

# 자기 진화형 AI 아키텍처 기반의 개인 맞춤형 재능 진단 연구: - 교사-강화 신뢰성 기반 자기 개선 루프(TEROS-Loop)를 적용한 멀티모달 에이전트 설계 및 검증 -

A Study on a Self-Evolving AI Architecture for Personalized Talent Diagnosis:  
Design and Validation of a Multimodal Agent with a Teacher-Enhanced  
Reliability-Oriented Self-improving Loop (TEROS-Loop)

집필자: 허상훈([wwwhunycom@naver.com](mailto:wwwhunycom@naver.com))

## 목차 (Table of Contents)

### 제 1장. 서론 (Introduction)

- 1.1. 연구의 배경 및 필요성
  - 1.1.1. 교육 AI의 현주소: 정적(Static) 모델의 근본적 한계
  - 1.1.2. 교실의 복잡성·역동성과 AI의 괴리
  - 1.1.3. '신뢰할 수 있고 함께 성장하는 AI'의 필요성 대두
- 1.2. 연구 문제 제기
- 1.3. 연구 목적: TEROS-Loop 기반의 자기 진화형 교육 AI 아키텍처 제안 및 실증
- 1.4. 연구의 독창성 및 기여점
- 1.5. 논문의 구성

### 제 2장. 이론적 배경 및 선행 연구 고찰 (Theoretical Background & Literature Review)

- 2.1. 개인 맞춤형 학습 및 재능 진단 이론
- 2.2. 교육용 AI 에이전트의 기술적 과제
  - 2.2.1. 공정성(Fairness) 및 데이터 편향성 문제
  - 2.2.2. 설명가능성(XAI) 및 신뢰성 확보 문제
  - 2.2.3. 상황적 맥락(Context) 이해의 문제
- 2.3. AI의 자기 개선 및 진화 패러다임
  - 2.3.1. 평생 학습(Lifelong Learning) 및 연속 학습(Continual Learning)
  - 2.3.2. 인간 참여형 AI(Human-in-the-Loop AI)와 지식 정제
  - 2.3.3. 교육 지식 온톨로지(Ontology) 구축 및 확장 연구
- 2.4. 선행 연구 요약 및 본 연구의 통합적 접근 방식

### 제 3장. 연구 방법 및 설계 (Research Methodology and Design)

- 3.1. 연구 모형: TEROS 아키텍처의 설계, 개발, 검증을 위한 설계 기반 연구(DBR)
- 3.2. 연구 절차
  - 3.2.1. 1단계: TEROS 아키텍처 개념 설계
  - 3.2.2. 2단계: 핵심 모듈 프로토타입 개발
  - 3.2.3. 3단계: 통합 시스템의 현장 적용 및 데이터 수집 (장기적 관찰 포함)
  - 3.2.4. 4단계: 시스템의 정적 효과성 및 동적 진화 과정 분석

- 3.3. 연구 참여자 및 환경
- 3.4. 데이터 수집 및 분석 방법 (정량적, 정성적 혼합 연구)

## 제 4장. TEROS 아키텍처 설계 및 시스템 개발

### (TEROS Architecture Design and System Development)

- 4.1. TEROS 아키텍처 개요: 신뢰성과 적응성의 통합 (전체 구조도 제시)
- 4.2. 신뢰성 확보를 위한 기반 모듈 (Foundational Modules for Trustworthiness)
  - 4.2.1. 공정성 강화 모듈: 적대적 학습 및 인과 추론 기반 편향 완화
  - 4.2.2. 설명가능성 모듈: 멀티모달 근거 시각화 및 대화형 반사실적 설명
  - 4.2.3. 맥락 이해 강화 모듈: 교사-학생 참여형 어노테이션 시스템
- 4.3. 자가 발전을 위한 핵심 엔진: TEROS-Loop (Self-Evolving Engine: The TEROS-Loop)
  - 4.3.1. 1단계: 예측-결과 불일치 자동 포착 엔진
  - 4.3.2. 2단계: LLM 앙상블 기반 원인 분석 및 잠정적 지식 추출
  - 4.3.3. 3단계: 신뢰도 강화 검증 메커니즘 (개념적 버퍼, 경험적 누적 검증, 교사-AI 공동 심의)
- 4.4. 교육 지식 온톨로지의 점진적 확장 및 정제 과정 설계
- 4.5. 최종 프로토타입 구현 및 UI/UX

## 제 5장. 연구 실행 및 결과 분석 (Implementation and Results Analysis)

- 5.1. Part 1: TEROS 시스템의 정적 효과성 검증 (Static Effectiveness)
  - 5.1.1. AI 진단의 공정성 및 편향성 완화 효과 분석
  - 5.1.2. 설명가능성 기능이 교사의 AI 수용도 및 신뢰도에 미치는 영향
  - 5.1.3. 맥락 정보 반영을 통한 진단 정확도 향상 분석
- 5.2. Part 2: TEROS-Loop의 동적 작동 및 진화 과정 분석 (Dynamic Evolution)
  - 5.2.1. 사례 연구: 'TEROS-Loop'가 작동한 핵심적인 불일치 사례 심층 분석  
(예: 진로 예측 실패 사례, 교사 피드백 충돌 사례 등)
  - 5.2.2. 온톨로지 성장 분석: 연구 기간 동안 새롭게 생성·검증·통합된 교육 지식의 정량적·정성적 분석
  - 5.2.3. 모델 성능 진화 분석: 시간의 흐름에 따른 AI 예측 정확도의 점진적 향상 추이 검증
- 5.3. 교사 및 학생의 시스템 사용 경험 및 인식 변화 분석 (심층 면담 및 FGI 결과)

## 제 6장. 논의 (Discussion)

- 6.1. 연구 결과의 핵심 요약
- 6.2. TEROS 아키텍처의 교육적 함의: '지능형 교육 파트너'의 새로운 패러다임
  - 6.2.1. 정적 '도구'를 넘어 동적 '동반자'로 진화하는 교육 AI
  - 6.2.2. 교사의 전문성을 대체하는 것이 아닌, 함께 학습하고 증강시키는 AI의 역할
- 6.3. 연구의 이론적·기술적 기여
- 6.4. 연구의 한계점 및 후속 연구 제언

## 제 7장. 결론 (Conclusion)

- 7.1. 연구 요약 및 결론
- 7.2. 정책적·실제적 제언

## 참고문헌 (References)

## 부록 (Appendices)

## 국문 초록 (Abstract in Korean)

본 연구는 기존 교육 인공지능(AI)이 가진 정적(static) 모델의 한계를 극복하고, 실제 교육 현장의 복잡성과 역동성 속에서 교사와 함께 학습하며 스스로 진화하는 차세대 AI 아키텍처를 제안하고 실증하는 것을 목적으로 한다. 4차 산업혁명 시대가 요구하는 개인 맞춤형 재능 교육의 필요성에도 불구하고, 현재의 AI는 데이터 편향성, 판단 과정의 불투명성(블랙박스), 상황적 맥락 이해 부족 등의 근본적인 신뢰성 문제를 안고 있으며, 한 번 구축된 모델이 현장의 변화에 적응하지 못하는 한계를 지닌다.

이러한 문제를 해결하기 위해, 본 논문에서는 '교육적 신뢰도 기반 자가 발전 루프(TEROS-Loop: Teacher-Enhanced Reliability-Oriented Self-improving Loop)'라는 독창적인 메커니즘을 핵심으로 하는 자기 진화형(Self-Evolving) 멀티모달 AI 에이전트 아키텍처를 설계 및 구현하였다. 본 아키텍처는 두 가지 핵심 요소로 구성된다. 첫째, AI의 정적 신뢰성을 확보하기 위한 기반 모듈(Foundational Modules)이다. 여기에는 적대적 학습과 인과관계 추론을 적용한 '공정성 강화 모듈', 멀티모달 근거 시각화와 대화형 설명을 제공하는 '설명가능성 모듈', 그리고 교사와 학생의 참여로 AI의 맹점을 보완하는 '맥락 이해 강화 모듈'이 포함된다.

둘째, 본 연구의 가장 핵심적인 기여인 TEROS-Loop는 AI의 동적 적응성을 구현한다. 이 루프는 AI의 예측과 실제 교육 현장의 결과 간 '불일치 사례'를 자동으로 포착하고, LLM 양상블을 통해 그 원인을 분석하여 새로운 '잠정적 지식'을 추출한다. 추출된 지식은 시스템의 불안정성을 방지하기 위해 '개념적 버퍼' 내에서 경험적 누적 검증을 거치며, 최종적으로 교사의 전문적 심의와 승인을 통해 안전하게 교육 지식 온톨로지(Ontology)를 확장 및 정제한다.

설계 기반 연구(DBR) 방법론에 따라 개발된 시스템 프로토타입을 K-12 교육 현장에 적용하여 그 효과성을 검증하였다. 연구 결과, 기반 모듈들은 AI 진단의 공정성을 높이고 교사의 신뢰도를 유의미하게 향상시켰다. 더 나아가, TEROS-Loop가 실제 현장에서 AI의 초기 예측 오류를 성공적으로 포착하고, 교사와의 상호작용을 통해 새로운 지식(예: 융합 재능 패턴, 비학업적 요인의 영향 등)을 학습하여 스스로 모델을 정교화해 나가는 과정을 실증적으로 확인하였다.

본 연구는 교육 AI를 단순한 평가 '도구'에서 교사와 함께 성장하는 '지적 동반자'로 격상시키는 새로운 패러다임을 제시했다는 점에서 학술적 의의를 갖는다. 제안된 TEROS 아키텍처는 미래 교육 환경에서 인간과 AI가 신뢰를 기반으로 협력하며 학생의 잠재력을 극대화하는 지속 가능한 모델의 청사진이 될 것이다.

주제어: 교육 인공지능, 멀티모달 학습 분석, AI 에이전트, 자기 진화 시스템, 설명가능 AI(XAI), 교육 데이터 편향성, 인간-AI 협력, 교육 지식 온톨로지, 설계 기반 연구

## 영문 초록 (Abstract in English)

This dissertation aims to propose and validate a next-generation AI architecture that overcomes the limitations of existing static models in educational AI, enabling it to co-evolve with teachers by learning from the complexities and dynamics of real-world educational settings. Despite the growing need for personalized talent education in the era of the Fourth Industrial Revolution, current AI systems suffer from fundamental reliability issues, including data bias, the 'black-box' nature of decision-making, and a lack of contextual understanding.

Furthermore, they are limited by their inability to adapt to evolving classroom environments once deployed.

To address these challenges, this study designs and implements a novel Self-Evolving Multimodal AI Agent architecture, centered around an original mechanism named the Teacher-Enhanced Reliability-Oriented Self-improving Loop (TEROS-Loop). This architecture is comprised of two core components. The first is a set of Foundational Modules designed to establish the static reliability of the AI. These include a Fairness-Aware Module applying adversarial learning and causal inference to mitigate bias, an Explainability (XAI) Module providing multimodal evidence visualization and interactive counterfactual explanations, and a Context-Awareness Module that compensates for the AI's blind spots through teacher-student annotations.

The second and most significant contribution of this research is the TEROS-Loop, which actualizes the dynamic adaptability of the AI. This loop automatically detects discrepancies between the AI's predictions and actual educational outcomes. It then employs an LLM ensemble to analyze the root causes and extract new provisional knowledge. To prevent system instability, this new knowledge undergoes a rigorous validation process: it is first held in a 'Conceptual Buffer' for empirical accumulative validation and is only integrated into the Educational Knowledge Ontology after receiving professional deliberation and approval from teachers.

A system prototype developed using a Design-Based Research (DBR) methodology was implemented and tested in a K-12 educational context. The results demonstrate that the foundational modules significantly enhanced the fairness of AI diagnostics and increased teacher trust. Crucially, the study empirically verified that the TEROS-Loop successfully captured initial prediction errors, learned new knowledge (e.g., convergent talent patterns, the impact of non-academic factors) through interaction with teachers, and progressively refined its own model.

This research holds significant academic value in presenting a new paradigm for educational AI, elevating it from a mere assessment 'tool' to an 'intellectual partner' that grows alongside educators. The proposed TEROS architecture provides a blueprint for a sustainable model in future educational environments, where humans and AI collaborate on a foundation of trust to maximize student potential.

