정체성, 인적 정보, 위치 정보와 같은 메타데이터가 공유될 수 있는 문제, 사용자가 자기가 공유하는 정보의 범위를 제대로 이해하지 못한 채 동의하게 되는 문제, 공교육 시스템에서 AI 알고리즘 사용이 요구될 경우 학생과 학부모는 선택의 여지없이 해당 AI 알고리즘을 사용해야 하는 선택권 부재와 같은 문제가 있다고 보고한 연구가 있다(Akgun & Greenhow, 2022). AI가 생성한 가짜 뉴스나 딥 페이크 영상이 미디어에 유포되어 대중을 오도하는 허위 정보 및 딥 페이크 문제의 심각성도 보고하고 있다(Ma et al., 2025).

교육 현장에서 AI 알고리즘 편향은 학습 데이터의 대표성 부족으로 불공정한 판단, 소수 집단 학생의 낙인화, 편향의 불가시성, 언어 및 문화적 차이에 따른 평가 불이익, 책임주체의 불명확성과 같은 윤리 쟁점을 야기시킨다(Akgun & Greenhow, 2022; Gouseti et al., 2025; Lodhi & Lodhi, 2025). 알고리즘은 객관적이거나 가치 중립적인 도구로 보일 수 있지만, 실제로는 개발자의 가치관과 기존 사회의 구조적 편향을 반영한다(Akgun & Greenhow, 2022). 예를 들어 AI 번역 시스템에서는 사회적 고정관념을 반영하여 의사는 남성형으로 간호사는 여성형으로 변환하는 등 사례가 보고되었으며, 안면 인식 소프트웨어 인종적 편향 문제가 발생하기도 한다. 또한 데이터의 다양성 부족으로 자연어 처리 기반 평가 도구가 학생들이 주류가 아닌 비 주류의 작문 스타일이나 방언을 사용할 때 불리한 점수를 주는 사례도 있다(Lodhi & Lodhi, 2025). 또한 교사가 AI가 생성한 데이터를 무비판적으로 받아들이게 되면 교사의 평가 기술이 약화되고 알고리즘의 판단이 강화되는 결과를 초래할 수 있다.

그 외에도 Gouseti 외 3인(2025)은 교사들이 ChatGPT와 같은 생성형 AI 도구로 인한 학생들의 표절이나 인지적 게으름(cognitive laziness)을 가장 큰 윤리적 위험 중 하나로 인식하고 불안함을 느끼며, AI 알고리즘의 편향에 대한 이해 부족으로 학생들이 갖고 있는 오 개념을 파악하지 못해 적절한 교육을 제공하지 못하거나, AI 기술적 이해 부족으로 인해 AI 윤리를 가르칠 준비가 되어 있지 않아 무력감을 호소하고 있다고 보고하였다. 최근에는 학생들이 과제 수행 시 생성형 AI를 사용할 때 발생하는 표절이나 윤리적 책임과 같은 학문적 정직성의 문제가 교육 현장에 중요한 이슈로 떠오르고 있다(S. Kang, 2025; Ma et al., 2025; Oh et al., 2023).

이처럼 교육 현장에서 발생하는 AI 윤리 문제는 단순한 기술활용의 문제가 아니라 미래 시민이 마주하게 될 도덕적 판단과 가치 선택의 문제로 확장된다. 따라서 교사와 학생이 이러한 윤리적 쟁점과 잠재적 위험을 비판적으로 인식하고 성찰할 수 있도록 지원하는 체계적인 AI 윤리 교육이 요구된다(Akgun & Greenhow, 2022; Jeong & Park, 2023; Jun, 2023). 이는 AI 기술을 사용하는 능력을 넘어 책임 있게 판단하고 활용하는 역량을 기르는 교육으로 이어져야 함을 시사한다. 더 나아가서 AI 시대 학생들의 올바른 윤리관 형성을 위해서는 교사교육에서 AI 윤리 교육을 강화해야 함을 시사한다.

AI 윤리에 대한 인식연구

AI 윤리에 대한 선행 연구는 주로 대학생과 일반 학습자를 대상으로 AI 기술이 야기하는 윤리적 문제에 대한 인식 수준과 도덕성 민감성을 탐색하는데 초점을 두고 있다. 국내 연구에서는 생성형 AI 활용과 관련하여 출처 표기, 표절, 윤리 준수 인식 등을 분석한 연구가 보고되었으며, 전반적으로 학습자들의 윤리 인식 수준이 중간 정도에 머물러 있음을 지적하고 있다. 또한 생성형 AI 활용 경험이 윤리 인식과 일정 관련성이 있는 것으로 나타났으나 교육 맥락에서의 구체적인 윤리 원칙을 종합적으로 분석하는 데는 한계가 있다(S. Kang, 2025). AI 윤리 교육을 다루는 연구들은 ChatGPT와 같은 특정 도구를 중심으로 윤리적 문제를 도출하고 교육 방향을 제안하는 데 중점을 두고 있다.

M. Ryu 외 2인(2025)은 ChatGPT 사용 과정에서 나타난 공정성, 개인정보 보호, 정보 신뢰성 문제를 분석하며 AI 윤리 교육의 필요성을 강조하였다. 그러나 이러한 연구들은 주로 특정 도구나 사례에 초점을 두고 있어 교육현장에서 활용되는 AI 시스템 전반을 포괄적으로 다루는 데에는 한계를 보인다. Ghotbi와 Ho(2021)는 일본 및 다국적 대학생을 대상으로 AI 윤리 인식을 조사한 결과 개인정보 침해, 알고리즘 편향, 사회적 통제와 같은 구조적 윤리 문제에 대한 인식 수준이 낮다고 보고하고 있다. 최근 생성형 AI의 확산 이후 AI 윤리를 교육 맥락에서

탐색하려는 연구가 증가하고 있다. Yang 외 3인(2025)은 대학생을 대상으로 생성형 AI 윤리에 대한 인식을 분석한 결과, 대학생들은 AI 활용에서 인간의 자율성과 선택권 보장에는 긍정적으로 응답하였으나 정의성, 공정성 측면에는 상대적으로 AI에 대한 신뢰 수준이 낮다고 응답하였다. 이와 유사하게 요르단 대학생을 대상으로 AI 활용과 윤리적 우려를 조사한 연구에서는 다수의 학생이 AI를 과제 수행과 학습 지원에 활용하고 있으나, 윤리적 가이드라인에 대한 인식과 제도적 지원에 대한 인식은 제한적이었다고 보고하고 있다(Alnsour et al., 2025).

이처럼 일반 대학생을 대상으로 한 AI 윤리 인식 연구는 축적되고 있음에도 불구하고, 예비교사를 대상으로 한 연구는 상대적으로 제한적인 수준에 머물러 있다. 예비교사는 향후 학교 현장에서 AI 기술을 교육적으로 활용하고, 학생들의 AI 사용을 지도해야 하는 핵심 주체라는 점에서 예비교사들의 AI 윤리에 대한 교육 및 인식 연구를 지속적으로 수행할 필요가 있다. 인도의 예비교사를 대상으로 AI에 대한 태도를 분석한 연구에서 예비교사들이 AI에 대해 긍정적인 태도를 보이면서도 윤리적 우려와 기술 활용에 대한 불안이 공존하고 있음을 보고하고 있다(Konika et al., 2025). 특히 이 연구는 AI에 대한 예비교사의 태도가 연령, 전공계열, 지역성에 따라 유의미한 차이가 있음을 밝히고 교사 교육 단계에서의 맞춤형 AI 윤리 교육의 필요성을 시사하였다. Pei 외 2인(2025)은 예비교사들이 AI 기술의 작동 원리와 메커니즘에 대한 이해 수준이 전반적으로 낮았으며, 기술적 이해 부족이 AI 윤리에 대한 인식과 교육적 활용 가능성에 영향을 미치고 있음을 언급하였다. 국내의 경우 M. S. Kang(2024)은 인터뷰를 통해 AI 교육과 AI 교사에 대한 인식을 심층적으로 분석한 결과, AI를 교육의 효율성을 높이는 보조 도구로 인식하는 유형, AI가 교사를 대체할 수 없으며 정의적 측면에서 한계를 보인다고 인식하는 유형, AI 도입을 흥미로운 변화이자 새로운 교육적 기회로 인식하는 유형으로 구분된다고 보고하면서 인간 다음을 지키고 윤리적 판단력을 기르는 방향으로 AI 윤리 교육이 실시되어야 함을 시사하였다. Song과 Ko(2023)는 예비 도덕 교사를 대상으로 AI 윤리 인식 유형을 분석한 결과, AI 공정성, 책임성, 통제 가능성에 대한 인식이 예비교사 집단내에서 인공지능의 위험성 경계형, 인공지능 윤리의 선결 조건 강조형, 프라이버시 보호형, 인공지능과 공존 추구형으로 각기 다른 인식유형으로 나타났으며, 이러한 결과는 성공적인 AI 윤리 교육을 위해서 교사 양성 과정에서부터 인식의 균형을 맞추는 노력이 필요함을 시사한다고 언급하였다. 즉 예비 교사들이 자신의 인식 유형을 성찰하고 편향되지 않은 교육적 의사결정을 내릴 수 있는 메타적 교육이 교사 양성프로그램 교육과정에 포함될 필요성을 언급하고 있다.

