컴퓨터교육학회 논문지 2026년 제29권 제5호
https://doi.org/10.32431/kace.2026.29.5.007

비판적 인식 기반 AI 교육으로의 전환: AI 리터러시, 지각된 윤리적 위험, 책임있는 사용 의도 간의 구조적 관계

Transitioning to Critical Awareness-Oriented AI Education: The Structural Relationships among AI Literacy, Perceived Ethical Risk, and Responsible Use IntentionGenerative AI-Based Vibe Coding: Applying Design-Based Research Methodology

편은진†
Eunjin Pyeon†

요약

본 연구는 AI 기반 평가 맥락에서 AI 리터러시, 지각된 윤리적 위험, 책임있는 사용 의도 간의 관계를 분석하였다. 대학(원)생 418명을 대상으로 경로분석을 실시한 결과, AI 리터러시는 지각된 윤리적 위험과 책임있는 사용 의도에 직접적인 영향을 미쳤다. 그러나, 지각된 윤리적 위험의 전체 수준은 AI 리터러시와 책임있는 사용 의도 사이에서 유의한 매개 역할을 하지 않았다. 반면, 책임성 및 구제 불가능 위험 차원에서만 유의미한 간접 효과가 나타나 위험 차원 간 차별적 작동이 확인되었다. 이는 AI 리터러시 교육이 윤리적 위험에 대한 선별적 접근을 바탕으로 공정성 점검, 책임성 이해, 이의제기 절차 학습을 포함하는 비판적 참여 중심으로 전환될 필요가 있음을 시사한다.

주제어 AI 리터러시, 지각된 윤리적 위험, 책임있는 사용 의도, AI 기반 평가, 비판적 인식

ABSTRACT

This study examined the relationships among AI literacy, perceived ethical risk, and responsible use intention in the context of AI-based assessment. Based on data from 418 university and graduate students, path analysis was conducted. The results showed that AI literacy had direct effects on both perceived ethical risk and responsible use intention. However, perceived ethical risk did not show a significant mediating effect in the relationship between AI literacy and responsible use intention. In contrast, significant indirect effects were observed only for the accountability and irremediability risk dimensions, suggesting that risk dimensions operate differently. These findings suggest that AI literacy education should consider a selective approach to ethical risk, including fairness auditing, understanding accountability, and learning procedures for contestation and recourse.

Keywords AI Literacy, Perceived Ethical Risk, Responsible Use Intention, AI-based assessment, Critical Awareness

†정회원 연세대학교 AI혁신연구원 연구교수

논문투고 2026년 01월 07일
심사완료 2026년 03월 19일
게재확정 2026년 04월 02일
발행일자 2026년 05월 20일

1. 서론

인공지능(AI) 기술의 급속한 발전과 함께 교육 현장에서는 자동 채점, 피드백, 과제 평가 등 AI를 활용한 교수학습 및 평가 방식이 빠르게 확산되고 있다. 특히, AI 기반 평가는 학습자의 학업 성취뿐만 아니라 진로, 향후 기회에까지 영향을 미칠 수 있는 고위험 맥락으로서, AI가 생성한 결과를 학습자가 어떻게 이해하고 판단하며 활용하는지가 중요한 교육적 과제로 부상하고 있다. 그러나 이러한 중요성에도 불구하고 AI가 제공하는 평가 결과가 학습자의 인식과 판단에 어떠한 메커니즘을 통해 영향을 미치는지에 대한 실증적 연구는 아직 충분히 축적되지 않은 상황이다.

기존의 AI 리터러시 교육은 AI의 개념 이해와 도구 활용 등 기술적·도구적 측면을 중심으로 논의되어 왔다[1, 2]. 그러나 이러한 논의는 실제 교육 현장에서 요구되는 수준, AI 시스템의 오류 가능성, 알고리즘 편향, 불투명성, 사회적 영향과 같은 윤리적 쟁점을 인식하고 이에 비판적으로 대응하는 역량까지 충분히 포괄하지 못한다는 한계가 지적되고 있다[3, 4]. 특히, AI 기반 평가 맥락에서는 이러한 윤리적 위험이 학습자의 학습 성취뿐 아니라 향후 진로 기회에도 영향을 미칠 수 있으며, 평가 영역에서는 그 파급력이 더욱 확대될 수 있으므로 보다 신중한 접근이 요구된다.

최근 연구들은 기술의 수용이나 활용을 넘어, 학습자의 AI 리터러시를 윤리적 쟁점과 위험을 비판적으로 이해하고 책임있게 대응하는 역량으로 확장하여 논의하고 있다[1, 5, 6]. 그러나 AI 기반 평가 맥락에서 AI 리터러시와 지각된 윤리적 위험, 책임있는 사용 의도 간의 구조적 관계를 통합적으로 분석한 실증 연구는 아직 매우 제한적이다.

이에 본 연구는 AI 기반 평가 상황에서 학습자의 AI 리터러시, 지각된 윤리적 위험, 책임있는 사용 의도 간의 영향 관계와 매개 효과를 분석하는 것을 목적으로 한다. 이를 통해 AI에 대한 지각된 윤리적 위험이 책임있는 참여로 이어지는 메커니즘을 실증적으로 규명하고, 비판적 인식에 기반한 AI 리터러시 교육 전환을 위한 이론적·실천적 시사점을 제시하고자 한다.

2. 이론적 배경

2.1 비판적 인식 기반의 AI 리터러시

AI 기술의 급속한 확산과 함께 AI 리터러시는 필수 핵심 역량으로 강조되고 있다. 초기 AI 교육은 AI의 개념 이해와 활용 역량을 중심으로 논의되어 왔으나, 최근에는 AI의 기술적 한계와 윤리적·사회적 영향을 이해하고 이에 비판적으로 참여하는 역량의 중요성이 더욱 강조되고 있다[2, 4]. 이러한 관점에서 AI 리터러시는 AI의 사회적 영향을 이해하고 결과를 비판적으로 해석하는 능력을 포함하는 다차원적 역량으로 규정할 수 있다[2]. 또한, 리터러시는 단순히 무엇을 아는 수준이 아니라, 실제 사용 상황에서 판단과 행동을 조절하는 수행적 능력이다[7]. 즉, AI의 작동 원리와 한계를 이해하고 이를 바탕으로 사용자가 AI의 윤리적 위험을 어떻게 인식하고 대응하는지를 결정하는 핵심적인 인지적 기반으로 작용할 수 있는 역량인 것이다[1, 2].