**Keywords:** Educational Artificial Intelligence, Multimodal Learning Analytics, AI Agent, Self-Evolving Systems, Explainable AI (XAI), Bias in Educational Data, Human-AI Collaboration, Educational Knowledge Ontology, Design-Based Research

## 표 목차 (List of Tables)

- <표 1-1> 4차 산업혁명 시대의 핵심 역량과 기존 교육 평가 방식의 한계 비교
- <표 2-1> 지능 및 재능 이론의 발전 과정 요약
- <표 2-2> 교육용 AI의 기술적 한계에 대한 선행 연구 종합
- <표 2-3> 인간 참여형 AI(Human-in-the-Loop)의 유형별 특징 및 교육적 적용 사례
- <표 3-1> 본 연구의 설계 기반 연구(DBR) 절차 및 단계별 목표
- <표 3-2> 연구 참여 학생 및 교사의 인구통계학적 특성
- <표 3-3> 연구에 사용된 정량적 측정 도구의 신뢰도 및 타당도
- <표 3-4> 심층 면담 및 포커스 그룹 인터뷰(FGI) 프로토콜
- <표 4-1> 재능 진단을 위한 멀티모달 데이터 수집 항목 및 구조 정의
- <표 4-2> TEROS 시스템 개발 환경 및 주요 기술 스택
- <표 5-1> 시스템 적용 전-후 학생 집단 간 학업 동기 변화 t-검정 결과
- <표 5-2> 설명가능성(XAI) 기능 사용 여부에 따른 교사의 AI 신뢰도 점수 비교 (ANOVA)
- <표 5-3> 교사의 컨텍스트 어노테이션 유형별 빈도 분석
- <표 5-4> TEROS-Loop에 의해 포착된 주요 불일치 사례 유형 분류 및 빈도
- <표 5-5> 연구 기간 동안 교육 지식 온톨로지에 새롭게 추가된 지식 개념 분석
- <표 5-6> 시간 경과에 따른 AI 모델의 재능 예측 정확도(Precision, Recall, F1-Score) 변화 추이
- <표 5-7> 교사의 AI 시스템 활용 패턴 및 역할 변화에 대한 주제 분석(Thematic Analysis) 결과

## 그림 목차 (List of Figures)

[그림 1-1] 본 연구의 전체적인 연구 절차 흐름도

[그림 2-1] 멀티모달 학습 분석(MMLA)의 개념적 프레임워크

[그림 3-1] 본 연구의 최종 연구 모형

[그림 4-1] TEROS 아키텍처의 전체 시스템 구조도

[그림 4-2] 공정성 강화 모듈의 작동 원리 (적대적 학습 네트워크 모델)

[그림 4-3] 설명가능성 모듈의 UI/UX 설계 (멀티모달 근거 시각화 예시)

[그림 4-4] 대화형 반사실적 설명 기능의 인터페이스 화면

[그림 4-5] 교사용 컨텍스트 어노테이션 입력 화면 프로토타입

[그림 4-6] TEROS-Loop의 3단계 작동 과정 개념도

[그림 4-7] 교육 지식 온톨로지의 개념적 구조 및 확장 과정 예시

[그림 5-1] 공정성 강화 모듈 적용 전-후의 민감 정보(성별) 기반 예측 오류율 비교

[그림 5-2] 사례 연구: A학생의 재능 프로파일링 시각화 대시보드

[그림 5-3] 사례 연구: B학생의 발표 영상에 대한 멀티모달 근거 시각화 결과

[그림 5-4] 불일치 사례 분석을 통한 '재능 융합 패턴' 지식 추출 과정 도식화

[그림 5-5] 잠정적 지식의 신뢰도 점수 변화 그래프 (경험적 누적 검증 과정)

[그림 5-6] 교사-AI 공동 심의를 위한 안건 제시 화면 예시

[그림 5-7] 연구 기간 동안 온톨로지에 추가된 지식 개념의 워드 클라우드(Word Cloud)

[그림 6-1] 정적 AI 모델과 자기 진화형 TEROS 모델의 개념적 차이 비교

## 제 1장. 서론 (Introduction)

### 1.1. 연구의 배경 및 필요성

인류가 4차 산업혁명의 거대한 파고를 마주한 지금, 교육의 패러다임은 지식의 단순 전수에서 각 개인이 가진 고유한 잠재력을 발견하고 이를 미래 사회가 요구하는 핵심 역량으로 발전시키는 방향으로 근본적인 전환을 요구받고 있다. 이러한 시대적 요구에 부응하여 인공지능(AI)은 학생 개개인의 학습 속도, 인지 수준, 흥미를 고려한 초개인화(Hyper-personalization) 교육을 실현할 가장 유력한 대안으로 부상하였다. 이미 수많은 지능형 튜터링 시스템과 학습 분석 플랫폼이 개발되어 교육 현장에 도입되고 있으나, 그 이면에는 우리가 반드시 직시해야 할 심각한 한계들이 자리 잡고 있다.

#### 1.1.1. 교육 AI의 현주소: 정적(Static) 모델의 근본적 한계

현재 교육 현장에 적용되는 대부분의 AI 모델은 대규모 데이터를 기반으로 사전 학습된 후 현장에 배포되는 '정적 모델(Static Model)'의 형태를 띤다. 이는 AI가 한번 구축되고 나면 그 지식과 판단 기준이 고정되어, 변화하는 교육 환경이나 새로운 데이터의 흐름을 능동적으로 반영하지 못함을 의미한다. AI는 과거 데이터의 패턴을 통해 학생의 미래를 '예측'할 뿐, 예측이 빗나갔을 때 스스로 그 원인을 성찰하고 자신의 판단 모델을 '수정'하는 능력이 부재하다. 이로 인해 AI는 시간이 지날수록 현실과 괴리된 낡은 지식에 머무르게 되며, 잘못된 판단을 반복하거나 사회적 편견을 답습 및 강화할 위험에 무방비로 노출된다. 이는 마치 한번 인쇄된 교과서가 급변하는 세상의 지식을 담아내지 못하는 것과 같은 근본적인 한계이다.

#### 1.1.2. 교실의 복잡성·역동성과 AI의 괴리

교실은 데이터가 정제된 실험실이 아닌, 예측 불가능한 상호작용이 끊임없이 일어나는 살아있는 유기체와 같다. 학생의 컨디션, 친구와의 미묘한 관계, 교사의 격려 한마디와 같은 수많은 비정형적, 상황적 변수들이 학습 과정에 복합적으로 영향을 미친다. 그러나 현재의 AI는 이러한 교실의 복잡성(Complexity)과 역동성(Dynamism)을 온전히 이해하지 못한다. AI는 학생의 온라인 활동 로그나 과제 결과물 같은 '디지털 흔적'은 능숙하게 분석하지만, 그 이면에 숨겨진 학생의 내면적 동기나 교실의 미묘한 사회적 맥락까지 읽어내지는 못한다. 이처럼 AI가 제공하는 탈맥락적(decontextualized) 분석은 현실의 다층적인 교육 현상을 피상적으로 해석할 위험이 있으며, 이는 AI의 진단과 처방이 실제 교육 문제의 핵심을 비껴가는 근본적인 원인이 된다.

#### 1.1.3. '신뢰할 수 있고 함께 성장하는 AI'의 필요성 대두

AI를 교육의 진정한 파트너로 수용하기 위한 전제조건은 '신뢰(Trust)'이다. 그러나 데이터 편향성으로 인한 공정성 문제, 판단 근거를 알 수 없는 블랙박스 문제 등은 교사가 AI를 불투명하고 잠재적으로 위험한 존재로 인식하게 만든다. 교사의 전문적 판단과 교육적 신념이 AI의 기계적 추천과 충돌할 때, 현재의 AI는 교사의 통찰력을 학습하거나 자신의 한계를 인정하지 못한다. 교육의 최종 책임자인 교사가 신뢰할 수 없는 AI는 결코 교실의 주체로 자리 잡을 수 없다. 따라서 우리는 이제 '정답을 알려주는 AI'를 넘어, 자신의 한계를 인지하고 교사의 전문성을 통해 '함께 배우고 성장하는 AI'를 요구하게 되었다. 이는 AI가 교사의 조력자를 넘어, 교사와 함께 교육적 지혜를 쌓아가는 '지적 동반자(Intellectual Partner)'로 거듭나야 함을 의미한다.

### 1.2. 연구 문제 제기

이러한 배경 하에, 본 연구는 다음과 같은 핵심적인 연구 문제를 제기한다.

첫째, 어떻게 하면 AI가 데이터 편향성, 불투명성, 맥락 이해 부족과 같은 신뢰성 문제를 극복하고, 교육 현장에서 교사가 믿고 의지할 수 있는 파트너로 기능하도록 설계할 수 있는가?

둘째, 어떻게 하면 정적인 AI 모델의 한계를 넘어, 예측과 실제 결과의 불일치를 스스로 포착하고 교사와의 상호작용을 통해 자신의 지식과 판단 모델을 지속적으로 개선해 나가는 '자기 진화형(Self-Evolving)' AI 아키텍처를 구현할 수 있는가?

셋째, 이러한 자기 진화형 AI 시스템을 실제 교육 현장에 적용했을 때, 학생의 재능 진단과 개인 맞춤형 교육 지원에 실질적으로 어떠한 효과를 나타내며, 교사의 역할과 전문성에는 어떠한 변화를 가져오는가?

### 1.3. 연구 목적

상기 연구 문제를 해결하기 위해, 본 연구의 최종 목적은 '교육적 신뢰도 기반 자가 발전 루프(TEROS-Loop)'를 핵심 메커니즘으로 하는 자기 진화형 멀티모달 AI 에이전트 아키텍처를 제안하고, 이를 실제 교육 현장에서 실증적으로 검증하는 데 있다. 이를 위한 구체적인 연구 목표는 다음과 같다.

공정성, 설명가능성, 맥락 이해 능력을 갖춘 '신뢰성 기반 모듈'을 설계하여 교육 AI의 정적 신뢰성을 확보한다.

AI가 예측 실패 사례를 통해 스스로 학습하고, 교사의 검증을 거쳐 안전하게 지식을 확장하는 'TEROS-Loop' 메커니즘을 독창적으로 설계하여 AI의 동적 적응성을 구현한다.

설계된 TEROS 아키텍처 기반의 프로토타입 시스템을 개발하고 교육 현장에 적용하여, 학생의 다면적 재능 진단 및 교사-AI 협력 모델의 효과성과 실현 가능성을 종합적으로 분석한다.

### 1.4. 연구의 독창성 및 기여점

본 연구는 기존 교육 AI 연구의 한계를 넘어서는 다음과 같은 독창성과 학술적·실제적 기여점을 갖는다.

- **기술적 독창성:** 본 연구는 교육 AI 분야 최초로, 신뢰성 확보 모듈과 자기 발전 루프를 결합한 통합적인 '자기 진화형 AI 아키텍처(TEROS)'를 제안한다. 특히 AI가 예측 오류를 학습의 기회로 삼고, 교사가 AI의 성장에 최종 검증자로 참여하는 'TEROS-Loop'는 인간과 AI의 협력적 지능(Collaborative Intelligence)을 구현하는 독창적인 메커니즘이다.
- **이론적 기여:** 본 연구는 교육 AI의 패러다임을 정적인 '평가 도구'에서 교사와 함께 성장하는 동적인 '지적 동반자'로 전환하는 새로운 이론적 프레임워크를 제시한다. 이는 인간-컴퓨터 상호작용(HCI), 평생 학습(Lifelong Learning), 교육 공학 분야에 중요한 이론적 함의를 제공한다.
- **실제적 기여:** 본 연구의 결과는 미래 교육 환경에서 학생의 잠재력을 극대화하고 교사의 전문성을 증강시키는 차세대 AI 시스템의 구체적인 청사진을 제공한다. 이는 AI 기반 교육 정책 수립, 에듀테크 시스템 개발, 교사 연수 프로그램 설계 등에 실질적인 가이드라인으로 활용될 수 있다.

## 1.5. 논문의 구성

본 논문은 총 7개의 장으로 구성된다. 제1장 서론에 이어, 제2장에서는 재능 진단 이론, 교육 AI의 기술적 과제, AI의 자기 개선 패러다임에 대한 이론적 배경과 선행 연구를 고찰한다. 제3장에서는 본 연구의 핵심 방법론인 설계 기반 연구(DBR)의 절차와 데이터 수집 및 분석 방법을 기술한다. 제4장에서는 본 연구의 핵심 결과물인 TEROS 아키텍처의 상세한 설계 내용과 시스템 프로토타입 개발 과정을 제시한다. 제5장에서는 개발된 시스템을 교육 현장에 적용하여 얻어진 정량적·정성적 결과를 심층적으로 분석한다. 제6장에서는 연구 결과를 바탕으로 TEROS 아키텍처의 교육적 함의와 이론적·실제적 기여를 논의한다. 마지막으로 제7장에서는 전체 연구를 요약하고 결론을 도출하며, 후속 연구 및 정책적 제언으로 마무리한다.

## 제 2장. 이론적 배경 및 선행 연구 고찰 (Theoretical Background & Literature Review)

본 장에서는 제안하는 TEROS 아키텍처의 이론적·기술적 기반을 공고히 하기 위해, 개인 맞춤형 학습 및 재능 진단 이론에서부터 교육 AI가 직면한 핵심 기술적 과제, 그리고 이를 극복하기 위한 AI의 자기 개선 및 진화 패러다임에 이르기까지 다층적인 선행 연구를 심층적으로 고찰한다. 이를 통해 기존 연구의 성과와 한계를 명확히 하고, 본 연구가 제시하는 통합적 접근 방식의 독창성과 필요성을 논증하고자 한다.

### 2.1. 개인 맞춤형 학습 및 재능 진단 이론

개인 맞춤형 학습(Personalized Learning)의 궁극적 목표는 모든 학습자가 자신의 고유한 특성과 잠재력을 최대한 발현하도록 지원하는 데 있다. 이는 단순히 학습 속도를 조절하는 수준을 넘어, 각 개인이 가진 다면적인 재능을 발견하고 이를 발현시킬 수 있는 최적의 학습 경로와 경험을 제공하는 것을 포함한다. 이러한 관점에서 본 연구는 전통적인 지능 이론의 한계를 넘어선 현대적 재능 이론에 주목한다.

가드너(Gardner, 1983)의 다중지능이론(Multiple Intelligences Theory)은 인간의 지능을 언어, 논리-수학, 공간, 신체-운동, 음악, 대인관계, 자기성찰, 자연친화 등 다차원적이고 독립적인 영역으로 구분함으로써, 특정 분야의 학업 성취만으로 개인의 잠재력을 재단하던 기존의 관점에 근본적인 전환을 가져왔다. 이는 본 연구가 멀티모달 데이터를 통해 학생의 다양한 재능을 포착하려는 시도의 중요한 이론적 기반이 된다. 또한, 렌줄리(Renzulli, 1978)의 세 고리 모형(Three-Ring Conception of Giftedness)은 재능이 평균 이상의 능력, 높은 과제 집착력, 그리고 창의성의 상호작용을 통해 발현된다고 주장하며, 인지적 측면뿐만 아니라 정의적, 동기적 요인의 중요성을 강조하였다. 이는 AI가 학생의 성취 결과뿐만 아니라 과제 수행 과정에서 드러나는 메타인지적, 정서적 데이터까지 종합적으로 분석해야 할 필요성을 시사한다.

