

생성 AI 기반의 평가 체계 및 지식 추적 방법 설계 연구*

Research on Design for the Assessment System and Knowledge Tracing Methods Based on Generative AI

조용상†

Yong-Sang Cho†

요약

이 연구는 디지털 환경의 빠른 변화에 대응하여 교육 분야에서 생성 AI의 적용 가능성을 탐구하며, 특히 학습 평가 체계와 지식 추적 방법의 혁신에 중점을 둔다. 연구 목적은 생성 AI를 활용해 교육 과정과 성취 기준을 체계적으로 탐색하고 학습자의 지식을 정교하게 추적할 수 있는 평가 체계와 지식 추적 모델을 설계하는 것이다. 이를 위해, 연구에서는 교육 내용을 인공지능이 해석 가능한 형태로 구조화하는 방안을 제시하고, 지식 체계와 독립적인 학습자 지식 추적 알고리즘을 통해 시스템의 유연성과 확장성을 강화하는 모델을 개발한다. 또한, 생성 AI를 기반으로 한 자동 문항 생성이 동적인 평가 체계에서 어떻게 활용될 수 있는지를 실증적으로 논의한다. 주요 결과로는 AI가 학습자 개인의 지식 수준에 맞추어 개인화된 평가 문항을 생성하고, 이를 통해 학습 효율성을 증대시킬 수 있는 방안을 도출한다.

주제어: 생성 AI, 평가 체계, 과정 중심 평가, 지식 추적, 지식 맵, 자동 문항 생성

ABSTRACT

This study explores the prospects of generative AI in the field of education in response to rapid digital changes, with a focus on innovating assessment systems and knowledge tracing methods. The purpose of the research is to design an assessment system and a knowledge tracing model that utilizes generative AI to systematically explore educational curricula and achievement standards, and intricately track learners' knowledge. To this end, the study proposes a method to structure educational content in a form interpretable by artificial intelligence and develops a model that enhances the system's flexibility and scalability through an independent learner knowledge tracking algorithm. Furthermore, it empirically discusses how generative AI-based automatic item generation can be utilized in dynamic assessment systems. The main result is the derivation of a method by which AI can generate personalized assessment items tailored to the individual knowledge levels of learners, thereby increasing learning efficiency.

Keywords: Generative AI, Assessment System, Process-centered Assessment, Knowledge Tracing, Knowledge Map, Automatic Item Generation

1. 서론

최근 우리 사회는 디지털 전환을 맞아 법·제도적인 거시적인 변화뿐만 아니라 일상생활 속에서도 많은 변화를 경험하고 있다[1]. ChatGPT와 같은 생성 AI의 등장은 사회의 변화 속도와 방향을 예측하기 어려운 정

도로 뒤흔들고 있는 것이 사실이다. 최근 미국에서 개최된 에듀테크 박람회 겸 컨퍼런스에서는 예전에는 경험해보지 못한 다양한 인공지능 서비스들이 선을 보였고, 교육 현장에서도 이러한 혁신을 적극적으로 실험하고 있다[2]. 국내 교육 정책도 같은 맥락에서 많은 변화를 예고하고 있다. 예를 들어 AI를 활용한 디지털 교과

†정 회 원: 한국열린사이버대학교 인공지능융합학과 교수

논문투고: 2023년 08월 13일, 심사완료: 2023년 11월 13일, 게재확정: 2023년 12월 17일

* 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. RS-2023-00229780, 맞춤형 교육을 위한 과정 중심 평가(학습진단) 인공지능 기술 개발)

서는 2022 교육과정이 적용되는 시기와 맞춰 일선 학교에 단계적으로 보급될 예정이라고 한다[3].

이러한 디지털 변화의 중심에는 생성 AI가 놓여 있으며, 그 영향력은 교육 분야에서도 점차 커지고 있다. 특히, 평가 체계와 지식 추적 방법에 대한 연구는 학습 경험과 효율성을 높이는 중요한 역할을 하고 있다[4]. 생성 AI를 활용한 평가 체계는 학습자의 개별적인 필요와 능력에 맞춘 맞춤형 평가를 실현하며, 지식 추적 방법은 학습 과정을 세밀하게 분석하고 구체적인 처방과 추천을 제공한다.