개인은 위험을 인식할 때 위험의 심각성뿐 아니라 이를 통제·대응할 수 있는 능력을 함께 평가하여 행동을 결정한다는 보호동기 이론(Protection Motivation Theory)에 따르면[8], AI 리터러시는 위험에 대한 대응 판단을 가능하게 하는 인지적 자원으로 기능한다. AI의 작동 원리와 한계를 충분히 이해하지 못한 상태에서는 오류나 불공정성과 같은 윤리적 위험이 막연한 불안이나 거부감으로 인식되기 쉽지만, AI 리터러시가 높은 경우에는 평가 결과의 오류 가능성을 이해하고 위험을 감정적 위협이 아닌 검토·관리의 대상으로 구조화하여 인식하게 된다. 결국, AI 리터러시는 윤리적 위험을 단순히 낮추는 역할이 아니라, 위험을 보다 분별력 있게 인식하고 주체적으로 대응하도록 만드는 역할을 수행할 수 있다.

위험에 대한 이러한 접근은 학습자의 AI 사용과 판단에도 중요한 영향을 미친다. 가치 기반 수용 모형(Value-based Adoption Model: VAM)에 따르면, 사용자는 기술의 지각된 편익과 희생을 종합적으로 평가하여 지각된 가치를 형성하고, 이러한 가치가 수용의도에 영향을 미친다[9]. 이러한 관점을 AI 평가 맥락에 확장하면, 윤리적 위험은 수용 과정에서 고려되는 희생 요인으로 해석될 수 있다. 특히 AI 리터러시가 높은 학습자는 이러한 위험을 단순한 억제 요인이 아니라 관리 가능한 비용으로 인식함으로써, 기술의 편익과 위험을 함께 고려한 책임있는 사용 의도를 형성할 가능성이 있다.

AI 리터러시의 개념은 Long & Magerko[1]와 Ng et al.[2]이 제시한 포괄적 논의를 바탕으로 이해할 수 있다. 이들 연구는 AI 리터러시를 AI의 작동 원리와 한계, 그리고 사회적·윤리적 영향에 대한 이해를 포함하는 다차원적 역량으로 제시한다. 다만, 본 연구에서는 지각된 윤리적 위험 및 책임있는 사용 의도와의 개념적 중복을 최소화하기 위해, AI 리터러시의 포괄적 구성 요소 전체를 측정하기보다 AI 기반 평가 맥락에서 데이터 의존성에 대한 이해, 시스템의 한계 및 불투명성 인지, 결과에 대한 비판적 성찰과 고려를 중심으로 위험을 인식하고 판단을 조정할 수 있는 인지적·비판적 인식 기반의 역량으로 제한하여 조작적으로 정의하였다.

이는 AI 리터러시의 윤리적 태도 요소와 지각된 윤리적 위험 간의 개념적 중복을 피하고, 위험 인식과 인지적·기술적 이해 능력을 검증하기 위한 의도적인 전략에 따른 것이다. 더욱이, 리터러시는 추상적인 지식 보유 여부가 아닌 수행적 능력이라는 점에서, AI가 무엇인지 아는 일반적 지식 수준이 아닌 AI의 데이터 의존성, 오류 가능성, 결과에 대한 무비판적 수용 위험을 인식하고 이를 비판적으로 해석하는 역량을 중심으로 포착될 필요가 있다. 수행 중심의 AI 리터러시는 윤리적 태도나 위험 인식 그 자체와 구분되면서 인지적 기반의 선행 개념으로 정교화 될 수 있으며, AI를 책임

있게 사용하려는 의도로 이어지는 메커니즘을 이론적으로 뒷받침한다.

2.2 지각된 윤리적 위험 개념 및 매개 기제

지각된 위험 관점에 따르면, 개인은 새로운 제품·서비스·기술을 수용할 때 기대되는 효익만이 아니라, 선택의 결과에 수반될 수 있는 불확실성과 잠재적 부정적 결과를 함께 평가하여 행동 의도를 형성한다[10, 11]. 후속 연구에서는 지각된 위험이 기술 수용을 저해하는 중요한 요인으로 다루어져 왔으며, 결과에 대한 불안감과 불확실성은 사용이나 구매와 같은 긍정적 행동 의도를 감소시키는 방향으로 작동하는 것으로 보고되었다[12]. 또한 신뢰가 사용 의도의 주요 결정요인으로 작용하는 동시에 지각된 위험을 낮추는 방향으로 작동하는 것으로 나타났다[13]. 이러한 맥락에서 위험은 주로 회피, 거부, 사용 중단과 같은 부정적 반응을 설명하는 요인으로 간주되어 왔다. 그러나 학습자의 성과, 기회, 진로에 직접 영향을 미치는 평가 영역에서의 위험 인식은 기능적 문제를 넘어 공정성, 책무성, 불투명성 등 윤리적 차원으로 재개념화될 필요가 있다. 특히 AI 기반 평가는 학습자의 권리와 공정성을 위협할 수 있는 문제를 포함한다는 점에서, 위험은 단순한 기능적 불편이 아니라 윤리적 위험으로 이해되어야 한다[14].

그러나, 지각된 위험이 항상 기술 수용을 저해하는 방향으로만 작동하는 것은 아니다. 일부 연구들은 위험이 적절히 인식될 경우, 오히려 사용자의 보호적이고 신중한 행동 의도를 촉진할 수 있음을 시사한다. 즉, 학습자가 AI의 오류 가능성이나 불공정성을 인식할수록 자신을 보호하려는 동기가 활성화될 수 있으며, 이는 AI 결과를 보다 신중하고 책임 있게 검토·사용하려는 의도로 이어질 수 있다[15]. 따라서 AI 기반 평가 맥락에서는 학습자가 이러한 위험을 인식할 때, AI 결과를 무조건 수용하기보다 이를 비판적으로 검토하고 판단을 유보하려는 태도가 강화될 가능성이 있다.

이러한 논리는 가치 기반 수용 모델(VAM)과 맥락을 같이 한다. 즉, 학습자는 AI의 유용성뿐만 아니라 윤리적 결함으로 인해 수반될 수 있는 부담과 위험을 하나의 희생 요인으로 평가하며, 이를 바탕으로 기술의 가치를 판단한다[9]. 이 과정에서 AI 리터러시는 지각된 윤리적 위험을 형성하고 해석하는 핵심적인 인지적 기반으로 작용하여, 위험을 막연한 불안이 아니라 관리 가능한 대상으로 재구성하고, 이는 결과를 비판적으로 검토하고 책임있게 사용하려는 의도를 강화한다. 특정 기술에 대한 지식 수준이 높을수록 그 지식에 대한 위험 인식이 증가하고, 그 결과 수칙 준수 행동으로 이어진다는 선행 연구 결과는 이러한 메커니즘을 실증적으로 지지한다[2]. 결국, 지각된 윤리적 위험은 AI 리터러시가 학습자의 비판적 인식을 실천적 책임 의식으로 전환하는 과정에서 매개 기제로 기능할 수 있음을 시사한다.