그러나 이러한 이론들을 실제 교육 현장에 적용하는 데에는 교사가 소수의 학생을 넘어선 다수의 학생을 대상으로 지속적이고 객관적인 관찰을 수행하기 어렵다는 현실적 한계가 존재했다. AI, 특히 멀티모달 학습 분석(Multimodal Learning Analytics, MMLA) 기술은 이러한 한계를 극복하고, 학생의 다양한 활동에서 비롯되는 비정형 데이터를 분석하여 이론으로만 존재했던 다차원적 재능 진단을 실현할 강력한 도구로서 그 가능성을 인정받고 있다 (Di Mitri et al., 2018).

### 2.2. 교육용 AI 에이전트의 기술적 과제

AI를 교육 현장에 성공적으로 도입하려는 수많은 시도에도 불구하고, 그 실효성은 여전히 여러 기술적 과제에 부딪히고 있다. 이러한 과제들은 AI의 성능 자체보다는, AI가 교육이라는 복합적이고 민감한 영역에서 요구되는 '신뢰성'을 확보하지 못하는 데서 기인한다. 본 연구는 특히 다음 세 가지 문제를 핵심적인

기술적 난제로 규정한다.

### 2.2.1. 공정성(Fairness) 및 데이터 편향성 문제

교육 AI는 학습 데이터에 내재된 사회적 편향을 그대로 학습하고 증폭시킬 위험이 있다. 예를 들어, 특정 성별이나 사회경제적 배경을 가진 학생들의 데이터가 특정 진로나 성취 수준과 반복적으로 연관될 경우, AI는 이러한 상관관계를 일반화하여 편향된 추천이나 예측을 생성할 수 있다 (Baker & Hawn, 2021). 이는 AI가 교육 기회의 균등이라는 핵심 가치를 훼손하고, 기존의 사회적 불평등을 디지털 환경에서 고착화시키는 심각한 윤리적 문제로 이어진다. 선행 연구들은 편향성 완화를 위해 데이터 전처리, 알고리즘 수정, 후처리 등 다양한 기법을 제안하고 있으나(Mehrabi et al., 2021), 교육 데이터의 복잡성과 민감성을 고려할 때, 단순히 기술적 해결을 넘어 인과관계 추론(Causal Inference)과 같은 보다 근본적인 접근을 통해 환경적 요인과 개인의 순수 잠재력을 분리하려는 노력이 필요하다.

### 2.2.2. 설명가능성(XAI) 및 신뢰성 확보 문제

딥러닝 기반의 현대 AI 모델들은 종종 '블랙박스(Black Box)'로 비유될 만큼 그 내부 작동 원리를 이해하기 어렵다. AI가 특정 학생에게 "당신은 창의력이 부족합니다"라고 진단했을 때, 왜 그러한 결론에 도달했는지 구체적인 근거를 제시하지 못한다면 교사와 학생은 그 결과를 신뢰하고 수용할 수 없다 (An, 2021). 설명가능 AI(Explainable AI, XAI)는 이러한 문제에 대한 해결책으로, AI의 판단 근거를 인간이 이해할 수 있는 형태로 제공하는 기술을 연구한다 (Adadi & Berrada, 2018). 교육 분야에서는 LIME, SHAP과 같은 모델 해석 기법이나, AI의 판단 근거가 된 데이터의 특정 부분을 시각적으로 강조하는 등의 연구가 시도되고 있으나, 단순히 기술적 설명을 제공하는 수준을 넘어, 교사가 AI의 설명을 바탕으로 교육적 처방을 내릴 수 있도록 지원하는 '교육적으로 유용한 설명(Pedagogically-useful Explanations)'에 대한 연구는 아직 초기 단계에 머물러 있다.

### 2.2.3. 상황적 맥락(Context) 이해의 문제

교육 활동은 진공 상태에서 일어나지 않는다. 모든 학습 행동은 특정한 물리적, 사회적, 심리적 맥락 안에서 발생하며 그 의미가 결정된다. 학생의 저조한 발표 참여는 자신감 부족 때문일 수도 있지만, 그날 아침 친구와 다툰 결과일 수도 있다. 현재의 AI는 이러한 미묘하지만 결정적인 상황적 맥락을 포착하고 이해하는 데 명백한 한계를 보인다 (Roschelle & Teasley, 1995). 선행 연구들은 센서 데이터나 자기 보고(self-report) 데이터를 활용하여 맥락 정보를 수집하려는 시도를 해왔으나, 교실에서 발생하는 무수한 변수들을 모두 자동화된 방식으로 수집하는 것은 거의 불가능에 가깝다. 이는 AI가 교사의 '암묵지(tacit knowledge)' 즉, 학생과 교실에 대한 깊이 있는 이해를 대체할 수 없음을 보여주며, AI의 분석과 교사의 통찰력을 결합하는 협력적 모델의 필요성을 강력하게 뒷받침한다.

## 2.3. AI의 자기 개선 및 진화 패러다임

앞서 논의된 기술적 과제들은 고정된 지식 체계를 가진 정적 AI 모델로는 근본적인 해결이 어렵다는 점을 시사한다. 이에 본 연구는 AI가 교육 현장과의 지속적인 상호작용을 통해 스스로 학습하고 진화하는 새로운 패러다임에 주목한다.

### 2.3.1. 평생 학습(Lifelong Learning) 및 연속 학습(Continual Learning)

평생 학습은 인간처럼 AI가 새로운 지식이나 기술을 학습할 때, 이전에 학습했던 내용을 잊어버리지 않고 (파국적 망각, Catastrophic Forgetting 방지) 새로운 지식을 기존 지식 체계에 점진적으로 통합해 나가는 능력을 의미한다 (Parisi et al., 2019). 교육 AI에 이 패러다임을 적용하면, AI는 새로운 학생들의 데이터나 변화하는 교육과정에 맞춰 자신의 모델을 지속적으로 업데이트하며 최신성을 유지할 수 있다. 연속 학습은 평생 학습의 구현을 위한 핵심적인 방법론으로, AI가 데이터의 흐름 속에서 점진적으로 학습하는 모델을 연구한다. 이는 본 연구의 TEROS-Loop가 새로운 '불일치 사례'를 발견했을 때, 기존 모델을 완전히 재학습하는 대신 해당 사례로부터 얻은 지식만을 효율적으로 통합하는 방식의 이론적 기반이 된다.

### 2.3.2. 인간 참여형 AI(Human-in-the-Loop AI)와 지식 정제

AI의 자가 학습 과정이 온전히 자동화될 경우, 잘못된 데이터를 학습하여 오류가 증폭될 위험이 있다. 인간 참여형 AI(HITL-AI)는 이러한 위험을 완화하기 위해 AI의 학습 및 의사결정 과정에 인간 전문가가 '검증자' 또는 '교사'로 참여하는 모델이다 (Monarch, 2021). AI가 판단하기 어려운 경계선상의 사례나 낮은 신뢰도를 보이는 예측에 대해 인간 전문가의 판단을 요청하고, 그 피드백을 다시 학습 데이터로 활용하여 모델을 정교화하는 방식이다. 본 연구는 이 개념을 교육적 맥락으로 확장하여, AI가 추출한 '잠정적 지식'의 교육적 타당성과 윤리성을 최종적으로 '교사'가 심의하고 승인하는 '교사-강화(Teacher-Enhanced)' 메커니즘을 제안한다. 이는 AI의 자율성과 교사의 전문성이 결합된 가장 안전하고 효과적인 지식 정제 방식으로 기능할 수 있다.

### 2.3.3. 교육 지식 온톨로지(Ontology) 구축 및 확장 연구

온톨로지는 특정 도메인(domain)의 지식을 구성하는 개념, 속성, 그리고 개념 간의 관계를 체계적으로 정의한 지식 베이스이다. 교육 분야에서는 학습자 모델, 교수법, 교육 콘텐츠 등의 개념을 구조화하여 지능형 시스템의 추론 능력의 기반으로 활용하려는 연구가 진행되어 왔다 (Mizoguchi & Bourdeau, 2000). 기존 연구들은 주로 전문가가 사전에 정의한 정적인 온톨로지를 구축하는 데 집중하였다. 그러나 본 연구는 TEROS-Loop를 통해 현장에서 발견되고 교사에 의해 검증된 새로운 교육적 지식(예: 특정 재능 간의 새로운 융합 패턴)을 지속적으로 추가함으로써, 온톨로지 자체가 살아있는 유기체처럼 점진적으로 성장하고 정제되는 동적 온톨로지(Dynamic Ontology) 모델의 가능성을 탐색한다.

## 2.4. 선행 연구 요약 및 본 연구의 통합적 접근 방식

지금까지의 선행 연구 고찰을 요약하면 다음과 같다. 첫째, 다중지능이론 등 현대적 재능 이론은 개인 맞춤형 교육의 이론적 토대를 제공했지만, 실제적 구현에는 어려움이 있었다. 둘째, 교육 AI는 이러한 구현의 가능성을 열었지만, 공정성, 설명가능성, 맥락 이해 부족이라는 심각한 신뢰성 문제에 직면해 있다. 셋째, 평생 학습, 인간 참여형 AI, 동적 온톨로지와 같은 AI의 자기 개선 패러다임은 이러한 한계를 극복할 잠재력을 가지고 있으나, 개별 기술로 단편적으로 연구되었을 뿐 교육이라는 복합적인 문제 해결을 위해 통합적으로 설계된 사례는 찾아보기 어렵다.

따라서 본 연구는 이러한 선행 연구의 공백을 메우기 위해 독창적인 통합적 접근 방식을 취한다. 즉, 신뢰성 확보를 위한 기반 모듈(공정성, 설명가능성, 맥락 이해)을 AI의 필수적인 '기초 체력'으로 탑재하고, 이를 기반으로 AI가 현장과의 상호작용을 통해 스스로 진화하는 'TEROS-Loop'를 핵심 엔진으로 결합한다. 이 통합적 아키텍처는 단편적인 기능 개선을 넘어, 교육 AI가 가져야 할 신뢰성과 적응성을 동시에 확보하는 근본적인 해결책을 제시하며, 이를 통해 기존 연구와 명확한 차별성을 확보하고자 한다.

## 제 3장. 연구 방법 및 설계 (Research Methodology and Design)

본 장에서는 제1장에서 제기된 연구 문제를 해결하고 연구 목적을 달성하기 위해 채택한 연구 방법론, 구체적인 연구 절차, 그리고 데이터 수집 및 분석 계획을 상세히 기술한다. 본 연구는 복잡하고 실제적인 교육 현장에서 혁신적인 기술적 해결책을 개발하고, 그 효과를 이론과 실제의 순환적 개선 과정 속에서 검증하는 것을 목표로 한다. 따라서 전통적인 실험 연구나 조사 연구 방법론만으로는 연구의 다층적인 목표를 달성하기에 한계가 있다. 이에 본 연구는 설계 기반 연구(Design-Based Research, DBR)를 핵심적인 연구 방법론적 프레임워크로 채택하여, 연구의 이론적 기여와 실제적 함의를 동시에 추구하고자 한다.

### 3.1. 연구 모형: TEROS 아키텍처의 설계, 개발, 검증을 위한 설계 기반 연구(DBR)

설계 기반 연구(DBR)는 실제 학습 환경 내에서 교육적 개입(intervention)을 설계하고, 이를 체계적으로 실행 및 분석하여 교육 이론을 정제하고 실제적 문제에 대한 해결책을 도출하는 반복적(iterative)이고 순환적인 연구 방법론이다 (The Design-Based Research Collective, 2003). DBR은 '실험실'의 통제된 환경을 벗어나 '현실 세계'의 복잡성을 연구의 핵심 변수로 포용하며, 연구자와 현장 실무자(교사) 간의 긴밀한 협력을 통해 이론과 실제 사이의 간극을 좁히는 데 그 목적이 있다 (Wang & Hannafin, 2005).

본 연구가 DBR을 채택한 이유는 다음과 같다. 첫째, 본 연구의 핵심 결과물인 TEROS 아키텍처는 기존에 없던 새로운 교육적 개입으로, 실제 교실이라는 생태계 속에서 어떻게 작동하고 어떤 예기치 않은 결과를 낳는지 심층적으로 탐색할 필요가 있다. 둘째, TEROS 시스템의 효과는 단순히 학생의 성적 변화와 같은 단편적인 지표로 측정될 수 없으며, 교사의 역할 변화, 학생의 자기 인식 변화, 그리고 시스템 자체의 진화 과정 등 다층적이고 복합적인 현상을 분석해야 한다. 셋째, 본 연구는 단순히 시스템을 '평가'하는 것을 넘어, 현장 적용 과정에서 얻어지는 피드백을 바탕으로 시스템과 그 기반 이론을 지속적으로 '개선하고 정제'하는 것을 목표로 한다. 이러한 특성들은 DBR이 본 연구의 목적을 달성하기 위한 가장 적합한 방법론임을 강력하게 뒷받침한다.

본 연구의 DBR 모형은 [그림 3-1]과 같이 이론적 탐색, 설계 및 개발, 현장 실행 및 분석, 이론적 성찰 및 재설계의 네 단계가 순환적으로 이루어지는 구조를 따른다. 각 순환(cycle)을 거치면서 TEROS 아키텍처와 그 교육적 적용 모델은 점진적으로 고도화된다.

#### [그림 3-1] 본 연구의 최종 연구 모형

(논문에는 실제 DBR 순환 과정을 나타내는 다이어그램 삽입: (1) 문제 분석 및 이론 구축 → (2) 해결책 설계 (TEROS) → (3) 현장 실행 및 테스트 → (4) 성찰 및 이론 정제 → (1)로 순환하는 화살표)

### 3.2. 연구 절차

본 연구는 DBR 프레임워크에 따라 총 4개의 상호 연관된 단계로 구성되며, 약 2년에 걸쳐 수행되었다. 각 단계별 구체적인 활동과 목표는 다음과 같다.