AI 윤리 인식이 개인적, 교육적 배경 변인에 따라 다르게 나타날 수 있음을 보고하고 있는 연구도 다수 보고되고 있다. 성별과 관련하여 대부분의 연구에서 여학생이 남학생 보다 AI 윤리, 특히 공정성, 프라이버시 등에 대해 더 민감하고 높은 윤리 인식 수준을 보이고 있다고 보고하고 있다(Asiksoy, 2024; Jang et al., 2022; Mihoc et al., 2025; Ryu et al., 2025; Wang et al., 2025). 여학생은 남학생 보다 AI 공정성, 편향성, 차별 문제에 대해 더 민감하게 반응하고 우려하는 것으로 나타났다(Asiksoy, 2024; Jang et al., 2022; Ryu et al., 2025). 또한 여학생은 개인정보 보호와 AI가 해를 미치지 않아야 한다는 무해성 측면에서도 남학생 보다 더 엄격한 태도를 보였으나 투명성, 책임성 차원에서 성별간 유의미한 차이가 발견되지 않았다고 보고하고 있는 연구도 있다(Asiksoy, 2024; Jang et al., 2022). Mihoc 외 2인(2025)은 컴퓨터 과학 전공 대학생을 대상으로 한 연구에서 여학생이 남학생에 비해 AI 사회적 영향과 윤리적 고려에 대해 더 민감하게 반응하는 경향을 보였다고 보고하였으며, Alnsour 외 5인(2025)은 여학생이 남학생 보다 AI 사용 지식은 다소 낮으나 전반적인 윤리적 인식 수준은 더 높다고 보고하고 있다. 반면에 Konika 외 2인(2025)은 성별에 따른 AI 윤리 인식이나 태도에 통계적으로 유의미한 차이가 나타나지 않았다고 보고하였다. 연령(학년별)에 따른 차이를 살펴본 연구는 연구 맥락에 따라 결과가 상이하게 나타났다. 나이가 많을수록, 학위 과정이 높을수록 AI 윤리 지식과 인식 수준이 증가하는 경향을 보인다는 연구 결과(Alnsour et al., 2025), 고학년일수록 AI 활용 시 발생할 수 있는 부정행위 문제에 더 민감하게 반응한다는 연구 결과(Yang et al., 2025), 연령에 따른 윤리 인식의 차이가 없다는 연구 결과(S. Kang, 2025)가 있다. 특히 예비교원을 대상으로 인식을 조사한 연구에서는 연령이 낮을수록 AI 공정성과 의존성 문제에 대해 더 높은 우려를 나타낸다는 연구 결과(Ryu et al., 2025)와 디지털 네이티브 세대인 30세 미만의 예비 교원이 30세 이상 집단 보다 AI에 대해 통계적으로 유의미하게 더 긍정적인 태도를 보였다(Konika et al., 2025)고 보고하고 있다.

전공 계열에 따른 AI 윤리 인식 차이를 살펴본 선행연구에서는 다소 상반된 연구 결과를 보고하고 있다. 전공 특성에 따라 민감하게 반응하는 윤리 영역이 다르게 나타났다. Oh 외 2인(2023)에서는 인문사회 계열 학생들이 자연공학 계열 학생들보다 생성형 AI 활용 시 발생할 수 있는 문제점을 더 민감하게 인식하고 있었다. Alnsour 외 5인(2025)은 보건 및 이공계 계열 학생들이 인문학 전공 학생들 보다 AI 윤리 인식 수준이 더 높았다고 보고하고 있으며, Asiksoy(2024)는 교육학 전공 학생들이 공학 전공 학생들보다 투명성 차원에서 유의미하게 더 높은 인식을

보였다고 보고하였다. 반면에 전공 계열에 따른 AI 윤리 인식이나 태도에 통계적으로 유의미한 차이가 발견되지 않았다고 보고한 연구들도 있다(S. Kang, 2025; Wang et al., 2025). AI 사용 유무에 따른 윤리 인식 정도를 살펴본 선행연구에서는 일관되게 경험이 많을수록 AI 윤리 인식 수준이 높은 것으로 보고하고 있다. 생성형 AI 사용 경험이 있는 집단이 없는 집단보다 AI 활용 교육 및 인용법 교육에 대한 요구도가 높았으며(Oh et al., 2023), AI 교육 경험이 있는 학생들이 공정성 차원에서 유의미하게 더 높은 윤리적 태도를 보였으며(Jang et al., 2022), AI 도구를 자주 사용하는 집단이 모든 윤리적 성찰 차원(인식, 평가, 공익)에서 통계적으로 유의미하게 높은 점수를 나타냈다(Wang et al., 2025).

이와 같은 선행연구 결과는 성별, 연령, 전공, AI 사용 경험이 AI 윤리에 대한 인식에 영향을 미친다는 것을 시사하며 이러한 차이를 이해하고 이를 바탕으로 집단의 특성에 적합한 AI 윤리 교육 및 정책 수립이 필요함을 시사한다. 또한 미래교육의 주체인 예비교사를 대상으로 한 변인별 인식 차이 연구가 부족하므로 예비교사를 대상으로 이러한 변인이 AI 윤리 인식에 어떠한 영향을 주는지 종합적으로 살펴볼 필요가 있음을 시사한다. 이에 본 연구에서 예비교사를 대상으로 성별, 연령, 전공, AI 사용 경험 변인이 AI 윤리 인식에 미치는 영향을 종합적으로 분석하고자 한다.

연구방법

연구 대상

본 연구 조사 대상자는 서울 경기 지역 대학교 및 교육 대학원에 재학 중인 예비교사 총 283명이다. 연구에 참여한 예비교사들의 일반적 특성(성별, 연령, 전공)은 Table 1과 같다.

TABLE 1
Description of Participants

Category	Gender (N=283)			Age Group (N=283)			Major (N=283)			
	Male	Female	NR	20s	30s	40 & Above	Language Edu.	Natural Sciences Edu.	Non-TM	Others
F	88	193	2	190	53	40	91	63	110	19
%	31.1	68.2	0.7	67.1	18.7	14.1	32.2	22.3	38.9	6.7

Note. F: Frequency, NR: No Response, Non-TM: Non-Teaching Major (majors not directly related to subject-based teaching areas in schools)

예비교사들의 성별 분포를 살펴보면 남성은 88명(31.1%), 여성은 193명(68.2%)으로 여성의 비율이 더 높았으며, 무응답이 2명(0.7%) 있었다. 예비교사들의 연령대 분포는 20대가 190명(67.1%)으로 가장 많았으며, 30대가 53명(18.7%), 40대 이상이 40명(14.1%)로 나타났다. 전공 분포는 비교과(상담심리)가 110명(38.9%)으로 가장 높은 비율을 보였으며, 어문계열(영어교육, 불어 교육)이 91명(32.2%), 자연계열(수학교육, 소프트웨어 교육)이 63명(22.3%), 기타 19명으로 나타났다(6.7%; 어문계열, 자연계열, 비교과에 포함되지 않는 전공의 경우 전공별 분석에서 제외하고 전체 분석에는 포함).

조사 도구 및 자료 분석

본 연구에서 사용한 설문지는 크게 Part 1, 2로 구성되었다. Part 1은 기본 정보 3개 문항(전공, 성별, 연령), AI 도구 사용 경험 관련 2개 문항, AI 관련 강좌 수강 경험 관련 2개 문항으로 구성하였으며 Part 2는 AI 윤리에 관한 문항 24개로 구성하였다(Appendix 참조). AI 윤리 문항은 국내외 AI 윤리 기준과 선행연구에서 제시된 핵심 요소를 비교 종합하여 Table 2와 같이 AI 윤리 교육에 필요한 핵심요소를 도출하고 이를 기반으로 개발하여 사용하였다.