이에 본 연구에서는 지각된 윤리적 위험 개념을 정교화하기 위한 이론적 토대로 EU의 AI 윤리 가이드라인과 알고리즘 수용 관련 선행연구를 활용하였다. 이들 논의는 공정성, 책임성, 투명성, 프라이버시 보호, 인간 감독을 핵심 원리로 제시하며, 공정성과 설명가능성에 대한 우려가 형성되는 배경에 대한 이해를 제공한다[16, 17]. 또한 평가와 같은 고위험 영역에서는 이러한 윤리적 위험을 보다 체계적으로 다룰 필요가 있음을 시사한다. 특히, FATE (Fairness, Accountability, Transparency, Explainability)는 지각된 윤리적 위험의 하위 차원을 측정 가능한 구성 개념으로 구조화하는 데 필요한 개념적 근거를 제공하며[18], 원칙을 시스템 운영과 거버넌스 수준에서 관리해야 할 위험 요소로 구체화한다[19]. 이에 본 연구는 FATE와 AI RMF(Risk Management Framework) 1.0 프레임워크를 근거로, AI 기반 평가 맥락에서 지각하는 윤리적 위험을 4가지 차원으로 구체화하고 그에 대한 조작적 정의를 Table 1에 제시하였다.

4가지 차원의 지각된 윤리적 위험은 AI 기반 평가 시스템에 갖추어야 할 윤리적 토대를 구성하는 동시에, AI 리터러시의 핵심적인 매개 기제 역할을 보여준다. 예를 들어, 평가 결과에 대한 설명을 충분히 안내받지 못하거나 AI 시스템의 오류로 인해 학습 기회가 제한되는 상황은 학습자의 성장과 교육 기회, 미래 성장을 훼손할 수 있기 때문에, 사용자가 지각한 윤리적 위험은 단순한 부정적 인식에 그치지 않고 기술을 책임있게 사용해야 할 필요성을 지각하게 만드는 인지적·동기적 매개 효과로 작동할 수 있다. 아울러, AI 리터러시는 알고리즘의 한계를 인식하고 AI가 생성한 결과에 대해 인간 중심의 의사결정과 신중하고 비판적으로 검토하는 역량의 필요성을 제기한다. 이러한 역량을 갖춘 학습자는 윤리적 위험을 인식하지 못하는 상황에서도 AI 결과를 무조건 신뢰하지 않고 책임있는 사용 태도를 유지할 가능성이 높기 때문이다.

Table 1. Operational Definitions of Four Dimensions of Perceived Ethical Risk

FATE & AI RMF 1.0	Perceived Ethical Risk	Operational Definition
Fairness	Fairness and bias risk	Risk that the AI-based assessment system may produce unjustly unfavorable or discriminatory outcomes for certain individuals or groups
Transparency & Explainability	Opacity and incomprehensibility risk	Risk that learners find it difficult to understand and accept the process or rationale by which the AI-based assessment system produces results
Accountability	Accountability and irremediability risk	Risk of unclear responsibility and lack of remediation procedures when errors occur in the AI-based assessment system
Ethics & Privacy	Data privacy and violation risk	Risk of privacy infringement and inappropriate use of personal data in the AI-based assessment system

2.3 책임있는 사용 의도

AI 기반 평가 맥락에서 사용자의 판단과 행동을 설명하는 데에는 기존 기술 수용 연구에서 주로 사용되어 온 사용

의도 개념만으로는 한계가 있다. 기존 연구는 기술의 사용 여부에 초점을 두는 경향이 있으며, 지각된 위험 역시 대체로 수용을 저해하는 요인으로 다뤄져 왔다[13]. 그러나 AI 기반 평가에서 학습자가 지각하는 위험은 공정성, 책임성, 개인정보 보호 등 도덕적 판단을 요구하는 윤리적 차원을 포함한다는 점에서, 단순히 사용할 것인가의 문제를 넘어 AI를 어떻게 사용할 것인가가 더욱 중요해진다[14]. AI 윤리 논의는 책임있게 사용하는 방식에 초점을 두고 있으며, 기술 적용 맥락에 따라 책임있는 사용 의도의 중요성과 의미의 깊이가 달라질 수 있다. 위험을 안다고 해서 모두가 책임 있게 행동할 수 있는 것은 아니므로 학습자가 해당 위험에 어떻게 대응해야 할지에 대한 이해 여부에 따라 책임있게 사용하려는 의도가 달라질 수 있기 때문이다. 선행 연구에서도 AI를 많이 사용하는 것보다 사용 자체를 비판적으로 검토하려는 의도가 책임있는 사용에 필수적이며, AI가 산출한 결과를 무조건 따르지 않고 검토하며, 문제 발생 시 주체적으로 행동하려는 태도에 대해 집중함을 보여준다[20, 21].

이에 본 연구에서는 AI 책임성 및 거버넌스, 인간의 감시와 검토 가능성, 책임있는 정보시스템 사용, 그리고 자동화 평가에 대한 비판적 검토를 다룬 관련 선행 연구를 바탕으로[20-23], 책임있는 사용 의도를 단순한 AI 사용 여부나 사용 정도를 넘어, AI 기반 평가 상황에서 실제 사용 행동에 앞서 AI의 한계와 윤리적 위험을 인지하고 규범적 기준에 따라 판단하며 필요 시 개입하려는 규범적·성찰적 참여(Normative and Reflective Engagement) 의도로 조작적으로 정의하였다. 이는 AI를 도구적 역할로 인식하고 기술을 채택할 의향을 의미하는 기존의 사용 의도와 구별되는 고차원의 인지적 참여(engagement) 개념이다. 또한, 책임있는 사용 의도는 규범적 준수와 비판적 성찰을 포괄하는 단일 구성개념으로 개념화하였다. AI 평가 맥락에서 책임있는 사용 의도는 학습자가 위험을 어떻게 이해하고 대응하는지가 핵심이며, AI가 인간의 판단을 대체하기보다 강화시킨다는 인간 중심의 AI 교육 관점과 긴밀히 연결된다[24, 25].

교육적 맥락에서 바이브코딩의 핵심적 차별성은 세 가지이다. 첫째, 커스터마이징의 자유도이다. 노코드 플랫폼은 사전 설계된 기능 범위 내에서만 활용이 가능하지만, 바이브코딩은 교사의 자연어 요구사항을 AI가 완전한 소스 코드로 변환하므로 기존 도구로 충족할 수 없는 맥락 특화 기능의 구현이 가능하다. 본 연구에서 교사들이 요구한 6가지 핵심 기능을 단일 노코드 도구로 충족시킬 수 없었던 반면, 바이브코딩은 이를 통합적으로 구현할 수 있었다.

둘째, 반복적 수정의 신속성이다. DBR의 핵심 원리인 반복적 정교화를 실현하기 위해서는 사용자 피드백을 신속하게 반영할 수 있어야 한다. 노코드 플랫폼에서의 수정은 플랫폼이 허용하는 범위로 제한되지만, 바이브코딩에서는 자연어 프롬프트를 통해 코드 수준의 수정이 가능하므로 교사 피드백을 당일 또는 수일 내에 반영할 수 있었다.