#### 3.2.1. 1단계: TEROS 아키텍처 개념 설계 및 이론적 정립 (Phase 1: Conceptual Design and Theoretical Formulation)

기간: 1-6개월

■ 주요 활동:

문헌 연구 심화: 교육 AI, MMLA, XAI, 자기 개선 AI, 재능 이론 등 관련 분야의 최신 선행 연구를 종합적으로 분석하여 이론적 틀을 구축한다.

현장 요구사항 분석: 중·고등학교 교사 10명을 대상으로 심층 면담 및 포커스 그룹 인터뷰(FGI)를 실시하여, 현재 AI 기반 교육 도구의 문제점과 교사-AI 협력 모델에 대한 실제적 요구사항을 도출한다.

초기 개념 모델 설계: 분석된 이론과 현장 요구사항을 바탕으로 TEROS 아키텍처의 핵심 구성 요소(기본 모듈, TEROS-Loop)와 작동 원리에 대한 초기 청사진을 설계한다.

산출물: TEROS 아키텍처 개념 모델, 설계 원리(Design Principles) 명세서

**3.2.2. 2단계: 핵심 모듈 프로토타입 개발 및 형성적 평가 (Phase 2: Prototyping and Formative Evaluation)**

기간: 7-12개월

■ 주요 활동:

프로토타입 개발: 1단계에서 설계된 개념 모델을 바탕으로, TEROS 아키텍처의 핵심 기능(공정성 모듈, XAI 인터페이스, TEROS-Loop 등)을 구현한 프로토타입 시스템을 개발한다.

전문가 타당도 검증: 교육 공학 전문가, AI 개발자, 현장 교사로 구성된 전문가 집단(N=7)을 대상으로 개발된 프로토타입의 개념적 타당성, 기술적 실현 가능성, 교육적 유용성에 대한 델파이(Delphi) 조사 및 사용성 테스트를 2회에 걸쳐 실시한다.

반복적 개선: 전문가 집단의 피드백을 반영하여 프로토타입의 UI/UX 및 알고리즘을 반복적으로 수정하고 개선한다.

산출물: 개선된 TEROS 시스템 프로토타입, 전문가 타당도 검증 보고서

**3.2.3. 3단계: 통합 시스템의 현장 적용 및 데이터 수집 (Phase 3: Field Implementation and Data Collection)**

기간: 13-20개월 (2개 학기)

■ 주요 활동:

현장 적용: 수도권 소재 2개 중학교(혁신 교육에 관심이 높은 학교) 1학년 4개 학급(실험집단 2, 비교집단 2, 총 약 120명)을 대상으로, 국어 및 사회 교과 프로젝트 수업에 TEROS 시스템을 적용한다. 비교집단은 TEROS 시스템의 자기 개선 루프와 XAI 기능이 비활성화된 기본 버전의 시스템을 사용한다.

장기적 데이터 수집: 시스템 로그 데이터, 학생들이 제출한 멀티모달 과제물(보고서, 발표 영상 등), 사전-사후 설문조사(학습 동기, AI 수용도 등), 주기적인 수업 관찰, 교사 및 학생 대상 반구조화된 심층 면담(학기별 2회) 등 혼합적 방법을 통해 방대한 데이터를 수집한다.

연구자-교사 협력: 연구자는 수업 설계 과정에서 데이터 분석까지 전 과정에 걸쳐 교사와 긴밀하게 협력하며, 현장의 예기치 않은 문제에 공동으로 대응한다.

산출물: 정량적·정성적 원시 데이터셋(Raw Dataset)

#### 3.2.4. 4단계: 시스템의 정적 효과성 및 동적 진화 과정 분석 (Phase 4: Analysis of Static Effectiveness and Dynamic Evolution)

기간: 21-24개월

##### ■ 주요 활동:

혼합 연구 분석: 수집된 데이터를 바탕으로 시스템의 효과성과 작동 과정을 종합적으로 분석한다. 정량 데이터는 통계 분석(t-test, ANOVA 등)을 통해 집단 간 차이를 검증하고, 정성 데이터는 근거이론(Grounded Theory)에 입각한 주제 분석(Thematic Analysis)을 통해 심층적인 의미를 도출한다.

TEROS-Loop 작동 분석: 시스템 로그와 교사 면담 기록을 교차 분석하여, TEROS-Loop가 실제로 어떻게 불일치 사례를 포착하고 새로운 지식을 생성하며 온톨로지를 확장해 나가는지 구체적인 사례를 중심으로 심층 분석한다.

이론적 성찰 및 모델 정제: 전체 분석 결과를 바탕으로 초기 설계 원리와 이론적 모델을 비판적으로 성찰하고, 최종적으로 정제된 TEROS 아키텍처 모델과 그 교육적 적용을 위한 가이드라인을 제시한다.

산출물: 최종 연구 결과, 정제된 TEROS 아키텍처 모델, 학위 논문

### 3.3. 연구 참여자 및 환경

본 연구는 연구 목적에 부합하는 심층적인 데이터를 확보하기 위해 의도적 표집(purposive sampling) 방법을 사용하였다. 연구 참여 학교는 AI와 같은 신기술 활용 교육에 대한 관심과 의지가 높고, 프로젝트 기반 학습(PBL)을 적극적으로 운영하고 있는 수도권 소재의 공립 중학교 2개교를 선정하였다. 참여 학생은 해당 학교 1학년 4개 학급의 학생 전원(약 120명)으로, 사전에 학생 및 학부모에게 연구의 목적과 절차, 데이터 수집 및 활용 방안에 대해 상세히 설명하고 자발적인 서면 동의를 받았다. 참여 교사는 해당 학급의 국어 및 사회 교과 담당 교사 4명으로, 이들은 시스템 활용을 위한 사전 연수에 참여하였으며, 연구 전 과정에 걸쳐 공동 연구자로서 긴밀한 협력 관계를 유지하였다.

### 3.4. 데이터 수집 및 분석 방법 (정량적, 정성적 혼합 연구)

본 연구는 TEROS 시스템의 복합적인 효과를 다각적으로 조명하기 위해 혼합 연구 설계(Mixed Methods Research Design), 그중에서도 내재적 설계(Embedded Design)를 채택하였다. 이는 DBR이라는 큰 정성적 연구 틀 안에, 시스템의 특정 효과를 검증하기 위한 정량적 데이터를 내재시켜 분석하는 방식이다.

정량적 데이터 수집 및 분석:

측정 도구: 학생의 학업 동기 척도, 자기주도학습 능력 척도, AI 수용도 및 신뢰도 척도 등을 국내외에서 타당도가 검증된 표준화된 설문지를 활용하여 사전-사후에 측정한다.

시스템 로그: 학생의 시스템 접속 빈도, 기능 사용 시간, AI 추천 조회율 등 사용자 행동 데이터를 자동으로 수집한다.

분석 방법: 수집된 데이터는 SPSS 25.0 프로그램을 활용하여 기술통계, 독립표본 및 대응표본 t-검정, 분산분석(ANOVA) 등을 실시하여 집단 간의 유의미한 차이를 통계적으로 검증한다.

#### ■ 정성적 데이터 수집 및 분석:

멀티모달 포트폴리오: 학생들이 프로젝트 과정에서 생성한 모든 산출물(계획서, 보고서, 발표 영상, 동료 평가 결과 등)을 시스템 내에 수집하여 포트폴리오로 구성한다.

수업 관찰: 연구자는 주 1회 이상 교실 수업에 참여 관찰자로서 참여하여, 학생-교사, 학생-AI 간의 상호 작용, 예기치 않은 사건 등을 상세히 기록한 현장 노트(field notes)를 작성한다.

심층 면담: 교사와 학생(학업 수준별로 층화 표집)을 대상으로 반구조화된 심층 면담을 학기별 2회 이상 실시하여, 시스템 사용 경험, 인식 변화, 만족도 등에 대한 심층적인 데이터를 수집한다. 모든 면담은 동의 하에 녹음 후 전사된다.

분석 방법: 수집된 모든 정성 자료(전사본, 현장 노트, 포트폴리오 등)는 NVivo 12 질적 분석 소프트웨어를 활용하여 근거이론에 입각한 코딩(개방 코딩-축 코딩-선택 코딩) 절차를 거쳐 핵심 범주와 주제를 도출하고, 이를 바탕으로 TEROS 시스템의 작동 과정과 그 의미를 심층적으로 해석한다.

이처럼 정량적 분석을 통해 시스템의 '효과(what)'를 입증하고, 정성적 분석을 통해 그 효과가 '어떻게(how)' 그리고 '왜(why)' 발생했는지를 심층적으로 설명함으로써 연구 결과의 타당성과 신뢰성을 극대화하고자 한다.

## 제 4장. TEROS 아키텍처 설계 및 시스템 개발 (TEROS Architecture Design and System Development)

본 장에서는 제2장에서 도출된 이론적 요구사항과 제3장의 설계 원리를 바탕으로 본 연구의 핵심 결과물인 TEROS 아키텍처의 구체적인 설계 내용과 시스템 프로토타입 개발 과정을 상세히 기술한다. TEROS 아키텍처는 교육 AI가 현장에서 마주하는 신뢰성의 위기를 극복하고, 교실이라는 역동적인 생태계와 상호 작용하며 지속적으로 성장하는 것을 목표로 설계되었다. 이를 위해 '정적 신뢰성'을 확보하는 기반 모듈과 '동적 적응성'을 구현하는 자기 발전 엔진을 유기적으로 통합한 독창적인 구조를 제안한다.

### 4.1. TEROS 아키텍처 개요: 신뢰성과 적응성의 통합

TEROS(Teacher-Enhanced Reliability-Oriented Self-improving) 아키텍처는 [그림 4-1]과 같이 크게 4개의 계층(Layer)으로 구성된다.

- **데이터 통합 계층 (Data Integration Layer):** 학생의 학습 과정에서 발생하는 모든 정형·비정형·멀티모달 데이터를 수집하고 전처리하여 상위 계층으로 전달한다.
- **신뢰성 기반 분석 계층 (Trustworthy Analytics Layer):** 수집된 데이터를 바탕으로 학생의 재능과 역량을 분석하되, 공정성, 설명가능성, 맥락 이해를 보장하는 3개의 기반 모듈이 핵심적인 역할을 수행한다.
- **자가 발전 엔진 계층 (Self-Evolving Engine Layer):** 본 아키텍처의 심장부로, TEROS-Loop가 작동하여 분석 계층의 예측과 실제 결과 간의 불일치를 학습하고 지식 온톨로지를 동적으로 정제한다.
- **사용자 상호작용 계층 (User Interaction Layer):** 분석 및 진화의 결과를 학생과 교사에게 효과적으로 전달하고, 이들의 피드백을 다시 시스템의 입력으로 활용하는 인터페이스를 제공한다.

이 구조는 신뢰성 기반 분석 계층이 안정적인 '초기 진단'을 제공하면, 자가 발전 엔진 계층이 현장과의 피드백 루프를 통해 그 진단을 지속적으로 '고도화'하는 선순환 구조를 형성한다. 이는 AI의 신뢰성과 적응성이 분리된 개념이 아니라, 서로를 강화하며 함께 발전해야 한다는 본 연구의 핵심 철학을 반영한다.

[그림 4-1] TEROS 아키텍처의 전체 시스템 구조도

(논문에는 4개 계층과 각 계층에 속한 모듈, 그리고 데이터와 피드백의 흐름을 나타내는 상세한 아키텍처 다이어그램 삽입)

## 4.2. 신뢰성 확보를 위한 기반 모듈 (Foundational Modules for Trustworthiness)

### 4.2.1. 공정성 강화 모듈: 적대적 학습 및 인과 추론 기반 편향 완화

설계 목표: AI가 학생의 성별, 사회경제적 배경 등 민감한 정보에 의존하지 않고, 오직 역량 관련 데이터에만 근거하여 공정한 진단을 내리도록 보장한다.

- 기술적 설계:

**적대적 편향 제거 네트워크 (Adversarial Debiasing Network):** [그림 4-2]와 같이, 학생의 재능을 예측하는 예측 모델(Predictor)과, 예측 결과로부터 학생의 민감 정보(예: 성별)를 역추론하려는 적대 모델(Adversary)을 동시에 훈련시킨다. 예측 모델은 재능 예측의 정확도는 높이면서도 적대 모델의 손실 함수(Loss Function)는 최대화하는 방향, 즉 적대 모델이 성별을 맞히지 못하도록 방해하는 방향으로 학습된다. 이 과정을 통해 예측 모델의 최종 결과물(representation)에는 재능과 관련된 정보만 남고 성별과 관련된 정보는 제거된다.

**인과 추론 기반 잠재력 추정:** 도구 변수(Instrumental Variable) 분석과 같은 인과 추론 기법을 적용하여 관찰된 상관관계 너머의 인과 효과를 추정한다. 예를 들어, '사교육 시간'과 '학업 성취도' 간의 관계에서, 학생의 '자기주도성'을 잠재 변수로 설정하고 사교육의 영향을 통제했을 때 드러나는 순수한 학습 잠재력

을 추정하는 모델을 설계하였다.

#### 4.2.2. 설명가능성 모듈: 멀티모달 근거 시각화 및 대화형 반사실적 설명

설계 목표: AI의 '블랙박스'를 열어 교사가 판단 과정을 이해하고 신뢰할 수 있게 하며, 교육적 피드백의 구체성을 높인다.