교육 현장에서의 인공지능 활용은 단순히 기술적인 혁신을 넘어서 교육의 본질과 가치에 대한 새로운 이해를 제시하고 있다. 학습자 중심의 교육, 개인화된 학습 경로, 지속적인 학습 여정에 대한 추적 등은 교육의 미래 지향적인 비전을 현실로 만들어 가고 있다 [1]. 블룸의 2-시그마 문제를 현 시점에서 재조명해보는 것은 인공지능을 바람직한 방향으로 활용하는데 좋은 이정표가 될 것이다.

이 연구에서는 생성 AI 기반의 평가 체계 및 지식 추적 방법에 대한 설계 방안을 다룬다. 특히, 최근의 기술 발전과 선행 연구 분석을 통해 인공지능이 평가에 어떻게 통합되고 있는지, 그리고 이로 인한 맞춤형 학습과 교육 혁신에 어떻게 기여하고 있는지에 대한 심층적인 분석을 제공한다. 이 연구의 목적은 교육 현장에서 생성 AI의 효과적인 활용 방안을 제시하고, 특히 AI 기반의 평가 체계와 학습자 지식 추적을 통해 맞춤형 학습 경험을 촉진하는 데 있다.

2. 연구 관련 동향 분석

최근 생성 AI 기술을 주도하고 있는 미국에서는 다양한 미디어 도구에 생성 AI 기술을 접목하면서 다양한 교육적 실험들을 하고 있다[5]. 동시에 생성 AI의 발전 속도가 현재뿐만 아니라 미래 사회 변화를 주도할 것으로 예상되는 바, 교육 분야의 이해관계자들은 미래 사회의 주역인 학생들을 어떻게 준비시킬 것인지에 대한 논의도 활발히 이루어지고 있다.

이 장에서는 AI와 더불어 살아갈 교육 분야의 이해관계자들이 주목해야 할 주제와 핵심 스킬에 대해서 간략히 알아보고, 교육 분야에서 다양한 실험과 활용 가능성을 점검하고 있는 생성 AI 도구들에 대한 동향을 소개한다. 새로운 평가 체계를 지원하기 위한 동향 분석에 앞서, AI 기반 교육 환경에서 무엇을 가르치고, 새

로운 평가체계에 적용하기 위해 사전에 필요한 스킬들이 무엇인지부터 고려해본다.

2.1 AI 기반 미래 교육에서 필요한 핵심 스킬

ISTE Live는 미국에서 연례적으로 개최되는 대규모 에듀테크 박람회 겸 교사 컨퍼런스이다. ISTE Live에서는 주목할 만한 교육 분야의 트렌드와 교육 현상이 직면한 이슈들이 잘 조명되고 있는데, 2023년의 핵심 주제는 “AI와 더불어 살게 될 디지털 세상에서 필요한 사람의 지능(human intelligence)과 추구해야 할 가치(humanity)” 라고 요약할 수 있다.

전 세계가 공통적으로 경험하고 있는 변화, 다시 말해 우리 사회를 빠르게 변화시키고 있는 요인은 두 가지로 요약할 수 있을 것이다.

- 빠른 속도로 확장되고 있는 연결성(dramatic increase of connectivity)
- 급격하게 성장하고 있는 인공지능의 능력(dramatic increase in the capabilities of AI)

예를 들어 아프리카 초원에서 관광객이 촬영한 동영상 상에서 동물 개체를 추적할 때 인공지능 비전 기술을 활용할 수 있는데, 전통적인 표본 중심의 동물의 개체 조사 방식보다 효과적일 수도 있다. 중국에서는 안면 인식 기술을 활용해서 실종 또는 유괴된 아이들 4000명을 찾아낸 사례가 보고된 바 있다. 인공지능이 냄새를 맡아서 분석하는 기술도 개발되고 있는데, 사람의 호흡을 분석해서 잠재적인 질병을 예측할 수 있다고 한다. 또한 인공지능의 생성 기술은 가상 인간을 통해서도 잘 드러나는데, 고인이 된 유명인이나 가족과 대화를 하면서 색다른 경험을 할 수 있다[2].



Figure 1. Examples using AI introduced in ISTE Live 23

이와 같이 급변하는 사회 속에서 교육 현장을 위해 시급히 답을 찾아야 하는 질문들이 제기되고 있는데, 다음은 정책을 결정하고 학교를 운영하는 리더들이 진지하게 고민해야 할 주제들이다[5].

- (학생을 위해) 학교에서 학습을 지원하기 위해 어떻게 AI를 활용할 것인가?
- (교사를 위해) 교육 활동과 학생에게 집중할 수

있도록 어떻게 AI가 지원할 것인가?