셋째, 소스 코드의 소유와 이식성이다. 노코드 플랫폼은 벤더 종속(vendor lock-in)의 위험이 있어 서비스 중단 시 교육 도구의 지속성이 위협받는다. 바이브코딩은 완전한 소스 코드를 생성하므로 다른 환경으로의 이식이 가능하며, 이는 교육 기술의 장기적 지속가능성 측면에서 중요한 이점이다.

다만, 바이브코딩은 AI 환각(hallucination)으로 인한 코드 오류, 복잡한 기능에서의 예측 불가능한 동작, 비전공자의 디버깅 어려움 등의 한계를 가진다. 본 연구에서도 좌표 스케일링 문제, 드로잉 데이터 저장 실패 등의 기술적 문제가 발생하였으며, 이는 4.1절에서 상세히 기술한다.

3. 연구 방법

3.1 연구 모형 및 가설

본 연구의 목적을 달성하기 위해 설정한 연구 모형은 Fig. 1과 같다.

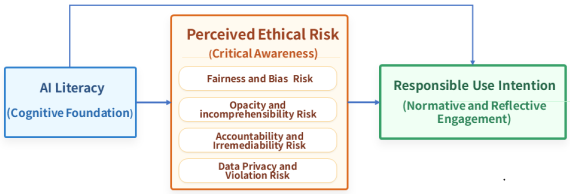

Figure 1. Research Model

본 연구 모형에서 AI 리터러시는 AI 기반 평가 상황에서 발생할 수 있는 윤리적 위험을 식별하고 평가할 수 있는 인지적 기반으로 개념화되며, 지각된 윤리적 위험은 4개 하위 요인을 통합한 상위 수준으로 개념화된다. 이에 본 연구에서는 지각된 윤리적 위험의 총효과를 개별 문항의 평균이 아닌 4개 하위 요인을 각 하위 요인의 평균 점수를 산출한 후, 이들 평균값의 평균을 계산한 합성 점수를 사용하였다. 이는 총효과 가설과 하위 요인별 가설 간의 개념적 일관성을 유지하기 위한 것이다. 책임있는 사용 의도 문항들은 AI를 실제 사용한 이후의 행동을 측정하기보다, AI 사용 행동이 발생하기 이전 단계에서 학습자의 비판적 감시, 규범적 판단과 인지적 참여 성향을 측정하도록 구성되었다.

연구 모형에 따른 연구 가설은 다음과 같다.

가설 1. AI 리터러시는 책임있는 사용 의도에 유의미한 영향을 미칠 것이다.

가설 2. AI 리터러시는 지각된 윤리적 위험에 유의미한 영향을 미칠 것이다.

가설 3. 지각된 윤리적 위험은 책임있는 사용 의도에 유의미한 영향을 미칠 것이다.

가설 4. 지각된 윤리적 위험은 AI 리터러시와 책임있는 사용 의도 간의 관계를 유의미하게 매개할 것이다.

가설 5. AI 리터러시는 지각된 윤리적 위험 각 요인에 유의미한 영향을 미칠 것이다.

가설 6. 지각된 윤리적 위험 각 요인들은 책임있는 사용 의도에 유의미한 영향을 미칠 것이다.

가설 7. 각 하위 요인은 AI 리터러시와 책임있는 사용 의도 간 관계를 유의미하게 매개할 것이다.

3.2 연구 대상

본 연구의 대상자는 AI가 학습 성과 평가나 피드백에 활용되는 상황을 경험했거나, 그러한 맥락을 충분히 이해할 수 있는 대학(원)생으로 구성하였다. 본 연구는 특정 AI 평가 시스템의 효과성 검증이 아니라 AI 기반 평가 환경에서 학습자가 자신의 평가에 직결되는 실제 상황에서 발휘되는 윤리적 태도와 AI 리터러시를 측정하는데 목적이 있다. 이에 대학(원)생은 이러한 평가 맥락과 밀접하게 연관된 적합한 연구 대상자로 판단하였다. 조사 대상자는 전국 대학교의 교수들을 통해 설문 참여자를 모집하였으며, 해당 교수들의 수업을 수강하는 대학(원)생들을 대상으로 설문을 실시하였다. 총 응답자 수는 418명이었다. 응답자의 일반적 특성은 Table 2와 같다.

Table 2. Demographic Characteristics of the Respondents (N=418)

Category		Frequency (%)
Gender	Male	212 (50.72)
	Female	206 (49.28)
Grade	University	342 (81.82)
	Graduate	76 (18.18)
AI System Experience	None	221 (52.87)
	Experienced	197 (47.13)

AI 기반 평가 시스템에 대해 경험이 없는 응답자 수가 경험이 있는 응답자 수보다 다소 높았다. 그러나, 본 연구는 특정 AI 기반 평가 시스템의 실제 사용 경험에 따른 차이를 비교하기보다, AI 기반 평가 맥락에서 학습자가 형성하는 인식과 책임있는 사용 의도 간의 관계를 분석하는 데 초점을 두었다. 또한 AI 기반 평가는 아직 현장에 점진적으로 도입되는 단계에 있어, 직접적인 사용 경험이 없더라도 교육이나 매체를 통해 관련 인식과 판단이 형성될 수 있다고 보았다. 이에 본 연구에서는 경험 유무와 관계없이 해당 맥락을 이해할 수 있는 학습자를 연구 대상에 포함하였다. 다만 경험 유무에 따라 인식과 의도 형성 양상이 달라질 가능성이 있으므로, 이는 본 연구의 한계로 함께 고려하였다.

3.3 측정 도구

지각된 위험 이론, AI 윤리 프레임워크 등 이론과 선행연구에 근거하여 측정 요소들을 추출하고 측정 지표를 AI 기반 평가 맥락에 맞게 재구성하였다. 연구 변수별 측정 지표와 관련 출처는 Table 3과 같다.

'AI 리터러시'는 Long & Magerko[1]와 Ng et al.[2]의 개념을 참고하여 구성하였으며, 본 연구에서는 AI 리터러시의 포괄적 역량 전체를 측정하기보다 AI 기반 평가 맥락에서의 데이터 의존성 이해, 시스템의 오류 가능성 및 불투명성 인지, 결과에 대한 비판적 고려와 같은 인지적·비판적 인식 측면에 초점을 두어 제한적으로 조작화하였다. 이에 따라 AI 리터러시는 해당 이해 수준을 반영하는 3개 문항으로 구성하였다(예: AI 기반 평가 시스템의 성능은 시스템이 학습한 데이터의 품질에 크게 의존한다는 것을 알고 있다). 또한 본 연구의 AI 리터러시는 객관적 수행 능력 자체를 평가하는 것이 아니라, AI 기반 평가 상황에서 학습자가 지닌 자기보고식 이해 수준과 비판적 판단 가능성을 측정한 것이다.