TABLE 2
Core AI Ethics Elements Used for Survey Development

UNESCO (2021)	OECD (2019)	EU (2019)	Ministry of Science and ICT (2020)	S. A Bak (2023)	Derived AI Ethics Factors
Respect for human dignity, human rights, and fundamental freedoms as the highest values of AI use	Human-centered values	Human agency & oversight	AI that supports human growth	Subjectivity: Recognizing AI as a tool to be controlled and utilized by humans	Human-centeredness
Protection of diversity; prevention of discrimination, exclusion, and bias	Fairness and non-discrimination	Diversity, non-discrimination, fairness	Ensuring equal educational opportunities and preventing discrimination	Fairness: Civic attitudes that guard against bias and discrimination	Fairness
Emphasis on explainability, transparency, and understandability	Transparency and explainability	Transparency	Requirement to explain AI decision-making and usage processes	Critical understanding: Ability to critically understand and evaluate AI outcomes	Transparency
Consideration of harm prevention, safety, and environmental sustainability	Robustness, security, and safety	Technical robustness and safety	Safe and appropriate use of AI in education	Safety: Safe and appropriate use of AI technologies	Safety
Governance and accountability of AI actors (states, organizations, individuals)	Accountability	Accountability	Clarification of responsibility among developers, operators, and users	Responsibility: Ethical responsibility for the outcomes of AI use	Accountability

선행연구의 공통 요소를 기반으로 Table 2에서처럼 AI 윤리 하위 요인(인간중심성, 공정성, 투명성, 안전성, 책임성) 다섯개를 도출하고 이를 토대로 AI 윤리 설문 문항을 개발하여 예비교사를 대상으로 AI 윤리에 대한 인식을 조사하였다. 본 연구에서 사용한 AI 윤리 인식 설문지 구성과 하위 요인별 신뢰도는 Table 3과 같다.

TABLE 3
Measurement Structure and Reliability

AI Ethics Factor	Definition	Items	Cronbach's α
Human-centeredness	AI should be designed to complement humans without infringing upon human dignity and autonomy.	Q 19, 20, 21	.787
Fairness	AI should ensure equal opportunities for all users by avoiding discrimination and bias.	Q 1, 2, 3, 16, 17, 18	.869
Transparency	The operational mechanisms and decision-making processes of AI should be explainable and understandable.	Q 22, 23, 24	.800
Safety	AI should be reliable in both physical and digital environments and protected from errors and security threats.	Q 7, 8, 9, 13, 14, 15	.711
Accountability	Clearly defined responsible actors should exist for the outcomes arising from the development, operation, and use of AI.	Q 4, 5, 6, 10, 11, 12	.661

Part 1의 기본 문항은 선택형 및 주관식 형태로 구성하였다. AI 도구 사용 경험 및 AI 관련 강좌 수강 경험을 유무로 묻고 사용경험이 있는 경우와 AI 관련 강좌를 수강한 경우에는 구체적으로 사용한 AI 도구와 강좌명을 주관식으로 작성하도록 구성하였다. Part 2의 AI 윤리 문항은 5점 Likert 척도(1: 전혀 그렇지 않다, 5: 매우 그렇다)로 응답하도록 하였다. 내적 일관성을 확인한 결과 Cronbach α 는 인간중심성 .787, 공정성 .869, 투명성 .800, 안전성 .711로 양호한 수준이었으며, 책임성 요인은 .661로 상대적으로 다른 요인보다 낮은 신뢰도를 보였다. Cronbach α 가 .60 이상이면 탐색적인 연구에서 수용 가능하다는 점을 고려하여(Hair et al., 2010) 본 연구에서 책임성 요인도 활용 가능한

수준으로 판단되어 해석에 포함하였다.

온라인 및 면대면(교직과목이나 전공 수업시간 활용)으로 학생들의 동의를 받아 설문을 실시하고 수합하였다. 수합한 자료는 SPSS를 사용하여 반복측정 분산분석, *t*검정, 일원배치분산분석, Scheffé 사후검정을 사용하여 통계분석을 실시하였다.

결과

예비교사의 AI 도구 활용 실태 및 AI 관련 강좌 수강 유형

연구질문 1에 응답하기 위해 예비교사의 AI 도구 활용 경험과 AI 관련 강좌 수강 경험을 분석하였다(Table 4 참조). 전체 응답자 283명 중 75.6%(n=214)는 AI 도구 사용 경험이 있다고 응답하였으며, 24.4%(n=69)는 사용 경험이 없었다. AI 관련 강좌 수강 경험은 약 4분의 1 수준으로, 26.2%(n=73)가 수강 경험이 있다고 보고하였다(무응답 1.4%, n=4).

TABLE 4
Experience with AI Tools and AI-related Courses

Category	Experience Using AI Tools (N=283)		Experience Taking AI Course (N=283)		
	Yes	No	Yes	No	NR
n	214	69	73	206	4
%	75.6	24.4	26.2	72.8	1.4

Note. NR: No Response

AI 도구를 사용한 경험이 있다고 응답한 214명을 대상으로 구체적으로 어떤 AI 도구를 사용하였는지 주관식으로 응답하도록 하였으며, 중복 응답을 허용하였다. 응답에서 언급된 도구명을 개별 항목으로 분리하여 빈도를 집계한 뒤에 분석의 간결성을 위해 상위 10개 도구를 Table 5에 제시하였다.

TABLE 5
Frequency of AI Tools Used by Pre-service Teachers

AI Tool	Gender		Age			Major		
	Male (N=69)	Female (N=144)	20s (N=150)	30s (N=33)	40 & Above (N=32)	Language Edu. (N=78)	Natural Sciences Edu. (N=47)	Non-TM (N=75)
Google Translate	31	91	87	21	15	36	27	54
ChatGPT	42	66	84	12	13	47	19	35
Chatbot	12	34	25	13	9	8	22	15
Duolingo	4	18	16	6	1	6	8	9
Papago	8	9	16	1	0	8	0	8
Teachable Machine	5	11	4	7	5	2	9	0
Wrtn	4	9	8	3	2	7	3	1
Machine Learning for Kids	1	6	0	4	3	0	5	0
Wordtune	1	5	1	2	3	4	1	1
Perplexity	3	3	5	0	1	5	0	0

Table 5에 제시된 바와 같이, 예비교사들이 가장 빈번하게 사용한 AI 도구는 Google Translate(n=122)와 ChatGPT(n=108)로 나타났다. 이어서 Chatbot(n=46), Duolingo(n=22), Papago(n=17)등의 순서로 나타났다. 그 외에도 Teachable Machine, Wrtn, Machine Learning for Kids, Wordtune, Perplexity와 같은 다양한 AI 기반 학습 도구를 사용하고 있는 것으로 나타났다. 성별에 따른 활용 양상을 살펴보면, 남녀 모두 Google Translate와 ChatGPT를 사용한다는 빈도가 가장 높았으며 남성보다 여성 예비교사 응답자 비중이 상대적으로 높게 나타났다. 연령대별로는 20대 예비교사의 AI 도구 활용 경험이 가장 많이 보고되었으며 30대와 40대 이상에서 보고하고 있는 양상과 유사하게 Google Translate와 ChatGPT를 많이 활용하는 것으로 응답 빈도가 나타났다. 전공별 분석 결과, 어문계열과 비교과계열 예비교사들이 AI 도구 활용 경험을 자연계열 예비교사들 보다 많이 보고하였으며, 모든 전공에서 Google Translate와 ChatGPT를 중심으로 활용 양상을 보이고 있다. 이와 같은 결과를 통해 예비교사들은 AI 도구 언어 번역지원과 정보 탐색 중심의 범용 AI 도구를 주로 사용하고 있음을 알 수 있다. 또한 성별, 전공별, 연령에 따른 AI 도구 활용 양상이 유사함을 알 수 있다.

AI 관련 강좌를 수강한 경험이 있다고 응답한 참여자에게 수강한 강좌명을 주관식으로 응답하도록 하였으며, 응답된 강좌를 Table 6에 제시한 분류 기준에 따라 유형화하였다.

TABLE 6
Classification of AI-related Courses

Type	Course Titles
Type 1: Embedded AI in Existing Courses	TESOL; Logic and Essay Writing; Teaching Practice; Instructional Methods and Educational Technology; Educational Technology; Curriculum; Understanding Language and Educational Technology; Understanding Language and Career Exploration; Research on English Language Teaching Materials
Type 2: AI Literacy and Interdisciplinary Courses	AI and Music; AI and Philosophy; AI and Computational Thinking; Understanding AI; Big History and Artificial Intelligence; Artificial Intelligence and Future Education; The Present and Future of Artificial Intelligence
Type 3: AI- Applied Education Courses	AIDT Pre-service Teacher Training; Introduction to AIDT; AI-Based Education; Understanding AI-Based Educational Applications; AI-Based Mathematics Education; Theory and Practice of AI-Assisted English Education; Duolingo-Based Spanish Language Course; Intelligent Information-Based English Education (Capstone Design)
Type 4: AI/Data Science Technical Courses	Fundamentals of AI and Content Development; Data Mining; Deep Learning Programming; Machine Learning; Understanding Big Data; Mathematical Programming; Algorithms; Algorithms and Game Content; Intelligent Information Science; Computational Thinking and Data Literacy; Computer Vision

AI 관련 강좌명은 주관식으로 수집되었으나, 응답에 강좌명만 제시되어 있어 수업 목표, 평가 방식, 세부 내용까지 확인하기에는 제한적이었다. 따라서 본 연구에서는 강좌명에 드러난 핵심 키워드(예: AI, AIDT, 교육, 교과교육, 머신러닝 등)와 강좌의 맥락을 기준으로 AI 관련 강좌를 4가지 유형으로 분류하였다(Table 6 참조). 두명의 연구자가 독립적으로 분류한 후, 불일치 항목은 논의를 통해 최종 합의하였다. 구체적으로 Type 1은 기존 교직 또는 교과교육 강좌 내에서 AI가 보조적 주제 또는 일부 내용으로 통합되어 다루어지는 강좌, Type 2는 AI의 개념적 이해와 인문, 예술, 역사 등 인문학적 맥락에서의 융합적 탐구를 목적으로 하는 AI리터러시 및 교양 성격의 강좌, Type 3는 AIDT 및 AI 기반 교육/교수·학습 도구 등 교육 현장에의 AI 활용을 핵심 목적으로 하는 강좌(교과교육 맥락의 AI 활용 포함), Type 4는 머신러닝, 딥러닝, 데이터마이닝 등 AI/데이터과학의 핵심 원리와 기술을 심화적으로 다루는 강좌로 정의하였다.