- (필요한 스킬) AI 기반 사회에서 학생들이 성공할 수 있는 스킬을 어떻게 준비시킬 것인가?

마지막 질문인 스킬에 관한 주제는 더 많은 현상 분석과 해법에 대한 실증이 필요한 것이지만, 현 시점에서 대략 다섯 가지 핵심 스킬로 정리해 볼 수 있다.

- skill 1. AI가 실제로 어떻게 동작하는지 이해
- skill 2. 인공지능이 새로운 아이디어를 떠올리게 어떻게 도울 수 있는지 배우기
- skill 3. 사람과 AI로 구성된 하이브리드팀이 일하는 방법
- skill 4. AI의 생성 능력을 활용하는 것은 창작보다는 큐레이션 활동이라는 점 이해하기
- skill 5. 더 나은 사람이 되는, 사람의 고유한 스킬 (공감, 정직, 창의, 친절, 사랑 등)

마지막에 언급한 스킬은 결코 컴퓨터로 대체될 수 없는 것들이므로 학생들에게 강조해야 할 스킬이 되어야 할 것이다.

2.2 생성 AI를 활용한 교육 서비스

2020년을 기점으로 거대 언어 모델(Large Language Model, LLM)에 대한 연구가 전 세계적인 주목을 받기 시작했는데, 그중에서 ChatGPT는 범용 언어 모델로써 가장 폭넓게 활용되고 있다. 그림 2는 한 문헌조사[6]에서 공개된 시기 별로 정리한 거대 언어 모델들인데, 거대 언어 모델은 기본적으로 100억 개의 파라미터 이상으로 구성된 모델을 말한다. 현재 일반에 공개된 대부분의 거대 언어 모델들은 범용적인 용도로 활용되고 있다. 소설, 시, 시나리오와 같은 문예창작활동으로도 활용되면서 동시에 다양한 직무의 업무용으로도 쓰이고, 또한 여가생활 같은 일상적인 목적으로도 활용되고 있다.

인공지능 분야에서는 ChatGPT에 처음 적용된 GPT-3부터 거대 언어 모델로 분류를 하는데, 거대 언어 모델은 거대한 학습 규모 외에도 이전 세대의 인공지능 모델들과 다르게 행동을 한다는 점에 주목할 필요가 있다. 예를 들어 거대 언어 모델들은 몇 가지 예시를 보여주면 스스로 예시를 통해 학습을 해서 예시와 같은 방식으로 새로운 답을 만들어 내는 능력을 갖게 되었는데, 이것을 급발현 능력(emergent abilities)라고 부른다[6]. 아직까지 거대 언어 모델에 급발현 능력이 어떻게 생기게 되었는지 명확히 설명할 수 있는 연구는 부족하지만, ‘프롬프트’ 라고 부르는 텍스트로

된 지시문과 답변의 맥락을 통해 스스로 학습한다는 점에서 분명 거대 언어 모델은 정해진 알고리즘으로만 학습하던 이전 세대의 인공지능과는 확연히 구분된다.

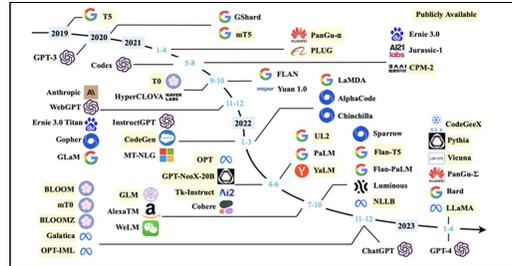


Figure 2. A Survey of Large Language Models[6]

그런데 교육 분야, 특히 학교에서 범용적인 거대 언어 모델을 활용하는 것은 정확성과 안전성 이슈 외에도 부정직하게 사용하는 것을 방지하는 윤리성 이슈 등 다양한 문제점들이 지적되고 있으므로 바람직한 AI 활용 방안에 대한 논의가 필요하다. 예를 들어 어떤 학생은 AI를 활용해서 작성한 과제 결과물의 출처와 방법을 밝히지 않고 자신의 결과물로 제출하는 사례는 AI 활용에 대한 윤리 교육의 주제가 될 수 있다.

교육 분야 비영리단체인 ISTE는 한 가지 대안으로 학교에서 활용할 수 있는 교육용 생성 AI인 ‘Stretch AI’ 라는 서비스를 개발하고 있다. 교사의 필요와 교육 과정과의 연계, 신뢰할 수 있는 교육자료 추천 및 링크 제공 등 정확성과 안전성을 담보할 수 있는 생성 AI를 곧 선보일 예정이라고 한다.