Table 3. Measurement Variables and Indicators

Variable	Measurement Indicators (Number of Items)	Source
AIL	· the capacity to actively and critically engage with AI-based assessment systems (3)	Long & Magerko[1] Ng et al.[2]
Risk	· Risk of fairness and bias (3) · Risk of opacity and incomprehensibility (3) · Risk of accountability and lack of redress (3) · Risk of data and privacy infringement (3)	Zhu et al.[26] Featherman & Pavlou[12]
RUI	Intention for normative and ethical compliance, critical monitoring and intervention (6)	Lowry & Moody[27] Floridi et al.[28] Veale et al.[29]

Note. AIL = AI literacy; Risk = perceived ethical risk; RUI = responsible use intention.

'지각된 윤리적 위험'은 Zhu et al. [26], Featherman & Pavlou[12]의 선행연구 문항을 바탕으로 AI 기반 평가 맥락에 맞게 수정·보완하여 총 4개 차원, 차원별 3문항씩 12개 문항으로 재구성하였다(예: AI 기반 평가 시스템이 어떤 기준과 방식으로 평가를 진행하는지 이해하기 어렵다). 본 연구에서 지각된 윤리적 위험은 공정성 및 편향성 위험(p1)은 pr1-pr3, 불투명성 및 이해불가능성 위험(p2)은 pr4-pr6, 책임성 및 구제 불가능성 위험(p3)은 pr7-pr9, 데이터 및 프라이버시 침해 위험(p4)은 pr10-pr12로 구성되었다.'책임있는 사용 의도'측정은 총 6개 문항을 사용하였다. 문항 구성에는 Lowry & Moody[27]의 규범 및 절차 준수 의도, Floridi et al.[28]이 제시한 책임있는 AI 사용과 인간 감독의 중요성, 그리고 Veale et al.[29]이 논의한 알고리즘 의사결정에서의 책임성과 이의제기 가능성을 참고하였다(예: 나는 AI 기반 평가 시스템 이용 시 제시된 윤리적 규범이나 절차를 충실히 준수할 것이다).

설문 문항은 설문 동의 확인 문항 2개, 응답자 의 일반적 특성(성별, 학위 과정, 시스템 경험 여부 등) 5개 문항, 연구 모형 검증을 위한 문항 21로 구성하였다. 그리고, 모든 문항은 5점 척도(1점 전혀 그렇지 않다~5점 매우 그렇다)이다. 설문지의 내용 타당도를 확보하기 위해 AI 기반 평가 시스템 개발 또는 심사, 수업 경험이 있는 교수, 박사급 책임연구원 등 전문가 6명에게 검토를 의뢰하였다. 전문가 정보는 Table 4와 같다.

Table 4. Expert Group Information

ID	Degree	Current Position Affiliation	Experience	Major
A	Ph.D.	Professor, University H	15 years	Computer Engineering
B	Ph.D.	Professor, University N	10 years	Educational Technology
C	Ph.D.	Senior Researcher, University B	13 years	Educational Technology
D	Ph.D. Candidate	Researcher, University C	13 years	Educational Evaluation
E	Master's	Teacher, Elementary School Y	16 years	Moral Education
F	Master's	Manager, Company E	7 years	AI Education

전문가들은 각 문항의 적합성을 4점 척도(1=부적합, 4=매우 적합)로 평가하였으며, 문항 내용 타당도 지수(CVI)를 산출한 결과, 최종 선정된 문항은 CVI=0.83 이상이었다. 또한, AI 기반 평가 시스템 경험이 있는 피평가자 42명을 대상으로 파일럿 테스트를 실시하였고, 설문지의 신뢰도 Cronbach's α값은 0.901이었고, 경험이 없는 응답자를 고려하여 '~같다', '~이다' 등의 표현을 수정하였다.

3.4 자료 수집 및 분석

자료 수집은 MS Form을 이용한 온라인 설문으로 조사했으며, 결측치는 발견되지 않았다. 본 연구의 가설 검증을 위해 jamovi (v.2.7.11) 통계 프로그램을 활용하여 다음과 같은 절차로 자료를 분석하였다. jamovi는 R의 lavaan 패키지를 기반으로 구조모형 분석을 수행하는 통계 프로그램이다[30].

첫째, 연구 모형에 포함된 변수들의 이상치, 결측치, 정규성을 검토하고, 연구 대상자 특성 파악을 위해 기술통계와 주요 변수 간의 상관관계를 분석하였다. 둘째, Anderson & Gerbing(1988)[31]이 제안한 2단계 접근법에 따라, 구조모형 분석에 앞서 확인적 요인분석을 실시하여 측정 모형의 타당성과 신뢰도를 먼저 검증하였다. 셋째, 연구 가설에 따른 구조적 관계를 규명하기 위해 초기 분석 단계에서 구조방정식모형의 적용 가능성을 검토하였다. 그러나 문항 수준의 잠재변수 모형을 적용한 결과, 전반적인 모형 적합도가 수용 기준을 충족하지 못하였다. 이에 확인적 요인분석을 통해 측정 타당성을 우선 검증한 후, 구성 개념별 합성 점수를 활용한 경로분석을 최종 분석 방법으로 채택하였다. 이는 구성 개념 간 관계를 보다 안정적으로 추정하기 위한 것이다. 넷째, 두 가지 모형을 설정하여 경로분석을 실시하였다. 모형 A는 지각된 윤리적 위험을 하나의 단일 요인 모형으로, 지각된 윤리적 위험의 전체 효과를 검증하기 위해 4개 하위 요인을 통합한 합성 점수를 활용하였다. AI 리터러시가 윤리적 위험 전체를 거쳐 책임있는 사용 의도에 미치는 매개 효과(가설 1~4)를 검증하였다. 모형 B는 지각된 윤리적 위험의 4개 하위 요인을 각각 매개변수로 설정, 종속 변수에 미치는 개별적 영향력과 다중 매개 효과(가설 5~7)를 분석하였다. 본 연구의 경로 모형은 모든 가능한 경로가 포함된 포화모형이므로 모형 적합도 지수는 산출되지 않으며, 개별 경로 계수의 통계적 유의성과 모델의 설명력()을 중심으로 가설을 검증하였다. 아울러 매개 효과의 유의성을 검증하기 위해 부트스트래핑(5,000회)을 실시하고, 95% 신뢰구간 내에 0을 포함하지 않는 경우를 통계적으로 유의한 것으로 판단하였다.

4. 연구 결과

4.1 주요 변수의 기술통계 및 상관관계

4.1.1 기술 통계

기술통계 분석 결과는 Table 5와 같다.