수강 경험 집단에서 비교적 반복적으로 언급된 강좌를 중심으로(빈도 4회 이상) 성별, 연령, 전공별로 수강 양상을 탐색적으로 살펴보았으며, 그 결과는 Table 7과 같다. 이 분석은 표본 수가 제한적이고 강좌명이 주관식으로 수집되었다는 점에서 기술통계 수준의 해석으로 한정하였다.

TABLE 7
Frequency of AI Related Course Taken by Pre-service Teachers

AI related Courses	Gender		Age			Major		
	Male (N=33)	Female (N=41)	20s (N=50)	30s (N=10)	40 and above (N=14)	Language Edu. (N=37)	Natural Sciences Edu. (N=33)	Non-TM (N=4)
Understanding AI	4	5	6	3	0	0	9	0
Theory and Practice of AI-Assisted English Education	5	1	0	0	5	5	0	0
AI-Based Mathematics Education	1	4	3	1	1	1	4	0
Intelligent Information-Based English Education (Capstone Design)	4	2	4	0	0	4	0	0

빈도 4회 이상으로 보고된 강좌는 ‘AI의 이해(Understanding AI)’, ‘AI 활용 영어 교육의 이론과 실제(Theory and Practice of AI-Assisted English Education)’, ‘인공지능 기반 수학 교육(AI-Based Mathematics Education)’ 등으로, AI 리더러시 강좌(Type 2)와 교과교육 맥락의 AI 활용 강좌(Type 3)가 상대적으로 자주 언급되었다.

예비교사의 AI 윤리 인식

AI 윤리 하위 영역별 인식

연구문제 2에 응답하기 위해 먼저 예비교사의 AI 윤리 인식을 하위 요인별로 분석하였으며, 그 결과는 Table 8과 같다. 전체 평균은 3.18($SD = .55$)로 중간 수준으로 나타났으며, 구체적으로 인간중심성은 3.81, 공정성은 3.24, 투명성은 3.01, 안전성은 2.81, 책임성은 3.12로 나타났다.

TABLE 8
Overall AI Ethics Awareness among Pre-service Teachers

AI Ethic Factor	Mean	SD	RM ANOVA(GG)
Human-centeredness	3.81	.67	
Fairness	3.24	.81	
Transparency	3.01	.79	
Safety	2.81	.64	
Accountability	3.12	.57	
Total	3.18	.55	$F(3.60, 1010.51)=160.61, p<.001, \eta^2 = .364$

AI 윤리 5개 하위 요인간 평균 차이를 검증하기 위해 반복측정 분산분석을 실시하였다. Mauchly의 구형성 검정 결과 구형성 가정이 위배되어 Greenhouse-Geisser 보정을 적용하였다. 분석 결과, 요인간 평균 차이는 통계적으로 유의한 차이가 있는 것으로 나타났다($F(3.60, 1010.51)=160.61, p<.001$). 평균은 인간중심성이 가장 높았으며 이어서 공정성, 책임성, 투명성으로 나타났으며, 안전성이 상대적으로 가장 낮게 나타났다. 이와 같은 결과는 예비교사들이 AI 시스템의 안전성 및 잠재적 위협에 대한 이해가 부족함을 시사하며 교원양성 과정에 AI 활용 교육과 더불어 AI 윤리교육에서 특히 안전성에 대한 교육을 보완할 필요가 있음을 시사한다.

예비교사의 배경 변인에 따른 AI 윤리 인식 차이

AI 윤리의 하위 요인에 대한 인식이 성별, 연령별, 전공별에 따라 차이를 보이는지 분석하였다. 먼저 성별에 따른

AI 윤리 하위 요인별 차이를 독립표본 *t*-검정을 통해 분석한 결과는 Table 9와 같다.

TABLE 9
Differences in AI Ethics Awareness by Gender among Pre-service Teachers

AI Ethics Factor	Gender	<i>n</i>	Mean	<i>SD</i>	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Human-centeredness	Male	88	3.90	0.70	1.508	.133	0.19
	Female	193	3.77	0.66			
Fairness	Male	88	3.41	0.84	2.399*	.017	0.30
	Female	193	3.17	0.79			
Transparency	Male	88	3.21	0.78	2.828**	.005	0.37
	Female	192	2.92	0.80			
Safety	Male	88	2.95	0.63	2.657**	.008	0.35
	Female	193	2.73	0.64			
Accountability	Male	88	3.30	0.54	3.732***	.000	0.46
	Female	193	3.04	0.57			

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

성별에 따른 AI 윤리 인식 차이를 분석한 결과, 인간중심성 요인에서는 남성과 여성의 평균간 차이가 통계적으로 유의하지 않았다($t=1.508, p=.133, d=.19$). 이는 성별에 따라 AI의 인간 중심적 가치에 대한 인식 수준이 비교적 유사하며, 효과크기 또한 작은 수준임을 의미한다. 반면, 공정성 요인에서는 남성이 여성보다 평균이 유의미하게 높게 나타났다($t=2.399, p=.017, d=.30$). 이는 남성이 여성에 비해 AI 활용 맥락에서 공정성 인식을 더 강하게 보이는 경향이 있음을 시사하며, 효과크기는 작은 수준으로 확인되었다. 투명성 요인에서도 남성이 여성보다 유의미하게 높은 점수를 나타냈다($t=2.828, p=.005, d=.37$). 이는 남성이 AI 작동 원리와 의사결정 과정의 투명성에 대해 보다 긍정적으로 인식하는 경향이 있음을 보여주며, 효과크기는 작은 수준에서 중간 수준에 근접하였다. 안전성 요인 역시 남성이 여성보다 유의미하게 높게 나타났다($t=2.657, p=.008, d=.35$). 이와 같은 결과는 남성이 여성보다 AI 기술의 안전성에 대해 상대적으로 높은 신뢰를 가지고 있음을 의미한다. 책임성 요인에서도 남성이 여성보다 유의미하게 높은 평균 점수를 보였다($t=3.732, p=.001, d=.46$). 책임성 요인의 효과크기는 중간 수준에 근접하였다.

요약하면 성별에 따른 AI 윤리 인식은 공정성, 투명성, 안전성, 책임성 요인에서 남성이 여성보다 유의하게 높게 나타난 반면, 인간중심성 요인에서는 유의한 차이가 확인되지 않았다. 또한 효과크기 분석 결과, 인간 중심성은 작은 효과크기, 나머지 하위 요인들은 작은 수준에서 중간 수준에 근접한 효과크기를 보여 일부 AI 윤리 하위 요인에서 성별에 따른 인식 차이가 존재함을 시사한다.