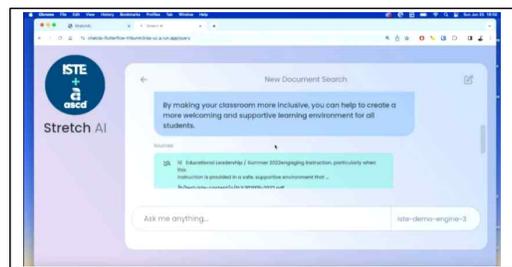


Figure 3. Stretch AI introduced at ISTE Live 23

이러한 교육용 생성 AI도 범용언어모델을 기반으로 교육 분야에 최적화된 답을 생성하도록 훈련을 시킨 결과물이라는 것은 흥미롭다. 완전히 새로운 교육용 생성 AI를 만드는 것이 아니라 범용언어모델을 기반으로 범위를 제한하고 답을 생성하는 방식과 생성된 답을

자체적으로 정제하고 검증하는 과정을 반영하는 의미로 이해할 수 있다. Stretch AI는 시범운영 기간 동안 교사들의 활용하는 과정에서 발생하는 질문과 생성되는 답을 데이터로 모아서 지속적으로 교육용 생성 AI에 훈련을 시킬 예정이라고 한다[2].

그 밖에 다양한 민간의 교육 서비스에도 생성 AI 기술이 적용되고 있다. 전형적인 모습은 미디어 저작 도구에서 콘텐츠를 생성하는데 생성 AI를 적용한 것인데, 사용자가 주제와 작성 형식을 생성 AI에게 프롬프트로 지시를 하면 결과물을 바로 생성해주는 방식이다. 그중에는 교사를 위해 수업설계서 작성과 수업자료 생성을 지원하는 기능도 포함된다. 아래 표 1은 미국의 교육 현장에서 활용중인 서비스들을 정리한 것이다.

Table 1. Educational services adopted generative AI

Services	Description
Canva	Canva, an online media authoring tool, utilizes Image Generator to create images for use in classes, and Magic Write to generate text content or ideas.
QuestionWell	From the website, “generates an endless supply of questions so teachers can focus on what matters. Just input some reading, and the AI will write Essential Questions, Learning Objectives, and aligned multiple choice questions.”
Curipod	From the website, “Create interactive lessons: Use polls, wordclouds, open questions, drawings and Q&A to capture the student voices. Make your own slides in Curipod, or upload your existing presentation.”
TLDR This	From the website, “TLDR This helps you summarize any piece of text into concise, easy to digest content so you can free yourself from information overload.”
Recast AI	From the website, “A new way to consume articles: Recast takes the hassle out of reading long articles, by turning them into entertaining, informative, and easy-to-understand audio conversations.”
ChatGPT	From the website, “We’ve trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.”
DALL-E	Generative AI that creates images using prompts. If you briefly describe the picture you want to draw in text, it is created in the form of a photo or drawing.

POE	When a request for lesson plan creation is made, it generates in the format of a lesson plan, then creates a recommended query for the next necessary action, and with just a click, it provides convenience by continuing to create teaching materials according to the request.
-----	---

살펴본 바와 같이 생성 AI를 활용한 교육 서비스들의 공통적인 지향점은 대략 세 가지로 정리할 수 있다. 첫째, AI는 교육이라는 구체적인 목적으로 사용하도록 범위를 제한하면서 안전성과 신뢰성을 담보로 운영되어야 한다. 둘째, 교육과정 및 성취기준과 같은 교육의 단계 또는 역량 수준에 맞춰서 대화 또는 자료를 제시한다. 복잡한 문제나 창의적인 접근이 필요한 문제를 다룰 때는 사람과 같은 수준의 대화 또는 상호작용을 제공한다. 이와 같은 생성 AI의 교육적 활용 가능성은 새로운 평가 체계에 대한 실험에도 영향을 주고 있는데, 이 연구를 시작하게 주요 배경이다.