Table 5. Descriptive Statistics

V	M.V	Mean	S.E	Skewness	Kurtosis
AIL	a1	3.96	0.04	−0.76	0.76
	a2	4.00	0.04	−0.83	0.94
	a3	3.83	0.04	−0.46	−0.20
Risk	pr1	3.37	0.06	−0.35	−0.84
	pr2	2.91	0.06	0.10	−0.79
	pr3	3.16	0.06	−0.01	−0.87
	pr4	3.27	0.05	−0.23	−0.66
	pr5	3.28	0.05	−0.21	−0.76
	pr6	3.26	0.06	−0.30	−0.89
	pr7	3.83	0.05	−0.79	−0.02
	pr8	3.67	0.05	−0.60	−0.39
	pr9	3.13	0.05	−0.26	−0.82
	pr10	3.21	0.06	−0.24	−0.85
	pr11	3.05	0.06	−0.02	−0.85
	pr12	3.31	0.04	−0.31	−0.78
RUI	rui1	4.10	0.04	−0.80	0.52
	rui2	4.19	0.04	−0.90	0.76
	rui3	4.21	0.04	−0.89	0.70
	rui4	4.17	0.04	−0.81	0.40
	rui5	4.04	0.04	−0.87	0.82
	rui6	4.22	0.04	−0.70	−0.14

Note. AIL = AI literacy; Risk = perceived ethical risk; RUI = responsible use intention; V = Variable; M.V = Measurement Variable; S. E. = Standard Error

모든 문항에서 평균과 표준오차는 적절한 분포를 보였다. 정규성 검토를 위해 왜도와 첨도를 확인한 결과, 왜도 절댓값이 2 미만, 첨도 절댓값이 7 미만으로[32], 정규성 가정을 충분히 만족하는 것으로 판단하였다.

4.1.2 상관관계

주요 변수 간 상관관계는 Table 6과 같다.

Table 6. Correlation Matrix

	AIL	Risk	RUI	Cronbach's α
AIL	—			0.733
Risk	.183***	–		0.857
RUI	.530***	.099	—	0.903

Note. AIL = AI literacy; Risk = perceived ethical risk; RUI = responsible use intention. ***p < .001, *p < .05.

상관관계 분석 결과, AI 리터러시는 지각된 윤리적 위험, 책임있는 사용 의도와 유의미한 정적 상관을 보였다. 이는 AI 기반 평가 맥락에서 AI 리터러시 수준이 높을수록 윤리적 위험을 더 많이 인식하는 경향이 있고, 책임있는 사용 의도 역시 높아지는 경향이 있음을 시사한다. 반면, 지각된 윤리적 위험과 책임있는 사용 의도 간의 상관은 통계적으로 유의하지 않았다. 이는 지각된 윤리적 위험이 책임있는 사용 의도에 직접적으로 연결되지 않을 가능성을 시사한다. 한편, 각 변수의 Cronbach's α값 모두 수용 가능한 수준의 내적 일관성을 보였다.

4.2 확인적 요인분석

본 연구에서는 측정 모형의 타당성 검증을 위해 확인적 요인분석을 실시하여 문항의 요인 적재량과 수렴 타당성을 확인하였다. 각 요인의 신뢰도, 타당도 검증 결과는 Table 7과 같다.

확인적 요인분석을 실시한 결과, AI 리터러시의 모든 문항은 요인적재량이 0.40 이상으로 나타났으며, 모두 $p < .001$ 수준에서 통계적으로 유의하였다. 지각된 윤리적 위험과 책임있는 사용 의도 역시 모든 문항이 해당 잠재변수를 안정적으로 설명하는 것으로 나타났다. 다만, pr3 문항은 상대적으로 낮은 요인적재량을 보였으나, 공정성 및 편향성 차원의 내용타당도를 확보하기 위해 유지하였다. 또한 평균분산추출(AVE)이 일부 기준치에 미치지 못하였으나, 구성신뢰도(CR)값이 기준을 충족하고 요인적재량이 전반적으로 양호한 수준을 보여 수렴 타당성은 수용 가능한 수준으로 판단하였다. 따라서 각 구성 개념의 측정 타당성은 전반적으로 확보된 것으로 해석된다.

Table 7. Results of Measurement Model Analysis

	M.V	S.F.L	Cronbach's α	CR	AVE
AIL	a1	0.647	0.733	0.63	0.37
	a2	0.665			
	a3	0.489			
Risk	pr1	0.502	0.857	0.90	0.44
	pr2	0.622			
	pr3	0.405			
	pr4	0.697			
	pr5	0.782			
	pr6	0.766			
	pr7	0.651			
	pr8	0.683			
	pr9	0.643			
	pr10	0.678			
	pr11	0.680			
	pr12	0.710			
RUI	rui1	0.603	0.903	0.82	0.42
	rui2	0.651			
	rui3	0.697			
	rui4	0.626			
	rui5	0.666			
	rui6	0.659			

Note. AIL = AI literacy; Risk = perceived ethical risk; RUI = responsible use intention; M.V = Measurement Variable; S.F.L = Standard Factor Loading; CR = Composite Reliability; AVE = Average Variance Extracted

4.3 경로분석

타당성과 신뢰도를 확보한 변수들을 바탕으로 연구 가설 검증을 위한 경로분석을 실시하였다. 경로분석은 지각된 윤리적 위험의 총효과를 분석하기 위한 모형 A(가설 1~4), 지각된 윤리적 위험의 하위 요인별 효과를 분석하기 위한 모형 B(가설 5~7)로 구분하여 수행하였다. 또한, 확인적 요인분석을 통해 타당성이 확보된 구성 개념의 합성 점수를 사용하여 변수 간 구조적 관계를 추정하였다. 이 경우 변수 간 가능한 모든 구조적 경로가 모형에 포함되므로 경로분석은 포화모형 형태로 추정되었다.

4.3.1 모형 A 분석

모형 A는 지각된 윤리적 위험을 하나의 매개 요인으로 설정하고 AI 리터러시와 책임있는 사용 의도 간의 관계와 매개 효과에 대한 것이다. 경로 계수의 크기와 통계적 유의성을 중심으로 모형 A에서 설정한 가설 1~4의 검증 결과는 Table 8과 같다.

Table 8. Hypothesis Testing Results for Model A

H	Path	B	S.E	β	p	Results
H1	AIL→RUI	0.504	0.066	0.531	〈.001***	Supported
H2	AIL→Risk	0.200	0.081	0.192	.014*	Supported
H3	Risk→RUI	−0.002	0.056	−0.003	.966	Rejected
H4	AIL→Risk→RUI	0.000	0.012	−0.000	.968	Rejected

Note. AIL = AI literacy; Risk = perceived ethical risk; RUI = responsible use intention; B = Unstandardized Coefficient; S.E = Standard Error; β = Standardized Path Coefficient, Hypothesis 4 represents the bootstrapped indirect effect, ***p 〈 .001, **p 〈 .01, *p 〈 .05.