예비교사를 대상으로 연령별로 AI 윤리의 하위 요인에 대한 인식 차이를 일원배치분산분석과 Scheffé 사후검정을 통해 분석한 결과는 Table 10과 같다. 분석에 앞서 등분산성 검정을 실시한 결과, 모든 하위 요인에서 분산의 동질성 가정이 충족되어($p>.05$) 일원분산분석을 실시하였다. 그 결과 첫째, 인간중심성 요인에서 연령대별 평균은 20대(3.89)가 가장 높았고, 40대 이상(3.73), 30대(3.57) 순으로 연령대 간 차이가 통계적으로 유의한 것으로 나타났다. 효과크기 분석 결과, $\eta^2=.034$ 로 작은 수준의 효과크기가 나타났으나 사후검정에서는 세 집단 간 차이가 나타나지 않아 평균값 차이가 실제로 명확한 집단 간 차이로 이어지지 않는 것으로 해석된다. 둘째, 공정성 요인에서도 연령대별 차이가 통계적으로 유의미하게 나타났다. 평균은 20대(3.38)가 가장 높았으며, 30대와 40대가 동일하게(2.95) 나타났다. 사후검정 결과 20대가 30대와 40대 이상의 예비교사들보다 공정성에 대한 인식이 높은 것을 확인할 수 있었으며 효과크기는 $\eta^2=.063$ 으로 중간 수준에 해당되어 연령대에 따른 공정성 인식 차이가 비교적 뚜렷함을 시사한다. 이와 같은 결과는 젊은 세대일수록 AI 활용 과정에서의 공정성 문제에 대해 더 민감하게 인식하고 있음을 의미한다. 셋째, 투명성 요인에서도 연령대에 따른 평균의 차이가 유의하게 나타났다. 사후검정결과 20대(3.12)가 30대(2.77)와 40대 이상(2.79)보다 평균이 높은 것으로 나타났다. 효과크기는 $\eta^2=.040$ 으로 작은 수준에서 중간 수준에 근접한 것으로 나타나 연령대에 따른 투명성 인식 차이가 존재함을 보여준다. 넷째, 안전성 요인의 경우 평균이 20대(2.90), 40대 이상(2.64), 30대(2.61) 순으로 나타났으며 통계적으로도 유의미하게 차이가 있는 것으로 나타났다. 효과크기도 $\eta^2=.040$ 으로 작은 수준의 효과 크기를 보였으나 사후검정

결과 세 집단간의 차이가 없는 것으로 나타났다. 이는 연령대별 평균 차이는 존재하나 실제로 모든 연령층이 AI 안전성에 대해 비교적 유사한 수준의 인식을 공유하고 있음을 의미한다. 마지막으로 책임성 요인에서도 연령대에 따른 차이가 있는 것으로 나타났다($p < .05$). 평균은 20대(3.19), 40대 이상(2.98), 30대(2.97)의 순서로 나타났으며 효과크기는 $\eta^2 = .032$ 로 작은 수준의 효과크기가 확인되었다. 그러나 사후 검정 결과 세집단의 차이가 나타나지 않았다. 이와 같은 결과는 책임성 인식이 연령에 따라 다소 차이를 보이는 하나, 전반적으로 연령과 관계없이 AI 윤리에서 공통적으로 중요하게 인식되는 요소임을 시사한다.

TABLE 10
Differences in AI Ethics Awareness by Age among Pre-service Teachers

AI Ethics Factor	Age	n	Mean	SD	F	p	η^2	Scheffé'
Human-centeredness	20s	190	3.89	0.63	4.949**	.008	.034	a=b=c
	30s	53	3.57	0.77				
	40 and above	40	3.73	0.69				
Fairness	20s	190	3.38	0.77	9.434***	.000	.063	a>b=c
	30s	53	2.95	0.81				
	40 and above	40	2.95	0.82				
Transparency	20s	190	3.12	0.78	5.874**	.003	.040	a>b=c
	30s	53	2.77	0.72				
	40 and above	39	2.79	0.90				
Safety	20s	190	2.90	0.65	5.861**	.003	.040	a=b=c
	30s	53	2.61	0.65				
	40 and above	40	2.64	0.56				
Accountability	20s	190	3.19	0.57	4.590*	.011	.032	a=b=c
	30s	53	2.97	0.57				
	40 and above	40	2.98	0.54				

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

요약하면 연령에 따른 AI 윤리 인식은 공정성, 투명성 요인에서만 20대 예비교사들이 다른 연령대 예비교사보다 유의하게 높게 나타났으며 다른 요인에서는 유의한 차이가 확인되지 않았다. 또한 효과크기 분석 결과, 공정성은 중간 효과크기를, 인간중심성에서는 작은 효과크기를, 나머지 하위 요인들은 작은 수준에서 중간 수준에 근접한 효과크기를 보여 일부 AI 윤리 하위 요인에서 연령에 따른 인식 차이가 존재함을 시사한다.

전공별로 AI 윤리의 하위 요인에 대한 인식 차이를 일원배치분산분석과 Scheffé 사후검정을 통해 분석한 결과는 Table 11과 같다. 분석에 앞서 등분산성 검정을 실시한 결과, 모든 하위 요인에서 분산의 동질성 가정이 충족되어($p > .05$) 일원분산분석을 실시하였다. 첫째, 인간중심성 요인에서는 어문계열 평균이(4.03) 가장 높았으며, 자연계열과 비교과 평균은 3.71로 동일하게 나타났다. 분산분석 결과 전공에 따른 차이가 통계적으로 유의하게 나타났으며, 효과크기 분석 결과 $\eta^2 = .036$ 으로 작은 수준에서 중간 수준에 근접하는 효과크기가 확인되었다. 사후검정 결과 어문계열의 예비교사들이 자연계열과 비교과 예비교사들보다 인간중심성 인식 수준이 높은 것으로 나타났다.

둘째, 공정성 요인에서도 전공별 차이가 통계적으로 유의하게 나타났다. 어문계열(3.54)의 평균이 가장 높았으며 자연계열(3.09)과 비교과(3.13)는 유사한 평균을 보였다. 효과크기는 $\eta^2 = .046$ 으로 중간 수준에 해당되며 사후검정 결과 어문계열에 속한 예비교사들이 자연계열과 비교과에 속한 예비교사들 보다 유의하게 AI 공정성에 대한 인식이 높은 것으로 나타났다. 셋째, 투명성 요인에서도 전공에 따른 유의한 차이가 나타났다. 구체적으로 사후 검정 결과 어문계열(3.37)이 자연계열(2.86)과 비교과(2.83) 보다 높게 나타났으며, 자연계열과 비교과 간의 차이는 없는 것으로 나타났다. 효과크기는 $\eta^2 = .067$ 로 중간 수준의 효과에 해당되어 전공에 따른 투명성 인식 차이가 비교적 뚜렷함을 보여준다. 넷째, 안전성 요인에서도 집단 간 차이가 통계적으로 유의미하게 나타났다. 어문계열의 평균(3.08)이 자연계열(2.72)과 비교과(2.62)보다 높았으며, 사후검정 결과로 이러한 차이를 확인할 수 있었다. 마지막으로 책임성

요인에서도 다른 요인과 유사하게 통계적으로 유의미하게 어문계열의 평균(3.39)이 자연계열(2.98)과 비교과(2.98)보다 높게 나타났으며, 사후 검정 결과로 이러한 차이를 확인할 수 있었다. 효과크기는 $\eta^2=.077$ 로 큰 효과크기가 확인되었다. 요약하면 어문계열 예비교사들이 인간중심성, 공정성, 투명성, 안전성, 책임성 모두에서 자연계열과 비교과 예비교사들보다 인식 수준이 높았으며, 자연계열과 비교과 예비교사간의 이러한 인식 수준 차이는 없는 것으로 나타났다. 또한 효과크기도 대부분의 요인에서 중간 크기의 효과 수준을 보여주고 있어 전공에 따른 인식 차이가 존재함을 시사한다.

TABLE 11
Differences in AI Ethics Awareness by Major among Pre-service Teachers

AI Ethics Factor	Major	n	Mean	SD	F	p	η^2	Scheffé'
Human-centeredness	Language Edu.	91	4.03	0.57	5.271***	.001	.036	a>b=c
	Natural Sciences Edu.	63	3.71	0.78				
	Non-TM	110	3.71	0.68				
Fairness	Language Edu.	91	3.54	0.82	6.812***	.000	.046	a>b=c
	Natural Sciences Edu.	63	3.09	0.83				
	Non-TM	110	3.13	0.75				
Transparency	Language Edu.	91	3.37	0.87	10.015***	.000	.067	a>b=c
	Natural Sciences Edu.	63	2.86	0.77				
	Non-TM	109	2.83	0.67				
Safety	Language Edu.	91	3.08	0.68	8.973***	.000	.060	a>b=c
	Natural Sciences Edu.	63	2.72	0.63				
	Non-TM	110	2.64	0.58				
Accountability	Language Edu.	91	3.39	0.59	11.686***	.000	.077	a>b=c
	Natural Sciences Edu.	63	2.98	0.54				
	Non-TM	110	2.98	0.51				

Note. *** $p<.001$

마지막으로 AI 도구 사용 경험 유무 및 AI 관련 강좌 수강 경험 여부에 따른 AI 윤리 하위요인별 인식 차이를 분석한 결과 모든 하위 요인에서 통계적으로 유의한 차이가 나타나지 않았으며($p>.05$), 효과크기 또한 Cohen's d 기준으로 -.09에서 .17 범위로 매우 작은 효과크기 수준으로 나타났다.