3. 선행 연구

생성 AI 기반의 평가 체계를 설계하는 목적은 맞춤형 학습을 구체적으로 실현하고 학습 성과를 향상시키기 위한 것이므로, 선행 연구들 중에서 참조할 수 있는 적합한 모델을 찾는 것은 중요한 출발점이다. 선행 연구 중에서 블룸의 2-시그마 문제에서 관찰한 일대일 지도 학습 모델은 전문가 수준의 AI를 활용한 개별 학습자의 정밀한 진단과 학습자 지식을 추적하는 모델과 매우 유사하다고 판단했다.

또한 생성 AI를 활용한 평가 체계를 설계하기 위해서는 전통적인 문제은행에 기반을 둔 평가방식과는 차별화 된 새로운 접근법이 필요한데, 선행 연구 중에서 문항을 자동으로 생성하는 기술과 학습자의 지식을 실시간으로 추적하는 기술은 학습 과정 중에 학습자의 지식의 변화를 측정할 수 있는 좋은 도구 겸 접근법이다. 이와 같은 배경에서 세 가지 선행 연구를 간략하게 소개하고 다음 장에서 이들 모델을 개선한 방안들을 제시한다.

3.1 블룸의 2-시그마 문제(Bloom’s 2-sigma problem)

블룸의 2-시그마 문제는 벤자민 블룸(Benjamin Bloom)에 의해 1984년에 관찰된 현상을 말하는데, 일대

일 지도 학습 방법을 사용하여 교육받은 학생들의 평균 학습 성과(achievement)가 30명 정도로 구성된 교실 환경에서 교육받은 학생들의 평균 보다 월등히 높았으며, 2 시그마 수준(98%)에 도달했다는 것이다. 그림 4를 보면 교사 1명당 30명의 학생 비율을 가진 동일한 조건에서 숙달 학습(mastery learning) 방법을 적용한 학생들도 전통적인 교실 환경의 학생들보다 학습 성과가 높았는데, 1 시그마 수준(84%)에 도달했다고 한다[7].

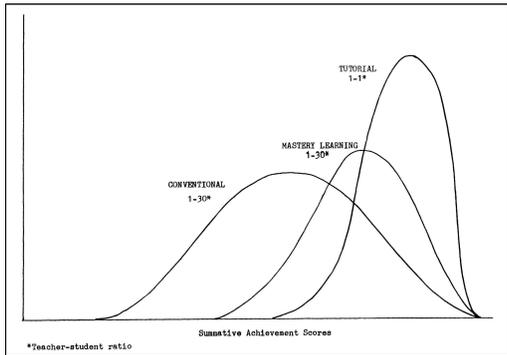


Figure 4. Achievement distribution for students under conventional, mastery, and tutorial instruction

숙달 학습은 학생이 특정한 지식 테스트에서 숙달 수준(예: 90%)을 달성해야 다음 주제를 학습할 수 있도록 어느 정도 통제를 하는 방식인데, 요구하는 지식수준에 도달하지 못하면 반복적으로 학습을 하거나 테스트를 다시 치러야 한다. 이와 같은 특성은 숙달 학습이 그룹 성과가 아닌 개인의 성과에 더 효과적일 수 있다는 점을 추리해볼 수 있다. 실제로 블룸의 실험에서도 숙달 학습 방법으로 일대일 지도 학습을 받은 학생들의 성취수준이 확연하게 높았던 이유도 숙달 학습과 일대일 지도 학습을 동시에 적용했기 때문이다. 따라서 교실에서 수행되는 그룹 단위의 교육 방법을 숙달 학습 및 일대일 지도 학습 방법과 같이 효과적으로 만들 수 있는 방법을 찾는 것이 앞으로 풀어야 할 과제이다.

3.2 자동 문항 생성(AIG) 기술

자동 문항 생성(Automatic Item Generation, AIG) 기술은 정해진 알고리즘 또는 인공지능을 활용해 제한 없이 문항을 자동으로 생성하는 기술을 말한다[8]. 자동 문항 생성 기술의 발전 과정은 두 건의 특허 분석을 통해서 간단하게 정리할 수 있다.

먼저 미국에서 개발된 자동 문항 생성기술인데, 심

리학적 평가와 컴퓨터 알고리즘을 결합한 모델을 기반으로 문항을 생성하는 과정이다. 이 방식은 평가 전문가가 ‘문항 모델’을 템플릿으로 제작을 하면, 컴퓨터 알고리즘이 문항 템플릿을 따라서 문항을 생성하는 과정으로 진행된다[4].