분석 결과, AI 리터러시는 지각된 윤리적 위험의 약 3.7%(=0.037)를 설명하였으며, 설명력은 크지 않으나, 95% 신뢰 구간이 [.003, .103]으로 0을 포함하지 않아 AI 리터러시가 지각된 윤리적 위험에 유의미한 영향을 미침을 확인하였다. 반면, AI 리터러시와 지각된 윤리적 위험은 책임있는 사용 의도의 약 28%(R2=0.281)를 설명하며 비교적 높은 설명력을 보였다.

경로분석 결과, AI 리터러시는 책임있는 사용 의도에 유의미한 정적 영향을 미쳐 가설 1이 채택되었으며(β = 0.531, p < .001), 지각된 윤리적 위험에 대해서도 유의미한 영향을 보여 가설 2가 채택되었다. 반면, 지각된 윤리적 위험이 책임있는 사용 의도에 미치는 영향은 유의미하지 않았고, 부트스트래핑 (5000회) 결과 매개 효과의 95% 신뢰구간이 [-0.024, 0.025]으로 0을 포함하여 가설 3과 4는 기각되었다. 이는 AI 리터러시가 윤리적 위험 인식을 높이지만, 높아진 인식이 책임있는 사용 의도로 직접 연결되지않음을 시사한다.

4.3.2 모형 B 분석

모형 B는 지각된 윤리적 위험의 하위 요인들이 어떤 경로를 통해, 얼마나 다르게 작동하는가를 밝히는 것이다. 분석 결과, AI 리터러시는 지각된 윤리적 위험 하위 요인의 설명력(R2)은 낮았으나, 하위 요인과 AI 리터러시가 책임있는 사용 의도의 약 38%(R2 = 0.377)를 설명하는 것으로 나타났다. 하위 요인별 경로 계수 및 가설 검증 결과는 Table 9와 같다.

AI 리터러시는 공정성 및 편향성 위험(p1), 책임성 및 구제 불가능 위험(p3)에만 유의미한 영향을 미쳤다. 이는 AI 리터러시가 모든 위험 요인을 동일하게 증폭시키지 않고, 특성 위험 차원에서만 선택적으로 작용함을 시사한다. 한편, 지각된 윤리적 위험이 책임있는 사용 의도에 미치는 영향은 데이터 및 프라이버시 침해 위험(p4)을 제외한 하위 요인 모두 유의미하게 나타났다. p1, p2 요인은 위험 인식이 높을수록 책임있는 사용 의도가 감소하였으나, p3 요인은 위험 인식이 높아질수록 책임있는 사용 의도가 오히려 증가하였다. 이는 AI 기반 평가 맥락에서 학습자들이 위험을 인식하더라도 회피하기보다, 결과에 대한 책임있는 사용과 개입을 강화하려는 태도를 형성함을 의미한다. 특히, 총효과 분석에서 확인되지 않았던 매개 기제가 특정 위험(p3) 차원에서 선택적으로 작동하였다.

Table 9. Hypothesis Testing Results for Model B

H	Path	B	S.E	β	p	Results
H5	AIL→p1	0.272	0.089	0.241	.002**	Supported
	AIL→p2	0.179	0.094	0.152	.057	Rejected
	AIL→p3	0.326	0.090	0.252	〈.001***	Supported
	AIL→p4	0.081	0.107	0.062	.445	Rejected
H6	p1→RUI	−0.110	0.051	−0.128	.030*	Supported
	p2→RUI	−0.130	0.061	−0.158	.032*	Supported
	p3→RUI	0.206	0.051	0.274	〈.001***	Supported
	p4→RUI	0.019	0.052	0.026	.710	Rejected
H7	AIL→p1→RUI	−0.030	0.018	−0.031	.089	Rejected
	AIL→p2→RUI	−0.023	0.018	−0.024	.197	Rejected
	AIL→p3→RUI	0.067	0.023	0.069	.003**	Supported
	AIL→p4→RUI	0.002	0.007	0.002	.823	Rejected

Note. AIL = AI literacy; Risk = perceived ethical risk; RUI = responsible use intention; B = Unstandardized Coefficient; S.E = Standard Error; β = Standardized Path Coefficient, ***p 〈 .001, ** p 〈 .01, *p 〈 .05.

5. 결론 및 논의

본 연구는 AI 기반 평가 맥락에서 AI 리터러시가 지각된 윤리적 위험을 매개로 책임있는 사용 의도에 미치는 관계를 분석하였다. 분석 결과, AI 리터러시는 지각된 윤리적 위험 및 책임있는 사용 의도와 유의미한 정적 관련을 보이는 것으로 나타났다. 이는 AI 리터러시가 단순한 기술 이해를 넘어, 윤리적 위험을 인식하고 판단하는 인지적 기반으로 기능할 가능성을 보여준다. 또한 이러한 결과는 윤리적 위험을 단일한 억제 요인으로 해석하기보다, 차원별로 상이하게 작용하는 구조적 변수로 이해할 필요가 있음을 시사한다.

모형 A 분석 결과, AI 리터러시는 윤리적 위험 인식과 책임있는 사용 의도에 모두 유의미한 영향을 미쳤으나, 윤리적 위험은 책임있는 사용 의도에 직접적 또는 매개 효과를 보이지 않았다. 이는 위험 인식 자체가 곧바로 책임있는 행동으로 이어지지 않으며, 위험이 단순한 억제 요인이 아니라 해석과 대응 과정을 거쳐야 의미를 갖는 인지적 자원으로 작용함을 시사한다. 결국 책임있는 사용은 위험 인식 그 자체보다, 이를 비판적으로 해석하고 대응할 수 있는 AI 리터러시에 의해 설명된다고 볼 수 있다.

모형 B에서는 AI 리터러시가 특정 위험 차원에만 선택적으로 영향을 미쳤으며, 위험 차원에 따라 책임있는 사용 의도에 미치는 영향 방향이 다르게 나타났다. 특히 일부 위험은 책임있는 사용을 감소시키는 반면, 책임성 관련 위험은 오히려 이를 강화하는 것으로 나타나, 학습자는 위험을 단순히 회피하기보다 특정 상황에서는 적극적 개입으로 대응함을 시사한다.

또한 각 하위 요인의 매개 효과를 분석한 결과, 전체 수준

에서는 매개 효과가 나타나지 않았으나, 책임성 및 구제 불가능 위험 차원에서만 유의미한 간접효과가 확인되었다. 이는 윤리적 위험이 단일한 구성으로 작동하지 않으며, 특정 위험 차원이 AI 리터러시를 책임있는 행동으로 전환하는 핵심 경로로 기능함을 시사한다. 이는 윤리적 위험을 단일한 억제 요인이 아닌, 차원에 따라 상이한 행동 반응을 유도하는 구조적 변수로 재개념화할 필요성을 제기한다.