논의 및 결론

본 연구는 인공지능 전환(AI) 환경에서 교육 현장의 변화를 주도할 예비교사들의 AI 윤리 인식을 다차원적으로 점검하고, 인구통계학적 변인 및 AI 관련 경험에 따른 인식 차이를 분석하기 위하여 크게 배경 변인(전공, 성별, 연령)에 따른 인식 차이와 AI 활용 경험에 따른 인식 차이라는 두 가지 핵심 연구문제를 중심으로 기술하였다. 국내외 주요 AI 윤리 프레임워크(OECD, 2019; UNESCO, 2021)를 기반으로 도출된 5대 핵심 원칙(인간중심성, 공정성, 투명성, 안전성, 책임성)을 중심으로 분석한 결과, 예비교사의 전반적 인식 수준은 중간($M=3.18$) 정도로 나타났으며 요인별로 뚜렷한 인식 차이가 확인되었다. 이러한 결과는 예비교사 교육에서 AI 윤리를 포괄적, 선언적으로 제시하는 접근만으로는 요인별 취약성을 보완하기 어렵고, 요인별 성취목표와 수행 장면 중심의 교수·평가 설계가 필요함을 시사한다.

첫째, 예비교사들은 '인간중심성' 가치에는 높은 공감대를 보였으나, '안전성' 요인에 대한 인식은 상대적으로 낮았다. 이는 예비교사들이 AI 활용의 규범적 지향점(인간 존엄, 자율성 존중 등)에는 동의하면서도, 실제 교육

현장에서 직면하게 될 데이터 거버넌스(수집·저장·공유), 보안 취약점, 환각 현상에 따른 허위정보 확산, 책임 소재 불명확성 등 실무적 위험 관리 영역은 충분히 내면화하지 못했음을 시사한다. K-12 맥락에서 AI 도입이 학습자 안전, 개인정보 보호, 알고리즘 편향 이슈와 직결된다는 점을 고려할 때(Akgun & Greenhow, 2022; Gouseti et al., 2025), 안전성 요인을 ‘기술적 안전(보안·데이터 보호·접근권한)’과 ‘교육적 보호(학습자 피해 예방, 디지털 웰빙, 허위정보·편향 대응)’로 세분화하여 다루는 체계적인 교육이 요구된다. 예를 들어, ‘수업 중 학생 데이터가 AI 서비스로 입력·저장되는 과정의 위험을 점검하는 절차’, ‘생성형 AI 답변의 출처·정확성 검증(교차검증·근거 확인)’과 ‘오류 발견 시 지도/기록/보고 흐름을 포함한 교실 수준의 안전 프로토콜’을 훈련 내용으로 포함할 필요가 있다.

둘째, 집단 간 차이 분석 결과, 성별, 연령, 전공에 따른 유의미한 인식 격차가 발견되었다. 성별 분석에서 남성이 주요 요인에서 높은 점수를 보인 결과는 여성이 AI 윤리에 더 민감하다는 선행연구(Asiksoy, 2024; Jang et al., 2022; Mihoc et al., 2025; Ryu et al., 2025; Wang et al., 2025)와 상반될 수 있는데, 이는 본 연구 문항이 ‘윤리적 우려(concern)’보다는 ‘규범적 원칙에 대한 동의(endorsement)’ 성격의 진술에 가깝기 때문으로 풀이된다. 즉 동일한 윤리 주제라도, 한 집단은 규범적 동의 수준이 높고 다른 집단은 위험 인식/우려 수준이 높을 수 있어, 측정 문항의 초점에 따라 결과가 달라질 수 있다. 연령대별로는 20대 예비교사가 다른 연령대 보다 공정성과 투명성에서 높은 인식을 보였으나, 인간중심성이나 책임성은 연령과 관계없이 비교적 모든 연령대에서 인식이 되고 있는 것으로 나타났다. 이는 예비교사의 연령이 낮을수록 AI 공정성과 의존성 문제에 더 민감하고, 30세 미만의 경우 AI에 더 긍정적인 태도를 보인다는 일부 연구와 맥을 같이 한다(Konika et al., 2025; Ryu et al., 2025). 이러한 결과는 예비교사의 연령에 따라 AI에 대한 윤리 인식 수준이 다르게 형성될 수 있음을 보여주며, 향후 교원양성 교육에서 이러한 특성을 반영한 차별화된 윤리 교육 설계가 필요함을 시사한다. 전공별로는 어문계열 예비교사의 인식이 가장 높았으며, 이는 생성형 AI를 활용한 텍스트 생성 및 번역 과정에서 평가 공정성, 표절, 저작권, 출처 표기 등 윤리적 쟁점을 비교적 직접적으로 경험할 기회가 많았던 결과로 해석된다(Oh et al., 2023). 이러한 격차는 모든 예비교사에게 동일한 콘텐츠를 제공하는 일률적 교육 방식에서 벗어나, 교과 맥락(평가 방식, 산출물 유형, 학습 활동의 특성)에 따라 윤리 쟁점이 어떻게 달라지는지를 반영한 차등화 설계가 필요함을 시사한다.

셋째, 예비교사의 AI 활용 경험이 상당히 높은 수준으로 나타났음에도 불구하고 윤리 인식이 중간 수준에 머무른 점은 중요한 시사점을 제시한다. 이는 AI 도구의 사용 경험 자체가 윤리적 이해나 비판적 성찰로 자동적으로 이어지지 않음을 의미한다. 구체적으로, AI 관련 강좌 수강이나 활용 경험의 ‘유무’ 및 ‘양’이 윤리 인식에 유의미한 차이를 만들지 못한 결과는 ‘경험의 질적 차이(quality of experience)’에 주목하게 한다. 특히 예비교사들이 주로 언어 번역이나 자료 탐색과 같은 제한적 기능 중심으로 AI를 활용하고 있으며, AI 관련 강좌 역시 주로 기술적 사용법(how-to)이나 기능 숙달에 치우쳐 있어 실제 윤리적 쟁점을 비판적으로 성찰하고 판단하는 기회를 충분히 제공하지 못했을 가능성이 크다. 따라서 AI 활용 교육과 윤리 교육은 분리된 영역이 아니라 실제 수업 맥락 속에서 활용 경험과 윤리적 성찰이 동시에 이루어지는 통합적 접근으로 설계될 필요가 있다. 이는 단순한 기술 노출이나 도구 활용 중심의 교육이 비판적·윤리적 성찰을 자동적으로 담보하지 않음을 의미하며(Holmes et al., 2022), ‘AI 관련 강좌’가 존재하더라도 윤리 내용이 명시적 학습목표로 구조화되어 있지 않거나, 다루는 범위·깊이가 상이할 경우 학습 효과가 분산될 수 있음을 시사한다. UNESCO(2023)가 강조하듯이, 미래의 교사 교육은 기술 훈련을 넘어 실제 교육 현장의 딜레마를 활용한 판단 훈련과 책임 소재를 구체화하는 사례 기반·수행 중심 모듈형 프로그램으로 재설계되어야 한다. 예컨대 ‘AI로 생성된 피드백을 교사가 그대로 제공했을 때 발생하는 오류 책임’, ‘자동채점/추천 결과의 편향 가능성 점검과 수정 절차’와 같이 수업·평가 장면에서의 의사결정을 중심으로 윤리 판단을 훈련하는 구성이 요구된다.

결론적으로 본 연구의 결과를 종합하여 볼 때, AI 활용 경험의 양보다 활용의 질과 성찰 과정이 중요함을 시사하며, 교원양성 교육에서 AI 활용 교육과 윤리 교육을 분절적으로 제공하기 보다는 통합된 역량 교육 체계로 재구조화 할 필요가 있음을 알 수 있다. 이는 AI 리터러시가 단순한 AI 도구 사용 경험이 아닌 이해, 평가, 윤리적 판단 능력을 포함하는 복합적 역량을 의미하므로 지식, 태도, 실천 능력이 함께 형성될 때 교육적 효과가 있다는 연구 결과와 맥을 같이 한다(Akgun & Greenhow, 2022; Kathala & Palakurthi, 2024; Long & Magerko, 2020). 또한 여러 교사 전문성 개발연구에서 언급하였듯이 AI 활용 교육은 기술 습득을 넘어 교수·학습 맥락과 윤리적 성찰을 통합적으로 다룰 때 그 효과가 높다는 점을 다시 한번 강조하는 결과로 볼 수 있으며(Fadel et al., 2019; UNESCO, 2021), 더 나아가 기술 활용 능력이 교수지식(pedagogical knowledge)과 결합될 때, 단순한 도구 사용을 넘어 실제 수업 개선과 학습 효과로 이어지는 교육적 실천이 가능해진다는 것이다(Koehler et al., 2013; Mishra & Koehler, 2006).