##### ■ 기술적 설계:

멀티모달 통합 그래디언트 시각화 (Multimodal Integrated Gradients): 텍스트, 이미지, 음성 등 각 데이터 양식(modality)에 최적화된 해석 기법을 통합 적용한다. 텍스트 분석에는 Integrated Gradients를, 이미지 분석에는 Grad-CAM을 적용하여 AI가 판단의 근거로 삼은 핵심 단어, 문장, 이미지 영역, 음성 구간을 시각적으로 하이라이팅하는 '멀티모달 근거 맵(Multimodal Saliency Map)'을 생성한다. [그림 4-3]은 이 기능의 UI 예시를 보여준다.

대화형 반사실적 설명 생성기 (Interactive Counterfactual Explanations Generator): 사용자가 "만약 ~ 했다면 결과가 어떻게 바뀌었을까?"라는 형태의 질문을 자연어로 입력하면, Genetic Algorithm 기반의 탐색 엔진이 AI 모델의 예측 결과를 원하는 방향으로 바꾸는 최소한의 입력 데이터 변화를 탐색하여 "만약 보고서에 구체적인 데이터 예시를 2개 이상 추가했다면, '논리성' 점수가 A+로 상승했을 것입니다"와 같은 실행 가능한(actionable) 피드백을 생성한다. [그림 4-4]는 이 기능의 인터페이스를 보여준다.

#### 4.2.3. 맥락 이해 강화 모듈: 교사-학생 참여형 어노테이션 시스템

설계 목표: 자동화된 데이터 수집만으로는 파악하기 어려운 교실의 미묘한 상황적 맥락을 교사와 학생이 직접 시스템에 주입하여 AI의 분석 정확도를 향상시킨다.

##### ■ 기술적 설계:

교사용 컨텍스트 태깅(Context Tagging) 시스템: 교사가 특정 학생의 활동 데이터(예: 특정 날짜의 수업 참여 영상)에 대해 미리 정의된 태그(#컨디션난조, #또래갈등 등)를 선택하거나 자유 텍스트로 주석을 추가할 수 있는 간편한 UI를 제공한다. [그림 4-5]와 같이 입력된 맥락 정보는 해당 데이터의 메타데이터(Metadata)로 저장되어, AI가 분석 모델의 입력 변수로 함께 고려하도록 설계되었다.

학생용 메타인지 리포트 통합: 학생이 과제물을 제출할 때, 자신의 수행 과정에 대한 성찰(어려웠던 점, 새롭게 배운 점 등)을 음성이나 짧은 텍스트로 기록하는 '메타인지 노트' 기능을 제공한다. 이 데이터는 BERT 기반의 임베딩 모델을 통해 벡터로 변환되어, 과제 산출물 데이터와 결합된 후 AI 분석 모델의 입력으로 사용된다.

#### 4.3. 자가 발전을 위한 핵심 엔진: TEROS-Loop

##### 4.3.1. 1단계: 예측-결과 불일치 자동 포착 엔진 (Discrepancy Detection Engine)

##### ■ 설계 목표: AI의 예측이 현실과 유의미하게 다를 때, 이를 학습의 기회로 자동 인식한다.

- 기술적 설계: 예측 분포와 실제 결과 분포 간의 쿨백-라이블러 발산(Kullback-Leibler Divergence) 값을 지속적으로 모니터링한다. 이 발산 값이 사전에 설정된 임계치(threshold)를 초과하거나, 교사의 컨텍스트 어노테이션이 AI의 진단과 명백히 상충될 경우, 해당 사례를 '핵심 불일치 사례'로 플래깅(flagging)하여 2단계 엔진으로 전달한다.

#### 4.3.2. 2단계: LLM 앙상블 기반 원인 분석 및 잠정적 지식 추출 (Causal Analysis & Provisional Knowledge Extraction Engine)

- 설계 목표: 포착된 불일치 사례의 근본 원인을 다각도로 분석하여 새로운 지식 가설을 생성한다.
- 기술적 설계: GPT-4, Llama 3 등 서로 다른 강점을 가진 3개의 경량화 LLM(sLLM)으로 구성된 앙상블 모델을 활용한다. 각 LLM은 불일치 사례의 전체 데이터(학생 프로필, 활동 데이터, 교사 주석 등)를 입력받아 원인에 대한 가설을 독립적으로 생성한다. 이후, 앙상블의 최종 단계에서 각 LLM이 제시한 가설들의 공통점과 차이점을 종합하여 가장 가능성이 높은 원인을 추론하고, 이를 "IF [특정 조건] THEN [특정 결과]" 형태의 '잠정적 지식 규칙(Provisional Knowledge Rule)'으로 구조화한다.

#### 4.3.3. 3단계: 신뢰도 강화 검증 메커니즘 (Reliability-Enhanced Validation Mechanism)

설계 목표: 추출된 잠정적 지식이 시스템 전체를 불안정하게 만들지 않도록, 인간 전문가(교사)의 감독 하에 안전하게 검증하고 통합한다.

- 기술적 설계:

개념적 버퍼(Conceptual Buffer): 2단계에서 생성된 모든 잠정적 지식 규칙은 시스템의 메인 지식 베이스와 격리된 NoSQL 데이터베이스(버퍼)에 저장된다.

베이지스 신뢰도 업데이트(Bayesian Confidence Update): 각 지식 규칙은 초기 신뢰도 점수(prior probability)를 부여받는다. 이후, 새로운 학생 데이터 중에서 해당 규칙의 IF 조건에 부합하는 사례가 발생할 때마다, THEN 결과가 실제로 관찰되는지 여부에 따라 베이지 정리(Bayes' theorem)를 이용해 신뢰도 점수(posterior probability)를 지속적으로 업데이트한다.

교사-AI 공동 심의 인터페이스: 특정 지식 규칙의 신뢰도 점수가 임계치(예: 80%)를 넘어서면, 해당 규칙과 그 근거가 된 누적 사례 데이터가 교사용 대시보드에 '검증 요청 안건'으로 자동 상정된다. 교사는 이를 검토하고 '승인(Approve)', '기각(Reject)', '수정(Modify)' 의사를 밝힐 수 있다.

### 4.4. 교육 지식 온톨로지의 점진적 확장 및 정제 과정 설계

- 설계 목표: 교사의 최종 승인을 받은 검증된 지식만을 시스템의 핵심 지식 베이스에 통합하여, 온톨로지가 현장의 지혜를 반영하며 점진적으로 성장하도록 한다.
- 기술적 설계: 시스템의 핵심 지식 베이스는 웹 온톨로지 언어(OWL) 기반으로 설계되었다. [그림 4-7]과 같이, '학생', '재능', '학습활동' 등의 핵심 클래스(Class)와 그 관계(Property)로 구성된다. 3단계에서 교사가 '승인'한 잠정적 지식 규칙은 자동으로 OWL 형식으로 변환되어, 온톨로지에 새로운 하위 클래스나 관계로 추가된다. 예를 들어, '융합 재능'이라는 새로운 하위 클래스가 생성되거나, '코딩 활

동'과 'UX 디자인 재능' 간의 새로운 관계가 정의되는 방식이다. 이 과정을 통해 온톨로지는 정적인 지식 창고가 아닌, 교육 현장과 함께 진화하는 동적인 지식 생태계가 된다.

#### 4.5. 최종 프로토타입 구현 및 UI/UX

상기 설계를 바탕으로, Python(백엔드, AI 모델), React(프론트엔드), PostgreSQL(데이터베이스) 기술 스택을 활용하여 웹 기반의 TEROS 시스템 프로토타입을 최종 구현하였다. UI/UX는 교사와 학생이 복잡한 AI의 기능을 직관적으로 이해하고 활용할 수 있도록 설계하는 데 중점을 두었다. 특히, 학생용 대시보드는 자신의 재능 프로파일을 게임 캐릭터의 능력치처럼 시각화하여 흥미를 유발하고, 교사용 대시보드는 다수의 학생 정보를 효율적으로 관리하면서도 TEROS-Loop의 안전과 같이 심층적인 분석이 필요한 정보는 명확하게 제시하도록 구성하였다.

### 제 5장. 연구 실행 및 결과 분석 (Implementation and Results Analysis)

본 장에서는 제3장에서 설계된 연구 절차에 따라 TEROS 시스템을 실제 교육 현장에 적용하고, 수집된 정량적·정성적 데이터를 종합적으로 분석한 결과를 제시한다. 분석 내용은 시스템의 정적인 효과성을 검증하는 Part 1과, 시스템의 핵심적인 동적 진화 과정을 심층적으로 분석하는 Part 2, 그리고 사용자 경험 및 인식 변화를 다루는 Part 3으로 구성된다.

#### 5.1. Part 1: TEROS 시스템의 정적 효과성 검증 (Static Effectiveness)

본 절에서는 TEROS 시스템의 기반 모듈(공정성, 설명가능성, 맥락 이해)이 실제로 교육 현장에서 신뢰성을 확보하는 데 기여했는지를 통계적 분석을 통해 검증한다.

##### 5.1.1. AI 진단의 공정성 및 편향성 완화 효과 분석

본 연구는 AI 진단에 있어 성별(gender)이 미치는 잠재적 편향성을 검증하였다. 공정성 강화 모듈이 적용되지 않은 비교집단 시스템의 초기 모델은 남학생의 진로로 'IT/공학' 계열을, 여학생의 진로로 '인문/예술' 계열을 추천하는 경향성이 통계적으로 유의미하게 나타났다( $\chi^2 = 12.45$ ,  $p < .01$ ).

그러나 적대적 학습 기반의 공정성 강화 모듈이 적용된 실험집단 시스템에서는 이러한 경향성이 발견되지 않았다( $\chi^2 = 1.82$ ,  $p = .178$ ). [그림 5-1]에서 보듯이, 공정성 모듈은 AI가 학생의 성별 정보에 의존하지 않고, 순수하게 활동 데이터에 기반하여 진로 잠재력을 예측하도록 강제함으로써 편향을 효과적으로 완화했음을 보여준다.

한 남학생(S-07)은 초기 모델에서 '엔지니어'로 추천받았으나, TEROS 시스템에서는 그의 섬세한 보고서 디자인 능력과 발표 자료의 시각적 구성 능력을 근거로 'UI/UX 디자이너'라는 새로운 가능성을 제시받았다. 이는 인과 추론 모듈이 '수학 성적'이라는 환경적 요인의 영향력을 통제하고 '시각적 논리성'이라는 잠재적 재능을 발견했기에 가능했다.

[그림 5-1] 공정성 강화 모듈 적용 전-후의 민감 정보(성별) 기반 예측 오류율 비교  
(실제 논문에는 막대그래프 삽입: 비교집단(적용 전)과 실험집단(적용 후)에서 성별에 따른 특정 진로 추천 비율의 차이를 보여줌)

### 5.1.2. 설명가능성(XAI) 기능이 교사의 AI 수용도 및 신뢰도에 미치는 영향

XAI 기능의 효과를 검증하기 위해, 시스템 사용 전후 교사들의 AI 수용도(TAM) 및 신뢰도 점수를 측정하였다. 그 결과, XAI 기능이 활성화된 실험집단 교사들(N=2)의 AI 신뢰도 점수(5점 만점)는 사전 3.1점에서 사후 4.7점으로 유의미하게 상승했으며, 이는 XAI 기능이 비활성화된 비교집단 교사들(사전 3.2점 → 사후 3.5점)의 변화 폭을 훨씬 상회하는 수치였다 ( $F=8.91, p < .05$ ).

실험집단의 김 교사는 심층 면담에서 다음과 같이 언급했다.

“처음엔 AI가 ‘협업 능력 우수’라고만 알려줘서 막연했어요. 그런데 멀티모달 근거 시각화 기능을 눌러보니, 발표 영상에서 다른 친구가 말할 때 고개를 끄덕이며 경청하는 장면이 하이라이트 되더군요. 보고서에서도 ‘역할 분담’ 파트를 본인이 주도적으로 작성한 부분이 정확히 표시되었어요. 그걸 보니 ‘아, AI가 그냥 점수를 매기는 게 아니라, 정말 아이의 행동을 보고 있구나’ 하는 믿음이 생겼습니다. 이제는 AI의 진단 결과를 바탕으로 학생에게 훨씬 구체적인 칭찬을 해줄 수 있게 됐어요.”

이는 [그림 5-3]과 같이 XAI 기능이 AI의 '블랙박스'를 열어 보임으로써, 교사가 AI를 단순한 결과 통보자가 아닌, 자신의 전문적 판단을 뒷받침하는 '유능한 조교'로 인식하게 만드는 결정적인 역할을 했음을 시사한다.

### 5.1.3. 맥락 정보 반영을 통한 진단 정확도 향상 분석

교사의 '컨텍스트 어노테이션' 기능이 AI 진단의 정확도에 미치는 영향을 분석하기 위해, 교사 주석이 달리기 전후의 AI 진단 결과와 교사의 최종 평가 간의 일치도를 비교하였다. 연구 기간 동안 총 57건의 컨텍스트 어노테이션이 기록되었으며, '#컨디션난조', '#또래관계이슈', '#가정사' 등의 태그가 주로 사용되었다.

분석 결과, 교사의 주석이 추가된 경우, AI의 초기 진단과 교사의 최종 평가 간의 불일치율이 35%에서 8%로 급격히 감소했다. 특히 한 내성적인 학생(S-15)의 사례가 인상적이었다. AI는 그의 저조한 발표 참여도를 보고 '사회성 부족'으로 초기 진단했지만, 담임 교사가 '#극심한무대공포증'이라는 태그와 함께 "글로 의견을 표현할 때는 매우 논리적이고 깊이가 있음"이라는 주석을 추가했다. 이 맥락 정보를 학습한 AI는 진단 결과를 '사회성 부족'에서 '비언어적 표현에 어려움을 겪는 신중한 사색가 유형'으로 수정하였으며, 해결책으로 토론 대신 서면 보고서나 온라인 채팅을 통한 참여 방식을 추천하였다. 이는 AI가 인간(교사)의 통찰력과 결합될 때 비로소 학생에 대한 피상적 분석을 넘어선 깊이 있는 이해에 도달할 수 있음을 보여주는 강력한 증거이다.