한편, 본 연구 결과는 AI 리터러시와 지각된 윤리적 위험의 측정 구성 특성을 고려하여 해석할 필요가 있다. 문항 수와 측정 밀도의 비대칭성은 AI 리터러시가 지각된 윤리적 위험을 설명하는 설명력이 낮게 나타난 결과에 영향을 미쳤을 가능성이 있다. 그러나 이러한 측정 비대칭성에도 불구하고 특정 위험 차원에서만 유의미한 경로와 기능적 매개 양상이 나타났다는 점은, AI 리터러시가 특정 윤리적 위험을 선별적으로 설명하고 있음을 시사한다.

이러한 결과를 바탕으로 구체적인 교육적·실천적 시사점을 제시하면 다음과 같다. AI 리터러시 교육은 단순한 기능 이해를 넘어, 공정성 및 편향성 위험을 식별하고 평가 결과의 형평성을 점검하는 활동을 포함할 필요가 있다. 또한 학습자가 AI 결과의 오류나 불공정성을 인지했을 때, 검토·이의제기·수정 요청과 같은 절차를 수행할 수 있도록 하는 교육이 요구된다. 아울러 AI 기반 평가 설계에서는 결과 제시에 그치지 않고, 학습자가 어떤 근거와 절차를 통해 결과에 개입할 수 있는지를 명확히 제시할 필요가 있다. 특히 윤리적 위험이 차원별로 다르게 작동한다는 점을 고려하여, 공정성 판단과 책임성 기반 절차 이해를 구분한 차별화된 교육 설계가 필요하다. 이러한 접근은 AI 리터러시를 실제 판단과 개입으로 연결하는 교육적 역량으로 확장하는 데 기여할 수 있다.

다만 본 연구는 AI 기반 평가 맥락에 한정되어 있어 일반화에 한계가 있으며, 경험 유무가 혼재된 표본을 사용하였다는 점에서 결과 해석에 제한이 있을 수 있다. 또한 본 연구는 합성 점수 기반 경로분석에 근거하고 있으므로, 확인된 매개 경로를 인과적 관계로 해석하는 데에는 신중할 필요가 있다. 후속 연구에서는 다양한 AI 활용 맥락에서 학습자의 실제 사용 행태와 행동 데이터를 결합하고, 경험 여부뿐 아니라 경험 수준을 보다 정교하게 구분하며, 균형 있는 측정 설계를 적용하여 AI 리터러시와 윤리적 위험, 책임있는 사용 의도 간의 관계를 보다 정밀하게 검증할 필요가 있다.

참고문헌

[1] 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* (CHI '20), 1-16. https://doi.org/10.1145/3313831.3376727

[2] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualising AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

[3] Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open, 6*, 100169. https://doi.org/10.1016/j.caeo.2024.100169

[4] Yim, I. H. Y. (2024). A critical review of teaching and learning artificial intelligence (AI) literacy: Developing an intelligence-based AI literacy framework for primary school education. *Computers and Education: Artificial Intelligence, 7*, 100319. https://doi.org/10.1016/j.caeai.2024.100319

[5] 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*, 100422. https://doi.org/10.1016/j.caeai.2025.100422

[6] Passlack, N., Hammerschmidt, M., Zimmer, M. P., & Durcikova, A. (2026). With great power comes great responsibility: What shapes AI literacy for responsible interactions of knowledge workers with AI, *Information Systems Frontiers, 28(1)*. https://doi.org/10.1007/s10796-025-10648-5

[7] Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. Routledge.

[8] Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology, 91*(1), 93-114. https://doi.org/10.1080/00223980.1975.9915803

[9] Kim, H., Chan, H. C., & Gupta, S. (2007). Value-based adoption of mobile internet: An empirical investigation. *Decision Support Systems, 43*(1), 111-126. https://doi.org/10.1016/j.dss.2005.05.009

[10] Bauer, R. A. (1960). Consumer behavior as risk taking. In R. S. Hancock (Ed.), *Dynamic marketing for a changing world: Proceedings of the 43rd national conference of the American Marketing Association* (pp. 389-398). American Marketing Association.

[11] Cox, D. F. (1967). Risk taking and information handling in consumer behavior. In D. F. Cox (Ed.), *Risk taking and information handling in consumer behavior* (pp. 604-624) Harvard University Press.

[12] Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human-Computer Studies, 59*(4), 451-474. https://doi.org/10.1016/S1071-5819(03)00111-3

[13] Choi, J. & Ji, Y. (2015). Investigating the importance of

trust on adopting an autonomous vehicle. I*nternational Journal of Human-Computer Interaction, 31*(10), 692-702. https://doi.org/10.1080/10447318.2015.1070549

[14] O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown Publishing Group.

[15] Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What do systems users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors. *MIS Quarterly, 39*(4), 837–864. https://doi.org/10.25300/MISQ/2015/39.4.5

[16] High-Level Expert Group on AI. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[17] Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review, 83*(2), 241–262. https://doi.org/10.1111/puar.13483

[18] 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

[19] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

[20] Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM, 59*(2), 56-62. https://doi.org/10.1145/2844110

[21] Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 599-609. https://doi.org/10.1145/3442188.3445921

[22] Schoenherr, F., & Al-Marzooq, A. (2024). Responsible social media use: How user characteristics shape the actualisation of ambiguous affordances. *European Journal of Information Systems, 34*(5), 799-821. https://doi.org/10.1080/0960085X.2024.2444249

[23] Perelman, L. (2014). When “the state of the art” is counting words. *Assessing Writing, 21*, 104-111. https://doi.org/10.1016/j.asw.2014.05.001

[24] UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

[25] Yang, S. J. H., Ogata, H., Matsui, T., & Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. Computers and Education: *Artificial Intelligence, 2*, 100008. https://doi.org/10.1016/j.caeai.2021.100008

[26] Zhu, W., Huang, L., Zhou, X., Li, X., Shi, G., & Ma, Y. (2025). Could AI ethical anxiety, perceived ethical risks and ethical awareness about AI influence university students' use of generative AI products? An ethical perspective. *International Journal of Human–Computer Interaction, 41*(6), 1555–1575. https://doi.org/10.1080/10447318.2024.2323277

[27] Lowry, P. B., & Moody, G. D. (2015). Proposing the control-reactance compliance model (CRCM) to explain opposing motivations to comply with organizational information security policies. *Information Systems Journal, 25*(5), 433–463.*https://doi.org/10.1111/isj.12043*

[28] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[29] Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI’18).* ACM. https://doi.org/10.1145/3173574.3174014

[30] The jamovi project. (2024). *jamovi (Version 2.7.11)* [Computer software]. https://www.jamovi.org

[31] Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423. https://doi.org/10.1037/0033-2909.103.3.411

[32] West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Sage Publications.

편은진

· 2002년 연세대학교 교육학과(교육공학 석사)
· 2007년 연세대학교 교육학과(교육공학 박사)
· 2024~25년 초등수학 AI 디지털교과서 검정심의위원
· 2025년~현재 연세대학교 AI혁신연구원 연구교수

관심분야 : 교수설계, AI평가, AI융합교육, AI교과교육

pej221242@gmail.com