위와 같은 시사점을 바탕으로 예비교사의 AI 윤리 역량 강화를 위해 다음과 같은 정책적·교육적 방향성을 제안하고자 한다. 먼저, 역량 기반 AI 윤리 교육 체계(competency-based design)로의 전환을 제안하는 바이다. 이는 AI 윤리를 단일 주제로 다루기보다 5개 하위 요인을 명시적 성취목표로 설정하고, ‘교실 내 데이터 보안 사고 대응’, ‘AI 산출물 검증 및 기록’, ‘환각/허위정보 발견 시 교수적 개입’ 등 실제 사례 기반 시나리오 학습과 수행평가를 결합해야 한다(Wieczorek et al., 2025). 수행평가에서는 단순 지식 확인을 넘어, 위험요인 식별, 대안 비교, 조치 계획 수립, 학생/학부모 커뮤니케이션 문안 작성 등 실제 실행 산출물을 요구함으로써 원칙을 실천 역량으로 연결할 수 있다.

또한 전공 특성화 및 차별화된 교육 설계(differentiated and contextualized modules)의 필요성을 강조하는 바이다. 이는 보편적 가치를 다루는 ‘기초 공통 모듈’(최소 준수 기준, 공정성·투명성·책임성의 공통 개념)과 전공별 교수·학습 특수성을 반영한 ‘심화 맞춤형 모듈’을 병행하는 이원화 전략이 필요하다. 특히 전공별 특화된 학습 장면(예: 서·논술형 평가, 탐구보고서, 프로젝트 산출물, 자동 채점/피드백 도구 활용)을 기준으로 윤리적 딜레마를 구조화하고, 이를 현장에서 즉시 활용 가능한 체크리스트·루브릭·수업 설계 템플릿으로 구체화할 것을 제안한다.

본 연구는 국제적 공통분모인 5대 원리를 통해 예비교사의 윤리적 취약 지점인 ‘안전성’을 경험적으로 확인하고, 인구통계학적 변인 및 경험 변인에 따른 차이를 제시함으로써 교원양성 교육과정 설계에 활용 가능한 근거 기반(evidence-based) 시사점을 제공했다는 점에서 학술적·정책적 기여가 있다. 다만 본 연구는 ‘AI 관련 강좌’의 내용·질(윤리 포함 여부, 교수 방식, 과제 구성)을 세분화하여 통제하지 못했다는 한계를 지닌다. 또한 예비교사의 전공 교과를 모두 포괄하지 못하고 어문계열, 자연계열, 비교과 영역을 중심으로 분석이 이루어져 다양한 전공 교과 특성을 충분히 반영하지 못하였다. 따라서 향후 연구에서는 동일 요인을 ‘규범적 동의’와 ‘윤리적 우려’로 구분하여 측정하는 모델 검증, 강좌/프로그램의 구체적 설계 요소, 포괄적인 전공 특성 등이 윤리 인식 및 실제 판단 수행에 미치는 영향을 확인하는 후속 검증이 필요하다. 또한 본 연구에서 도출된 인식의 특징이 실제 교수·학습 장면에서 어떻게 발현되는지 확인하기 위한 질적 사례 연구 역시 수행될 필요가 있다. 특히 저작권, 데이터 편향성, 평가의 공정성 등 실제 교실 상황에서 마주하게 될 구체적인 윤리적 딜레마를 예비교사들이 어떻게 인지하고 해결해 나가는지 그 의사결정 과정을 심층적으로 추적할 필요가 있다. 이러한 논의의 결과들이 향후 교원 역량 체계와 교사교육 프로그램의 실질적 변화로 이어지기를 기대하는 바이다.

References

- Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2(3), 431–440. <https://doi.org/10.1007/s43681-021-00096-7>
- Alnsour, M. M., Almond, H., Qouzah, L., Momani, M. Q., Alamoush, R. A., & AL-Omiri, M. K. (2025). Artificial intelligence usage and ethical concerns among Jordanian university students. *International Journal for Educational Integrity*, 21(1), 31. <https://doi.org/10.1007/s40979-025-00206-6>
- Asiksoy, G. (2024). An investigation of university students' attitudes toward artificial intelligence ethics. *International Journal of Educational Technology in Higher Education*, 21(18), 1–19. <https://doi.org/10.1186/s41239-024-00418-9>
- Awal, M. R. (2024). Curse of blessing? Students' experience form ChatGPT with an application of Colaizzi's phenomenological descriptive method of enquiry and content analysis. *Higher Education, Skills and Work-Based Learning*, 14(6), 1299–1313. <https://doi.org/10.1108/HESWBL-09-2023-0249>
- Bae, Jinah., Lee, Jeonghun., & Cho, Jungwon. (2022). An artificial intelligence ethics education model for practical power strength. *Journal of Industrial Convergence*, 20(5), 83–92. <https://doi.org/10.22678/JIC.2022.20.5.083>
- Bak, Sang A. (2023). A study on the contents of AI citizenship education: Focusing on AI ethics guidelines. *Journal of AI Ethics*, 2(2), 114–150. <https://scholar.kyobobook.co.kr/article/detail/4050067780319>
- Brown, T. B., Ryder, N., Subbiah, M., Kaplan, J. Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901. <https://arxiv.org/abs/2005.14165>
- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1819. <https://doi.org/10.2139/ssrn.3213954>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT, USA*. <https://arxiv.org/abs/1810.04805>
- EU. (2019). Ethics guidelines for trustworthy AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- Fadel, C., Holmes, W., & Bialik, M. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Pearson.

- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (Berkman Klein Center Research Publication No. 2020-1). Berkman Klein Center for Internet & Society, Harvard University. <https://doi.org/10.2139/ssrn.3518482>
- Ghotbi, N., & Ho, C. M. L. (2021). Moral awareness of college students regarding artificial intelligence. *AI & Society*, 36(2), 557–567. <https://doi.org/10.1007/s00146-020-01048-9>
- Gouseti, A., James, F., Fallin, L., & Burden, K. (2025). The ethics of using AI in K-12 education: A systematic literature review. *Technology, Pedagogy and Education*, 34(2), 161–182. <https://doi.org/10.1080/1475939X.2024.2428601>
- Gulchenko, V. (2024). *Navigating the risks: An examination of the dangers associated with artificial general intelligence and artificial superintelligence*. SSRN. <https://ssrn.com/abstract=4941716>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson Prentice Hall.
- Heo, Eu-Sun., Lee, Yeon-Hee., & Shim, Ji-Won. (2020). Why ethics is: A landscape of modern AI ethics debate, its features and limitations. *Human Beings, Environment and Their Future*, 24, 165–209. <https://doi.org/10.34162/hefins.2020..24.006>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. M. T., Cukurova, M., Bittencourt, I. L., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32, 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Hsu, H.P., Mak, J., Werner, J., White-Taylor, J., Geiselhofer, M., Gorman, A., & Torrejon Capurro, C. (2024). Preliminary study on pre-service teachers' applications and perceptions of generative artificial intelligence for lesson planning. *Journal of Technology and Teacher Education*, 32(3), 409–437. <https://www.learntechlib.org/p/224413>
- Huang, R., Yin, Y., Zhou, N., & Lang, F. (2026). Artificial intelligence in K-12 education: An umbrella review. *Computers and Education: Artificial Intelligence*, 10, Article 100519. <https://doi.org/10.1016/j.caeai.2025.100519>
- Jang, Yeonju., Choi, Seongyune., & Kim, Hyeoncheol. (2022). Development and validation of an instrument to measure undergraduate students' attitudes towards the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education. *Educational and Information Technologies*, 27, 11635–11667. <https://doi.org/10.1007/s10639-022-11086-5>
- Jeong, Dai., & Park, Hyoungbin. (2023). Development and implementation of an elementary-school AI ethics education program. *Korean Journal of Elementary Education*, 34(3), 39–54. <https://doi.org/10.20972/kjee.34.3.202309.39>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Jun, Soojin. (2023). Development and effectiveness of artificial intelligence ethics education program based on CAI model. *The Journal of Korean Association of Computer Education*, 26(1), 23–31. <https://doi.org/10.32431/kace.2023.26.1.003>
- Kang, Mun-Suk. (2024). A study on prospective teachers' perceptions of AI education and AI teachers. *Journal of Educational Community Research and Practice*, 6(4), 29–47. <https://doi.org/10.23108/decr.2024.6.4.29>
- Kang, Soomin. (2025). An exploratory analysis of university students' use of generative AI tools and their perception of citation and ethical compliance: Focusing on the case of K university. *Educational Research*, 47(2), 103–124. <https://doi.org/10.35510/JER.2025.47.2.103>
- Kathala, K. C. R., & Palakurthi, S. (2024). AI literacy framework and strategies for implementation in developing nations. *Proceedings of the 16th International Conference on Education Technology and Computers, Portugal*. <https://doi.org/10.1145/3702163.3702449>
- Kim, Cheol. (2024). Research trends related to “AI ethics and education”. *Journal of Education and Culture*, 30(1), 23–41. <https://doi.org/10.24159/joec.2024.30.1.23>
- Kim, Dong-Hwi. (2025). The effects of AI literacy and ethical awareness on K-12 teachers' AI learning behavioral intentions. *Journal of Educational Innovation Research*, 35(1), 301–322. <https://doi.org/10.21024/pnuedi.35.1.202503.301>
- Kim, Eungyeong., & Lee, Youngjun. (2023). Development of AI ethics dilemma to enhance AI ethical competence. *The Journal of Korean Association of Computer Education*, 26(5), 31–42. <https://doi.org/10.32431/kace.2023.26.5.003>
- Kim, Miyoung., & Park, Sunju. (2023). Analysis of AI ethics research trends using text mining. *Journal of The Korean Association of Information Education*, 27(1), 23–33. <https://doi.org/10.14352/jkaie.2023.27.1.23>
- Koehler, M. J., Mishra, P., Akcaoglu, M. & Rosenberg, J. M. (2013). *The technological pedagogical content knowledge framework for teachers and teacher educators*. Commonwealth Educational Media Centre for Asia.
- Konika, S., Devi, R., & Singh, L. (2025). A comparative analysis of preservice teachers' attitudes toward artificial intelligence across demographic variables. *Education and Information Technologies*, 30, 1–22. <https://doi.org/10.1007/s10639-024-12874-3>
- Lee, In-Jae. (2022). A study on ways of AI ethics education using AI ethics guidelines. *Journal of Korean Elementary Moral Education*, 76, 157–180. <https://doi.org/10.17282/ethics.2022..76.157>
- Lee, Yonghee. (2023). Utilizing ChatGPT in writing poem in English. *The Jungang Journal of English Language and Literature*, 65(2), 23–46. <https://doi.org/10.18853/jjell.2023.65.2.002>
- Lodhi, S. S., & Lodhi, S. (2025). *Integration of AI in STEM education: Addressing ethical challenges in K-12 settings* (arXiv preprint No. 2510.19196). arXiv. <https://doi.org/10.48550/arXiv.2510.19196>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, USA*, 1-16. <https://doi.org/10.1145/3313831.3376727>
- Ma, M., Ng, D. T. K., Liu, Z., & Wong, G. K. W. (2025). Fostering responsible AI literacy: A systematic review of K-12 AI ethics education. *Computers and Education: Artificial Intelligence*, 8, Article 100422. <https://doi.org/10.1016/j.caeai.2025.100422>
- Maeng, Unkyoung. (2025). Blending AI and instructor feedback in EFL writing: Learners' motivation, attitude, and self-efficacy.