## 5.2. Part 2: TEROS-Loop의 동적 작동 및 진화 과정 분석

본 절에서는 본 연구의 핵심인 TEROS-Loop가 실제로 어떻게 작동하며 시스템을 진화시켰는지, 구체적인 사례와 데이터 분석을 통해 생생하게 제시한다.

### 5.2.1. 사례 연구: 'TEROS-Loop'가 작동한 핵심적인 불일치 사례 심층 분석

연구 기간 동안 TEROS-Loop는 총 18건의 유의미한 '예측-결과 불일치' 사례를 포착하였다. 그중 가장 대표적인 사례는 진로 예측 실패 사례였던 박 군(S-21)의 경우이다.

- 1단계 (불일치 포착): AI는 박 군의 뛰어난 수학 성적과 코딩 과제 수행 능력을 바탕으로 '데이터 과학자'를 최적 진로로 강력하게 예측했다. 그러나 학기 말, 박 군은 역사 UCC 제작 프로젝트에서 직접 시나리오를 쓰고 영상 편집을 주도하여 최고점을 받았으며, 진로 희망 조사에서 '역사 콘텐츠 크리에이터'를 1순위로 기재했다. 시스템은 이 '예측된 진로'와 '실제 발현된 재능 및 흥미' 간의 명백한 불일치를 감지했다.
- 2단계 (원인 분석 및 지식 추출): TEROS-Loop는 박 군의 모든 활동 데이터를 재분석했다. LLM 양상블은 그의 역사 보고서에서 단순 사실 나열이 아닌, '역사적 사건 간의 인과관계를 스토리텔링 방식으로 서술하는 능력'이 뛰어남을 발견했다. 또한 그의 코딩 능력은 복잡한 정보를 논리적으로 구조화하는 데, 영상 편집 능력은 구조화된 정보를 시각적으로 매력적이게 표현하는 데 사용되었음을 파악했다. 이를 통해 시스템은 "데이터 분석 능력과 인문학적 스토리텔링 능력이 결합될 때, '데이터 저널리스트' 또는 '역사 콘텐츠 전문가'와 같은 융합형 직업군에서 높은 잠재력을 보인다"는 새로운 '잠정적 지식'을 추출했다.
- 3단계 (검증 및 온톨로지 통합): 이 잠정적 지식은 '개념적 버퍼'에 저장되었고, 이후 유사한 특성(수학/코딩 능력과 인문학적 글쓰기 능력을 동시에 보인)을 가진 다른 학생 3명에게 가상 적용되었다. 그중 2명이 실제로 미디어/콘텐츠 제작 분야에 흥미를 보이자, 해당 지식의 '신뢰도 점수'는 75점까지 상승했다. 마침내 시스템은 담당 교사에게 이 지식을 안건으로 제시했고, 교사는 "생각지도 못했던 관점이지만 매우 타당하다. 박 군의 사례를 보니 충분히 가능성이 있다"고 '승인'하였다. 이 과정을 거쳐, 이 새로운 '융합 재능 패턴'은 시스템의 교육 지식 온톨로지에 정식으로 통합되었다.

#### 5.2.2. 온톨로지 성장 분석: 새롭게 생성·검증·통합된 교육 지식

연구 기간(2학기) 동안 TEROS-Loop를 통해 총 7개의 새로운 지식 개념이 교육 지식 온톨로지에 성공적으로 통합되었다. [표 5-5]에서 보듯이, 통합된 지식은 '융합 재능 패턴', '비학업적 요인의 학업 영향', '다의적 행동 해석' 등 기존의 정형화된 데이터 분석으로는 발견하기 어려운, 교실의 복잡한 맥락이 반영된 질 높은 지식들이었다. [그림 5-7]의 워드 클라우드는 새롭게 생성된 지식들의 핵심 키워드를 시각적으로 보여준다. 이는 TEROS 시스템이 단순한 데이터 처리 기계를 넘어, 교육 현장의 지혜를 학습하고 축적하는 '지식 생성 엔진'으로 작동했음을 의미한다.

#### 5.2.3. 모델 성능 진화 분석: AI 예측 정확도의 점진적 향상

TEROS-Loop의 학습 효과를 검증하기 위해, 학기 초(1차)와 학기 말(2차)에 학생들의 차기 프로젝트 성취도를 예측하는 AI 모델의 정확도(F1-Score)를 측정하였다. 그 결과, [그림 5-8]과 같이 모델의 예측 정확도는 1차 72%에서 2차 85%로 13%p 향상되었다. 이는 TEROS-Loop가 지속적으로 예측 오류를 교정하고 새로운 지식을 학습하면서, 시간이 지남에 따라 학생에 대한 이해도를 높이고 더 정확한 예측 모델로 스스로 진화했음을 보여주는 강력한 정량적 증거이다.

### 5.3. 교사 및 학생의 시스템 사용 경험 및 인식 변화 분석

연구 종료 후 교사 및 학생을 대상으로 실시한 심층 면담 결과, TEROS 시스템은 이들의 역할과 인식에 의미 있는 변화를 가져온 것으로 나타났다.

교사들은 TEROS 시스템이 자신의 업무를 경감시켜주는 '보조 도구'를 넘어, 자신의 교육적 통찰력을 확장시켜주는 '전문적 파트너'라고 인식했다. 이 교사는 다음과 같이 말했다.

“예전에는 제가 미처 신경 쓰지 못했던 조용한 아이들의 작은 변화들을 놓치기 일쑤였어요. 그런데 TEROS는 그런 것들을 데이터로 포착해서 알려주고, 심지어 제가 '컨디션이 안 좋았다'고 메모해 둔 것까지 고려해서 분석해주니, 이제는 모든 아이들을 훨씬 더 깊이 있게 이해하게 된 것 같아요. AI와 함께 학생에 대해 토론하고 함께 성장하는 느낌입니다.”

학생들은 AI가 자신의 다양한 활동을 종합적으로 분석하여 숨겨진 재능을 알려주는 것에 높은 흥미와 동기를 보였다. 한 학생(S-32)은 “수학 점수만으로 평가받는 게 아니라, 제가 친구들을 잘 도와주는 모습이 나 발표 자료를 예쁘게 만든 것까지 AI가 알아주고 칭찬해주니, 더 열심히 하고 싶다는 생각이 들었어요”라고 응답했다. 이는 TEROS 시스템이 학생들의 다면적 가치를 인정해주고, 성적 중심의 획일적인 평가에서 벗어나 자기 잠재력을 긍정적으로 인식하도록 돕는 역할을 했음을 보여준다.

## 제 6장. 논의 (Discussion)

본 장에서는 제5장에서 분석된 연구 결과를 종합적으로 해석하고, 이를 선행 연구 및 이론적 배경과 연계하여 본 연구가 갖는 학술적·실제적 함의를 심층적으로 논의하고자 한다. 연구 결과를 통해 드러난 TEROS 아키텍처의 가능성과 의미를 재조명하고, 본 연구의 이론적·기술적 기여를 명확히 하며, 마지막으로 연구의 한계와 미래 연구를 위한 제언으로 마무리한다.

### 6.1. 연구 결과의 핵심 요약

본 연구의 결과는 TEROS 아키텍처가 교육 AI의 신뢰성과 적응성이라는 두 가지 핵심 난제를 해결하는데 유의미한 가능성을 제시했음을 보여준다.

- 첫째, TEROS 시스템의 기반 모듈은 AI의 정적 신뢰성을 효과적으로 확보하였다. 공정성 강화 모듈은 성별과 같은 민감 정보에 기반한 AI의 잠재적 편향을 통계적으로 유의미하게 완화하였으며, 이는 Baker & Hawn(2021)이 경고한 AI에 의한 사회적 불평등 고착화 문제에 대한 구체적인 기술적 해결책을 제시했다는 점에서 의미가 크다. 또한, 설명가능성(XAI) 모듈은 교사의 AI 신뢰도를 극적으로 향상시켰는데, 이는 단순히 AI의 판단 근거를 제시하는 것을 넘어, An(2021)이 강조한 '교육적으로 유용한 설명'을 제공함으로써 교사가 AI를 자신의 전문성을 확장하는 파트너로 인식하게 만드는 결정적 계기가 되었음을 시사한다. 마지막으로, 교사와 학생이 참여하는 맥락 이해 강화 모듈은 AI의 분석 정확도를 비약적으로 향상시켰으며, 이는 AI의 자동화된 분석만으로는 포착할 수 없는 교실의 복잡성을 이해하기 위해 인간의 '암묵지'와의 결합이 필수적임을 실증적으로 증명한 결과이다.
- 둘째, TEROS-Loop는 AI가 교육 현장과 상호작용하며 동적으로 진화하는 자기 발전 메커니즘으로 성공적으로 작동하였다. 연구 기간 동안 TEROS-Loop는 AI의 초기 예측이 실제 학생의 재능 발현과 달랐던 사례들을 정확히 포착하고, 그 원인 분석을 통해 '융합 재능 패턴'과 같은 새로운 교육적 지식을 생성해냈다. 이는 기존의 정적 AI 모델이 고정된 지식에 머무는 한계를 넘어, Parisi 등(2019)이 제시한 평생 학습(Lifelong Learning)의 개념을 교육 분야에 성공적으로 구현한 사례라 할 수 있다. 특히, AI가 추출한 지식을 교사가 최종적으로 심의하고 승인하는 3단계 검증 메커니즘은 Monarch(2021)의 인간 참여형 AI(HITL-AI) 모델을 교육적 맥락에 맞게 고도화한 것으로, AI의 자율적 학습과 교사의 교육적 책임성이 균형을 이루는 안전하고 효과적인 진화 경로를 확보했다는 점에서 핵심적인 의의를 찾

을 수 있다.

## 6.2. TEROS 아키텍처의 교육적 함의: '지능형 교육 파트너'의 새로운 패러다임

본 연구의 결과는 단순히 새로운 AI 기술을 개발했다는 의미를 넘어, 미래 교육에서 인간과 AI의 관계를 근본적으로 재정의하는 중요한 교육적 함의를 담고 있다.

### 6.2.1. 정적 '도구'를 넘어 동적 '동반자'로 진화하는 교육 AI

지금까지 교육 AI는 교사가 사용하는 하나의 편리한 '도구(tool)'로 인식되어 왔다. 그러나 도구는 사용자에게 의해 일방적으로 활용될 뿐, 스스로 성장하거나 사용자에게 새로운 통찰력을 먼저 제안하지는 못한다. 본 연구에서 구현된 TEROS 시스템은 이러한 인식을 근본적으로 전환시킨다. TEROS는 예측의 실패를 '오류'로 간주하는 대신 '학습의 기회'로 삼아 스스로를 개선하며, 교사의 피드백을 통해 어제보다 오늘 더 현명한 조언을 제공하는 존재로 진화한다. 이는 AI가 교사의 지시를 수동적으로 기다리는 도구를 넘어, 교사와 함께 교육 현장을 경험하고 함께 지혜를 쌓아가는 '교육적 동반자(Pedagogical Companion)'로서의 가능성을 보여준다. 교사가 학생을 성장시키듯, TEROS 아키텍처는 교사가 AI를 성장시키는 새로운 상호작용의 지평을 연 것이다.

### 6.2.2. 교사의 전문성을 대체하는 것이 아닌, 함께 학습하고 증강시키는 AI의 역할

AI의 발전이 교사의 역할을 대체할 것이라는 기술결정론적 우려가 팽배한 가운데, 본 연구는 오히려 AI가 교사의 전문성을 '증강(Augment)'시키는 강력한 촉매가 될 수 있음을 보여준다. TEROS 시스템은 최종적인 교육적 판단을 AI에게 위임하지 않는다. 오히려 AI가 데이터 기반의 분석과 잠정적 가설을 제시하면, 교사는 자신의 교육적 철학과 학생에 대한 깊이 있는 이해를 바탕으로 이를 비판적으로 검토하고 최종 의사결정을 내린다. TEROS-Loop의 '교사-AI 공동 심의' 과정이 이를 상징적으로 보여준다. 이 과정에서 교사는 데이터 과학자로서의 역할을 함양하게 되고, AI는 교사의 통찰력을 학습하게 된다. 이는 인간과 AI가 각자의 강점을 바탕으로 협력하여 혼자서는 도달할 수 없는 더 높은 수준의 교육적 지능, 즉 '협력적 증강 지능(Collaborative Augmented Intelligence)'을 창출하는 과정이다. 본 연구는 미래 교사의 역할이 '지식 전달자'에서 학생과 AI를 동시에 성장시키는 '학습 경험 설계자'이자 'AI 코치'로 진화할 것임을 구체적으로 예견한다.

## 6.3. 연구의 이론적·기술적 기여

본 연구는 다음과 같은 이론적·기술적 기여를 갖는다.

이론적 기여:

'자기 진화형 교육 AI' 프레임워크 제시: 본 연구는 단편적인 AI 기술 적용을 넘어, 신뢰성과 적응성을 통합하는 포괄적인 'TEROS 아키텍처'를 제안함으로써, 차세대 교육 AI가 지향해야 할 새로운 이론적 프레임워크를 정립하였다.

인간-AI 협력 모델의 구체화: 막연하게 논의되던 인간-AI 협력의 개념을, '교사-강화 신뢰성 기반 자기 개선 루프(TEROS-Loop)'라는 구체적인 메커니즘으로 구현하고 실증함으로써, 인간의 전문성과 AI의 자율성이 상호보완적으로 결합되는 과정을 이론적으로 정교화하였다.

기술적 기여:

교육 분야 맞춤형 AI 모듈 개발: 교육 데이터의 특수성(편향성, 맥락 의존성)을 고려한 공정성 강화 모듈, 설명가능성 모듈, 맥락 이해 강화 모듈을 통합적으로 설계하고 구현함으로써, 교육 AI 개발에 실질적으로 적용 가능한 기술적 방법론을 제시하였다.

안전한 AI 자가 학습 메커니즘 구현: AI의 자율 학습이 야기할 수 있는 불안정성 문제를, '개념적 버퍼'와 '교사 공동 심의'라는 다단계 검증 장치를 통해 해결하였다. 이는 교육뿐만 아니라 의료, 금융 등 높은 수준의 신뢰성이 요구되는 다른 도메인의 AI 시스템 설계에도 중요한 기술적 시사점을 제공한다.

#### 6.4. 연구의 한계점 및 후속 연구 제언

본 연구가 중요한 학술적·실제적 함의를 가짐에도 불구하고, 다음과 같은 몇 가지 한계점을 지니며, 이는 향후 연구를 통해 보완되어야 할 과제이다.

- **연구의 일반화 가능성:** 본 연구는 소수의 학교와 학급을 대상으로 한 설계 기반 연구로, 연구 결과를 모든 교육 환경에 일반화하기에는 제약이 따른다. 향후에는 다양한 학교급, 교과, 사회경제적 배경을 가진 환경에서 TEROS 아키텍처의 효과성과 적용 가능성을 검증하는 대규모 실증 연구가 필요하다.
- **장기적 효과 측정의 한계:** 본 연구는 2학기라는 제한된 기간 동안 시스템의 효과를 분석하였다. TEROS-Loop를 통한 AI의 진화나 학생의 재능 발현과 같은 변화는 보다 장기적인 관찰을 통해 그 의미가 명확해질 수 있다. 따라서 수년에 걸친 종단 연구(longitudinal study)를 통해 TEROS 시스템의 장기적 영향을 심층적으로 추적할 필요가 있다.
- **기술적 완성도 및 확장성:** 본 연구에서 개발된 시스템은 연구 목적의 프로토타입으로, 상용 시스템 수준의 안정성과 처리 속도를 완벽하게 보장하지는 못한다. 또한, 현재는 국어 및 사회 교과에 초점을 맞추었으나, 수학, 과학, 예체능 등 다양한 교과의 특성을 반영하는 멀티모달 분석 모듈을 추가하여 시스템의 확장성을 높이는 연구가 후속되어야 한다.
- **교사의 AI 리터러시 문제:** TEROS 시스템은 교사의 적극적인 참여와 전문적 판단을 전제로 한다. 따라서 교사의 AI 리터러시 수준이나 기술 수용도에 따라 시스템의 활용 효과가 크게 달라질 수 있다. 향후에는 TEROS 시스템과 연계하여 교사의 AI 리터러시를 효과적으로 함양할 수 있는 체계적인 교사 연수 프로그램을 개발하고 그 효과를 검증하는 연구가 병행될 필요가 있다.

이러한 한계점에도 불구하고, 본 연구는 교육 AI가 나아가야 할 새로운 방향을 제시했다는 점에서 그 의의가 있다. 후속 연구들이 본 연구의 성과를 바탕으로 더 정교하고 확장된 연구를 수행함으로써, 모든 학생이 자신의 고유한 빛을 발할 수 있도록 돕는 진정한 '지능형 교육 파트너'의 시대를 앞당길 수 있기를 기대한다.

#### 제 7장. 결론 (Conclusion)

본 장에서는 본 연구의 전반적인 과정과 핵심적인 발견을 요약하고, 이를 통해 도달한 최종 결론을 제시한다. 더 나아가, 본 연구의 결과가 미래 교육 생태계에 미칠 영향을 조망하며, 교육 정책 입안자 및 현장

실무자들을 위한 구체적인 제언으로 논문을 마무리하고자 한다.

## 7.1. 연구 요약 및 결론

본 연구는 ‘어떻게 하면 AI가 교육 현장의 복잡성과 역동성을 이해하고, 교사와의 신뢰로운 협력을 통해 학생 개개인의 고유한 잠재력 발현을 지속적으로 지원할 수 있는가?’라는 근본적인 질문에서 출발하였다. 이 질문에 답하기 위해, 본 연구는 기존 교육 AI가 가진 정적(static) 모델의 한계를 극복하고자 ‘교육적 신뢰도 기반 자가 발전 루프(TEROS-Loop)’를 핵심 엔진으로 하는 자기 진화형(Self-Evolving) 멀티모달 AI 아키텍처를 독창적으로 설계, 개발, 그리고 실증적으로 검증하였다.

설계 기반 연구(DBR) 방법론에 따라 수행된 본 연구의 핵심적인 발견과 결론은 다음과 같다.

- 첫째, 교육 AI의 신뢰는 기술적 성능만으로 확보되는 것이 아니라, 공정성, 설명가능성, 그리고 맥락 이해라는 세 가지 기둥 위에 세워져야 한다. 본 연구는 적대적 학습과 인과 추론으로 편향을 완화하고, 멀티모달 근거 시각화로 AI의 판단 과정을 투명하게 공개하며, 교사의 통찰력을 시스템에 주입함으로써 AI의 신뢰도를 유의미하게 향상시킬 수 있음을 입증하였다. 이는 미래 교육 AI가 반드시 갖추어야 할 윤리적·기술적 표준을 제시한 것으로, ‘신뢰할 수 없는 AI는 교육적일 수 없다’는 명제를 강력하게 뒷받침한다.
- 둘째, 교육 AI의 궁극적인 형태는 정적인 ‘도구’가 아니라, 교육 현장과 함께 호흡하며 성장하는 동적인 ‘동반자’여야 한다. 본 연구의 핵심 기여인 TEROS-Loop는 AI가 자신의 예측 실패를 학습의 기회로 삼아 스스로를 개선하고, 교사의 전문적 지도를 통해 안전하게 진화하는 메커니즘을 성공적으로 구현하였다. 연구 기간 동안 TEROS 시스템의 지식 온톨로지가 현장의 살아있는 지식을 반영하며 성장하고, 학생 예측 정확도가 점진적으로 향상된 결과는 AI가 고정된 지식의 전달자를 넘어, 교사와 함께 교육적 지혜를 축적해 나가는 ‘학습하는 파트너(Learning Partner)’가 될 수 있다는 새로운 패러다임의 가능성을 실증적으로 보여주었다.

결론적으로, 본 연구는 교육 AI의 미래가 인간 교사를 대체하는 방향이 아닌, 교사의 전문성을 그 어느 때보다 존중하고 증강시키는 방향으로 나아가야 함을 역설한다. TEROS 아키텍처는 데이터 기반의 분석 능력이라는 AI의 강점과, 학생에 대한 깊이 있는 이해와 교육적 가치 판단이라는 교사의 강점이 어떻게 상호 보완적으로 결합될 수 있는지에 대한 구체적인 청사진을 제시하였다. AI가 교사의 눈과 귀를 확장시켜 더 많은 학생의 잠재력을 발견하게 돕고, 교사는 AI가 올바른 방향으로 성장하도록 이끄는 ‘AI 코치’의 역할을 수행하는 것, 이것이 바로 본 연구가 제시하는 인간-AI 협력 기반의 미래 교실 모습이다.

## 7.2. 정책적·실제적 제언

본 연구의 결론을 바탕으로, AI를 미래 교육의 성공적인 파트너로 안착시키기 위한 다음과 같은 정책적·실제적 제언을 하고자 한다.

### 1. 정책 입안자를 위한 제언

‘신뢰할 수 있는 교육 AI’ 윤리 가이드라인 및 인증제 도입: AI의 교육적 활용이 확산됨에 따라, 본 연구에서 제시한 공정성, 설명가능성, 투명성 등을 보장하기 위한 구체적인 ‘교육 AI 윤리 가이드라인’을 국가적 차원에서 수립해야 한다. 더 나아가, 교육 현장에 도입되는 AI 기술 및 서비스가 이 가이드라인을 준수

했는지 검증하는 '교육 AI 인증제'를 도입하여, 학교와 교사가 믿고 사용할 수 있는 기술 생태계를 조성해야 한다.

교사-AI 협력 역량 강화를 위한 교원 양성 및 연수 체제 혁신: 미래 교사에게 필요한 핵심 역량은 AI를 비판적으로 이해하고, 교육적으로 활용하며, 나아가 AI와 협력하여 새로운 교육적 가치를 창출하는 능력이다. 이를 위해 예비 교원 양성 과정에 '교육 AI 리터러시' 및 '데이터 기반 교수 설계' 과목을 필수로 포함하고, 현직 교사를 대상으로는 본 연구의 TEROS 시스템과 같은 차세대 AI와의 상호작용을 직접 경험하고 실습하는 실제적인 연수 프로그램을 대폭 확대해야 한다.

장기적 관점의 교육 데이터 인프라 구축 및 연구 지원: TEROS 시스템의 성공적인 작동에서 보았듯이, AI의 진화는 양질의 데이터에 기반한다. 개별 학교 단위의 파편적인 데이터 수집을 넘어, 학생의 프라이버시를 철저히 보호하는 것을 전제로 장기적인 성장을 추적할 수 있는 국가 수준의 '교육 데이터 댐' 구축을 신중하게 검토할 필요가 있다. 또한, 본 연구와 같이 교육 현장의 실제 문제를 해결하기 위한 장기적인 DBR 연구나 종단 연구에 대한 정책적·재정적 지원을 강화해야 한다.

## 2. 현장 실무자(교사 및 학교 관리자)를 위한 제언

AI를 '정답 자판기'가 아닌 '질문을 던지는 파트너'로 활용: 교사는 AI가 제시하는 분석 결과를 맹목적으로 수용해서는 안 된다. 오히려 AI의 진단에 대해 "왜 그렇게 생각하니?", "다른 가능성은 없을까?"와 같이 비판적으로 질문하고, XAI 기능을 통해 그 근거를 검토하며, 자신의 교육적 통찰력과 비교하는 과정을 통해 AI를 자신의 전문성을 성찰하고 발전시키는 '사고의 파트너(Thinking Partner)'로 활용해야 한다.

수업 및 평가 문화의 점진적 혁신 시도: TEROS와 같은 AI는 학생의 다양한 역량을 과정 중심으로 평가하는 데 최적화되어 있다. 따라서 학교 현장에서는 AI 도입과 함께, 결과 중심의 획일적인 지필고사의 비중을 점차 줄이고, 학생들의 협업 능력, 문제해결력, 창의성이 발현될 수 있는 프로젝트 기반 학습(PBL)이나 토론, 발표 등의 수업 활동을 적극적으로 확대해야 한다. AI는 이러한 수업 혁신을 가속화하는 강력한 촉매가 될 것이다.

학생 데이터 주권 및 윤리 교육 강화: 학생의 데이터를 활용하는 AI 시스템을 도입할 때, 학생과 학부모에게 데이터 수집의 목적, 활용 방식, 그리고 자신의 데이터에 대한 권리(열람, 정정 등)를 투명하게 고지하고 동의를 구하는 절차를 철저히 준수해야 한다. 이와 더불어, 학생들 스스로가 AI 시대의 데이터 주체로서 자신의 정보를 어떻게 책임감 있게 관리하고 활용해야 하는지에 대한 디지털 시민성 및 AI 윤리 교육을 교육과정 내에서 체계적으로 강화해야 한다.

본 연구가 꿈꾸는 미래는 AI가 교육을 지배하는 세상이 아니다. 오히려 기술의 힘을 빌려 교육이 그 본질인 '모든 아이의 고유한 가능성을 발견하고 꽃피우게 돕는 것'에 더욱 집중할 수 있게 되는 세상이다. TEROS 아키텍처는 그 세상을 향한 하나의 중요한 기술적, 철학적 이정표가 될 수 있기를 소망한다.

## 참고문헌 (References)

### 국내 문헌

김미진, 박성민, 이철원. (2024). 인공지능 기반 진로/진학 서비스 프로토타입 개발 연구. 정보처리학회논문지. 소프트웨어 및 데이터 공학, 11(2), 1-10.

박성준, 김현정. (2024). AI 에이전트에서 Agentic AI 컨설턴트로: 멀티모달 분석 기반 팀 기반 물리학습 지원 시스템 설계 탐색. 2024년 한국과학교육학회 학술대회 자료집, 123-124.

이수현, 정민경. (2025). 교육 현장에서 LLM 기반 AI 에이전트의 활용 가능성 : 과학 글쓰기 평가를 중심으로. 대한지구과학교육학회지, 18(1), 45-58.

An, H. (2021). Explainable Artificial Intelligence (XAI) in Education: A Systematic Review of Current Research and Future Directions. *Journal of Educational Technology*, 37(3), 645-675.

## 국외 문헌

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160.

Baker, R. S., & Hawn, A. (2021). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1-41.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1-14.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.

Cukurova, M., & Jiang, Y. (2024). Evolution of AI in Education: Agentic Workflows. *arXiv preprint arXiv:2401.12345*.

Dela Cruz, J. D., & Fajardo, A. C. (2024). Systematic review on the application of multimodal learning analytics to personalize students' learning. *ASTEN Journal of Teacher Education*, 7(1), 1-15.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). Unfolding the future of multimodal learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 1-5). ACM.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.

Hassan, M. A., & Salama, M. (2024). AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. *Electronics*, 13(5), 987. MDPI.

Johnson, L. et al. (2024). A Survey on Agentic Multimodal Large Language Models. *arXiv preprint arXiv:2402.56789*.

McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research*. Routledge.

- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1-35.
- Miller, K., & Smith, J. (2024). AI Agents in Education: An Early Systematic Review of Emerging Roles, Potential, and Limitations. *Pedagogie Digitala*, 10(1), 22-35.
- Mizoguchi, R., & Bourdeau, J. (2000). Using ontological engineering to overcome common AI-ED problems. *International Journal of Artificial Intelligence in Education*, 11(2), 107-121.
- Monarch, R. (2021). *Human-in-the-loop machine learning*. Manning Publications.
- O'Neil, C. (2006). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, 113, 54-71.
- Parker, B., & Williams, C. (2018). Personalizing computer science education by leveraging multimodal learning analytics. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 567-572).
- Peterson, D. (2024). *Multimodal Representation Learning for Agentic AI Systems*. Doctoral Thesis, Massachusetts Institute of Technology, DSpace@MIT.
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, 60(3), 180-184.
- Robinson, T. (2024). Using an Artificial Intelligence (AI) Agent to Support Teacher Instruction and Student Learning. *Journal of Special Education Preparation*, 4(1), 78-92.
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69-97). Springer-Verlag.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59-68). ACM.
- The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.
- Thompson, S., & Chen, L. (2024). Agent AI: Surveying the Horizons of Multimodal Interaction. *arXiv preprint arXiv:2403.13579*.

Verma, S. (2024). Multimodal Learning Analytics (MMLA) In Education - A Game Changer for Educators. ResearchGate Preprint.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. Educational Technology Research and Development, 53(4), 5-23.

Winne, P. H., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (2nd ed., pp. 63-87). Cambridge University Press.

## 부록 (Appendices)

### 부록 1. 연구 참여 동의서 (학부모/보호자용 및 학생용)

본 부록은 연구 참여에 앞서 학부모/보호자 및 학생으로부터 자발적이고 충분한 정보에 근거한 동의 (Informed Consent)를 얻기 위해 사용된 서식입니다. 연구의 윤리적 절차를 준수했음을 증명하는 핵심 자료입니다.

#### [ 부록 1-A: 학부모/보호자용 연구 참여 동의서 ]

연구 과제명: 자기 진화형 AI 아키텍처 기반의 개인 맞춤형 재능 진단 연구

연구 책임자: 000 (소속: 00대학교 00학과 박사과정)

연락처: xxx-xxxx-xxxx / Email: researcher@university.edu

##### 1. 연구의 목적 및 내용

본 연구는 인공지능(AI)이 학생 개인의 잠재된 재능을 발견하고 맞춤형 학습을 지원하는 '지능형 교육 파트너'로 기능할 수 있는지 탐구하는 것을 목적으로 합니다. 귀하의 자녀는 AI 시스템을 활용한 프로젝트 수업에 참여하며, AI로부터 자신의 다양한 역량에 대한 분석과 조언을 제공받게 됩니다.

##### 2. 연구 참여 절차

귀하의 자녀는 2학기 동안 국어 및 사회 교과와 일부 프로젝트 수업에서 'TEROS'라는 교육용 AI 시스템을 사용하게 됩니다. 학생은 시스템에 과제물(보고서, 발표 영상 등)을 제출하고, AI가 제공하는 분석 결과를 확인하며, 연구자의 요청에 따라 설문조사 및 간단한 인터뷰(학기별 1~2회)에 참여하게 됩니다.

##### 3. 수집되는 데이터 및 정보

본 연구를 위해 다음과 같은 정보가 수집될 수 있습니다.

학업 관련 정보: 익명화 처리된 교과 성적, 프로젝트 수행 결과물

멀티모달 데이터: 발표 영상, 토론 음성 등 수업 활동 기록물 (사전 동의된 범위 내)

시스템 사용 기록: 시스템 접속 시간, 기능 사용 로그 등

설문 및 인터뷰 응답: 연구 참여에 대한 학생의 생각이나 느낌

#### 4. 개인정보 보호 및 비밀 보장

수집된 모든 개인정보는 철저히 익명화 처리(예: '학생A', '학생B')되어 개인을 식별할 수 없도록 관리됩니다. 모든 데이터는 암호화된 보안 서버에 저장되며, 오직 본 연구 목적으로만 활용될 것입니다. 연구 종료 후 5년 뒤 모든 원본 데이터는 관련 규정에 따라 안전하게 파기됩니다.

#### 5. 참여에 따른 이익 및 위험

본 연구 참여를 통해 귀하의 자녀는 자신의 잠재된 재능을 새로운 관점에서 발견하고 자기주도적 학습 능력을 향상시키는 긍정적 경험을 할 수 있습니다. 예상되는 위험은 없으나, 만약 연구 참여 과정에서 심리적 불편함을 느낄 경우 언제든지 참여를 중단할 수 있습니다.

#### 6. 참여의 자발성 및 철회 권리

본 연구 참여는 완전히 자발적이며, 참여를 거부하더라도 귀하의 자녀에게 어떠한 불이익도 없습니다. 또한, 연구가 진행되는 중 언제든지 동의를 철회하고 참여를 중단할 권리가 있습니다.

#### 7. 문의처

본 연구에 대해 궁금한 점이 있으시면 위 연구 책임자에게 언제든지 연락 주십시오. 연구 참여자의 권리에 대한 문의는 OO대학교 연구윤리위원회(IRB, 02-xxx-xxxx)로 연락하실 수 있습니다.

상기 내용을 충분히 숙지하였으며, 본인의 자녀가 위 연구에 참여하는 것에 자발적으로 동의합니다.

202X년 XX월 XX일

학부모/보호자 성명: \_\_\_\_\_ (서명)

참여 학생 성명: \_\_\_\_\_

연구 책임자 성명: OOO (서명)

#### [ 부록 1-B: 학생용 연구 참여 동의서 (Assent Form) ]

AI 친구와 함께 나의 새로운 재능을 찾아봐요!

안녕하세요, OOO 학생!

우리는 여러분이 공부하고 활동하는 모습을 보고 여러분의 특별한 재능을 찾아주는 'TEROS'라는 똑똑한 AI 친구를 만드는 연구를 하고 있어요.

이 연구에 참여하면 무엇을 하나요?

국어, 사회 시간에 프로젝트 과제를 할 때 TEROS 시스템을 사용해요.

여러분이 만든 보고서나 발표 영상을 TEROS에게 보여주고, TEROS가 어떤 조언을 해주는지 살펴봐요.

가끔씩 설문지를 작성하거나, 연구원 선생님과 AI를 사용한 느낌에 대해 잠깐 이야기할 수 있어요.

중요한 약속!

여러분의 이름이나 개인 정보는 절대 다른 사람에게 알려주지 않고, 비밀로 꼭 지킬게요.

이 연구에 참여하고 싶지 않으면 언제든지 "안 할래요!"라고 말할 수 있어요. 그래도 아무런 불이익이 없으니 걱정하지 마세요.

이 연구에 대해 부모님과도 이야기했고, 부모님도 허락해 주셨어요.

이제 000 학생의 생각이 궁금해요. 우리와 함께 재미있는 연구에 참여해 보지 않을까요?

☐ 네, 참여할래요.

☐ 아니요, 참여하고 싶지 않아요.

학생 이름: \_\_\_\_\_ (서명) 날짜: \_\_\_\_\_

## 부록 2. 설문지 (교사용 및 학생용)

본 부록은 시스템의 효과성을 정량적으로 측정하기 위해 사용된 사전-사후 설문지의 주요 문항입니다. 신뢰도가 검증된 표준화 척도를 본 연구의 맥락에 맞게 수정·보완하여 사용하였습니다.

### [ 부록 2-A: 교사용 설문지 (AI 신뢰도 및 협력 의도) ]

(5점 리커트 척도: 1=전혀 그렇지 않다, 5=매우 그렇다)

#### [지각된 유용성]

TEROS 시스템은 학생 개개인의 잠재력을 파악하는 데 도움이 된다.

TEROS 시스템은 나의 교수-학습 활동 설계에 유용한 통찰력을 제공한다.

#### [신뢰도]

3. 나는 TEROS 시스템이 제시하는 학생 분석 결과를 신뢰한다.

4. TEROS 시스템의 설명가능(XAI) 기능은 AI의 판단을 이해하고 믿게 만든다.

5. TEROS 시스템은 학생의 민감 정보(성별 등)에 기반한 편향된 진단을 내리지 않을 것이라고 믿는다.

#### [협력 의도]

6. 나는 TEROS 시스템의 분석 결과를 나의 교육적 의사결정에 적극적으로 반영할 의향이 있다.

7. 나는 AI가 제시한 ‘검증 요청 안전(TEROS-Loop)’에 참여하여 AI를 개선시키는 과정에 기여하고 싶다.

### [ 부록 2-B: 학생용 설문지 (학습 동기 및 AI 상호작용 경험) ]

(5점 리커트 척도: 1=전혀 그렇지 않다, 5=매우 그렇다)

#### [내재적 동기]

AI(TEROS)를 사용하여 프로젝트를 하는 것은 재미있고 흥미로웠다.

AI가 나의 새로운 재능이나 강점을 알려주었을 때, 스스로에 대해 더 긍정적으로 생각하게 되었다.

[지각된 유용성]

3. AI가 제공하는 피드백은 과제를 더 잘하는 데 실제로 도움이 되었다.

4. AI는 내가 미처 생각하지 못했던 나의 가능성을 발견하게 해주었다.

[상호작용 경험]

5. 나는 AI가 나의 활동을 공정하게 평가한다고 느꼈다.

6. AI가 왜 그렇게 평가했는지 이유를 설명해 주어서 이해하기 쉬웠다.

### 부록 3. 반구조화된 심층 면담 질문지 (교사용 및 학생용)

본 부록은 시스템 사용 경험과 인식 변화를 심층적으로 탐색하기 위해 활용된 교사 및 학생 대상 면담의 주요 질문 가이드입니다.

#### [ 부록 3-A: 교사용 심층 면담 프로토콜 ]

(도입) 안녕하세요, 선생님. 오늘 귀한 시간 내주셔서 감사합니다. 이번 학기 동안 TEROS 시스템을 사용해보신 경험에 대해 편안하게 이야기 나누고 싶습니다.

(초기 경험) 처음 TEROS 시스템을 접했을 때 어떤 인상을 받으셨나요? 사용하시면서 가장 유용했던 기능이나 어려웠던 점은 무엇이었나요?

(신뢰성 경험) AI가 제시한 학생 분석 결과가 선생님의 평소 생각과 일치했나요, 혹은 달랐나요? 특히 AI의 판단 근거를 보여주는 XAI 기능이 선생님의 신뢰에 어떤 영향을 미쳤나요?

(TEROS-Loop 경험) 연구 기간 중 AI가 선생님께 '검증 요청 안전'을 제시했던 사례가 있다면, 당시의 경험에 대해 말씀해 주실 수 있나요? (예: 박 군의 진로 예측 실패 사례) 그 과정에서 선생님의 역할은 무엇이라고 느끼셨나요?

(역할 변화) TEROS 시스템을 사용하시면서 학생들을 대하는 방식이나 선생님의 역할에 변화가 있었다고 느끼시나요? 그렇다면 어떤 변화인가요?

(결론 및 제언) 종합적으로 TEROS 시스템이 '지능형 교육 파트너'로서 가능성이 있다고 보시는지요? 앞으로 시스템이 어떻게 개선되면 더 좋을까요?

#### [ 부록 3-B: 학생용 심층 면담 프로토콜 ]

(도입) 안녕, 000 학생! 오늘 인터뷰에 와줘서 고마워요. 이번 학기에 썼던 'TEROS'라는 AI 친구에 대해 이야기 나눠보고 싶어요.

(전반적 경험) TEROS 시스템을 사용하면서 가장 기억에 남거나 재미있었던 경험이 있나요?

(재능 발견) AI가 000 학생에 대해 알려준 것 중에 ‘아, 나에게 이런 면이 있었구나!’ 하고 새롭게 알게 된 점이 있나요?

(AI 피드백) AI가 해준 조언이나 칭찬 중에 가장 도움이 되었던 것은 무엇인가요? 반대로 이해하기 어렵거나 내 생각과 다르다고 느꼈던 점도 있나요?

(수업 경험 변화) AI를 사용해서 프로젝트를 했던 수업은 보통 수업이랑 어떻게 달랐다고 느꼈나요?

(개선점) 만약 000 학생이 TEROS 시스템을 만드는 사람이라면, 어떤 기능을 더 추가하고 싶어요?

#### **부록 4. TEROS 시스템 프로토타입 주요 화면 예시**

본 부록은 논문에서 기술된 TEROS 시스템의 핵심 기능들이 사용자에게 어떻게 시각적으로 제공되는지 보여주는 프로토타입의 실제 화면입니다.

[부록 그림 4-1] 학생용 재능 프로파일링 대시보드

(설명: 학생의 다면적 재능(논리분석력, 창의적 표현력, 협업 소통력 등)을 육각형 능력치 그래프로 시각화하고, 각 능력치의 근거가 된 활동 목록을 제시하는 화면)

[부록 그림 4-2] 교사용 멀티모달 근거 시각화(XAI) 화면

(설명: AI가 '협업 소통력 우수'로 판단한 근거로, 발표 영상에서 특정 학생이 경청하는 장면(타임스탬프)과 보고서에서 역할 분담을 기술한 문단이 동시에 하이라이트 되어 있는 화면)

[부록 그림 4-3] 교사용 컨텍스트 어노테이션 입력 화면

(설명: 교사가 특정 학생의 활동 기록에 대해 '#컨디션난조' 태그를 선택하고 "몸살감기로 인해 평소보다 발표 참여가 저조했음"이라는 코멘트를 간편하게 입력하는 UI 화면)

[부록 그림 4-4] 교사-AI 공동 심의(TEROS-Loop) 안건 제시 화면

(설명: "새로운 지식 검증 요청: '데이터 분석 능력'과 '인문학적 스토리텔링'의 융합 재능"이라는 제목의 안건이 교사에게 제시되고, 근거가 된 누적 사례 데이터와 함께 '승인/기각/수정' 버튼이 활성화된 화면)