- English Teaching*, 80(5), 11–39. <https://doi.org/10.15858/engtea.80.5.202512.11>
- Mihoc, A., Popescu, D., & Ionescu, C. (2025). Student views on artificial intelligence ethics and social impact. *Computers & Education: Artificial Intelligence*, 6, 100210. <https://doi.org/10.1016/j.caeai.2024.100210>
- Ministry of Education. (2022, January 27). *Ethical principles for artificial intelligence in education*. <https://www.moe.go.kr/search/front/Search.jsp>
- Ministry of Science and ICT. (2020, November 27). *Human-centered artificial intelligence (AI) ethics standards (draft)* [Press release].
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- OECD. (2019). *Artificial intelligence in society*. OECD Publishing.
- Oh, Sun-Kyung, Jang, Mijung, & Park, Jung-eun. (2023). Undergraduates' awareness of the ethics of generative AI utilization in college writing. *The Korean Journal of Literacy Research*, 14(4), 69–96. <https://doi.org/10.37736/KJLR.2023.08.144.03>
- Park, Bo-kyung., & Yun, Ga-young. (2025). Topic modeling-based analysis of research trends and key issues in AI ethics in Korea. *Journal of Educational Technology*, 41(2), 633–663. <https://doi.org/10.17232/KSET.41.2.633>
- Peters, T. F. (2025). Machine intelligence, artificial general intelligence, super-intelligence, and human dignity. *Religions*, 16(8), 975. <https://doi.org/10.3390/rel16080975>
- Pei, B., Lu, J., & Jing, X. (2025). Empowering preservice teachers' AI literacy: Current understanding, influential factors, and strategies for improvement. *Computers and Education: Artificial Intelligence*, 8, 100406.
- Raman, R., Kowalski, R., Achuthan, K., Iyer, A., & Nedungadi, P. (2025). Navigating artificial general intelligence development: societal, technological, ethical, and brain-inspired pathways. *Scientific Reports*, 15, 8443. <https://doi.org/10.1038/s41598-025-92190-7>
- Ramesh, A., Dhariwal, P., Nichol, A., Chue, C., & Chen, M. (2022). *Hierarchical text-conditional image generation with CLIP latents*. arXiv. <https://arxiv.org/abs/2204.06125>
- Ryu, Miyoung., Cheon, Jongpil., & Han, Sungwan (2025). A review of ChatGPT's ethical factors for AI ethics education. *Journal of The Korean Association of Artificial Intelligence Education*, 6(1), 37–46. <https://doi.org/10.52618/aied.2025.6.14>
- Saghiri, A. M., Vahidipour, S. M., Jabbarpour, M. R., Sookhak, M., & Forestiero, A. (2022). A survey of artificial intelligence challenges: Analyzing the definitions, relationships, and evolutions. *Applied Sciences*, 12(8), 4054. <https://doi.org/10.3390/app12084054>
- Song, Aeri., & Ko Junga. (2023). A study on pre-service ethics teachers' perception types of AI ethics-A Q-methodology approach. *Journal of Moral & Ethics Education*, 80, 105-136. <https://doi.org/10.18338/kojmee.2023..80.105>
- UNESCO. (2019). *Preliminary study on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000367823>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- UNESCO. (2024). *Artificial intelligence and education: Ethics and policy implications*. UNESCO Publishing.
- Wang, X., Li, Y., & Chen, Z. (2025). Assessment of artificial intelligence ethical reflection among university students. *Computers & Education*, 198, 104768. <https://doi.org/10.1016/j.compedu.2024.104768>
- Winfield, A. F. T., & Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180085. <https://doi.org/10.1098/rsta.2018.0085>
- Wieczorek, M., Hosseini, M., & Gordigin, B. (2025). Unpacking the ethics of using AI in primary and secondary education: A systematic literature review. *AI and Ethics*, 5, 4693–4711. <https://doi.org/10.1007/s43681-025-00770-0>
- Yang, T., Cheon, J., Cho, M-H., & Cusson, N. (2025). Undergraduate students' perspectives of generative AI ethics. *International Journal of Educational Technology in Higher Education*, 23, 35. <https://doi.org/10.1186/s41234-025-00533-1>
- Zhai, C., Wibowo, S. & Li, L.D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learn. Environ*, 11, 28. <https://doi.org/10.1186/s40561-024-00316-7>

Appendix

Survey Questionnaire

Part I. Demographic Information

Please write the year you entered Graduate school/Undergraduate school and your current semester.

1. What is your major?
2. What is your gender? ① Male ② Female
3. Which age group do you belong to? ① 25 or under ② 26-30 ③ 31-35 ④ 36-40 ⑤ 41 or above
4. Have you ever used AI tools or AI learning platforms? ① Yes (Go to 4-1) ② No (Go to 5)
- 4-1. Please specify which AI tools/platforms you have used.
5. Have you ever taken any AI-related courses? ① Yes (Go to 5-1) ② No
- 5-1. Please write the title(s) of the AI-related course(s) you have taken.

Part II. AI Ethics

(1 = Strongly disagree · 2 = Disagree · 3 = Neutral · 4 = Agree · 5 = Strongly agree)

1. AI systems are fair to all users.
2. AI systems do not discriminate against specific groups.
3. In AI decision-making processes, individuals' backgrounds or characteristics do not result in disadvantages.
4. The use of AI technology is carried out in a socially responsible manner.
5. The use of data and AI systems complies with ethical standards.
6. The institutions or individuals responsible for negative outcomes resulting from AI use are clearly identified.
7. AI systems handle personal data securely.
8. Individuals' privacy is respected in the data collection and usage processes of AI systems.
9. The use of AI technology causes safety concerns.
10. AI technology has a positive impact on society.
11. The development of AI technology negatively affects the environment.
12. AI technology is beneficial for future generations.
13. AI-based educational platforms securely protect students' personal information.
14. Students' consent is appropriately obtained in the process of collecting and using learning data.
15. The types of data collected on learning platforms are appropriate for educational purposes.
16. AI-based learning assessment is fair to all students.
17. AI-based recommendation systems provide equal learning opportunities to all students.
18. Educational AI systems are implemented without bias toward particular groups of students.
19. AI-based educational platforms personalize content according to individual learning styles and needs.
20. AI-supported education improves learning efficiency.
21. Personalized learning pathways based on learner data increase students' learning motivation.
22. The way AI-based educational tools operate and make decisions (i.e., how the tools process information and produce final outcomes or recommendations) is sufficiently explained to students.
23. Students and teachers have opportunities to provide input on the settings and decisions of AI educational platforms.
24. There is a process to regularly evaluate and share the impact of AI educational tools.