이 모델의 한계는 텍스트로 구성된 문항에만 적용할 수 있다는 점과 대부분 숫자나 일부 단어만 변경하는 문항 복제에 가까운 수준으로 문항이 생성된다는 점이다. 따라서 템플릿 문항에 그림을 삽입하거나 지문이나 지시문의 맥락을 변경하는 것은 어렵다.

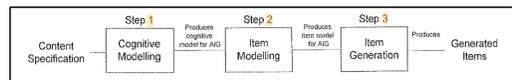


Figure 5. Automatic Item Generation (AIG) manufacturing process and system (US patent, US20140214385A1)

비교적 최근에 한국에서 특허로 등록된 자동 문항 생성 기술은 기존 기술의 문제를 해결하기 위해서 온톨로지 기술을 적용하였다. 온톨로지를 활용해서 지시문이나 지문에 있는 맥락을 유사한 맥락으로 대체하는 것이다. 예를 들어, “자전거가 시속 15km로 달릴 때.....” 라는 지문은 “자동차로 시속 80km로 달릴 때.....” 라는 형태로 자동으로 변형을 할 수 있다. 또한 교육과정의 내용 요소와 성취기준을 적용하여 교육과정에 기반한 문항을 생성할 수 있도록 조건을 설정한 것도 개선된 특징이다[4].

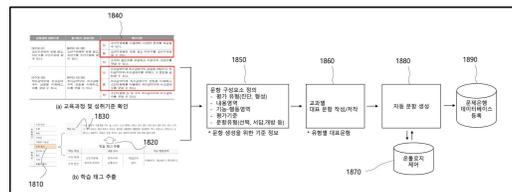


Figure 6. Automatic Item Generation (AIG) using curriculum standards and ontologies (Korea patent, 10-2197752)

이 모델 역시 실용화되지 못한 한계가 있는데, 온톨로지를 구축하는 시간과 비용이 만만치 않기 때문에 실제 서비스로 연결되지 못했다는 것과 생성의 기준이 되는 템플릿 문항에 그림을 포함하지 못한다는 것이다. 문항을 생성할 때 이미지를 포함해서 생성할 수 없기 때문에 자동 문항 생성의 범위가 제한될 수밖에 없는

한계는 여전히 극복하지 못한 문제로 남아 있다.

그림 7은 수학과 교육과정의 내용 일부를 ChatGPT와 대화하면서 확인해본 예시이다. 특정 주제에 대해서 문항을 생성하고, 생성한 문항을 풀 수 있는 대상 학년을 물어보고, 생성한 문항을 풀기 위해 필요한 기초 지식에 대해서도 확인을 할 수 있었다. 이 간단한 대화 속에는 언어모델에 축적된 지식의 구성요소를 추리해 볼 수 있는 단서들이 있다. 먼저 교육과정 분류체계 정보가 인공지능에 학습이 되어 있고, 각 분류체계에 따라서 성취기준에 해당하는 내용들이 추론할 수 있도록 연결되어 있다는 점이다. 이와 같이 교육 정보를 학습한 언어모델은 자연스럽게 일상적인 표현들로 문항을 생성함으로써 기존의 문항 생성 연구 수준을 단번에 뛰어넘었다[4].



Figure 7. Automatic Item Generation (AIG) results generated from ChatGPT

3.3 지식 추적 기술

3.3.1 심층 지식 추적(DKT)

자동으로 문항을 생성하는 목적 중 하나는 학습자의 지식을 세밀하게 진단하고, 적절한 처방을 하는 것이다. 학습자의 지식을 진단하기 위한 최근의 인공지능 기술은 심층 지식 추적(Deep Knowledge Tracing, DKT)

으로 알려져 있다.

심층 지식 추적은 그림 8에서 알 수 있듯이 딥러닝의 순환 신경망(Recurrent Neural Network, RNN)을 이용하는 장단기 기억(Long Shot-Term Memory, LSTM) 구조를 이용하여 학습자가 아직 풀지 않은 평가문항들에 대한 정답 확률을 예측한다[10].

이 모델은 객관식 문항에 적용할 수 있는데, 입력 x_t 는 학생의 행동이다. 학생의 행동에는 어떤 문제를 풀었는지, 어떤 답변을 선택했는지가 포함된다. 출력 y_t 는 각 문항을 맞출 확률을 벡터로 표현한 예측 벡터이다. 예를 들어, 다섯 개의 문항으로 구성된 시험이라면, 예측 벡터는 [0.7, 0.5, 0.9, 0.4, 0.6]과 같이 각 문제를 맞힐 확률로 표현된다.

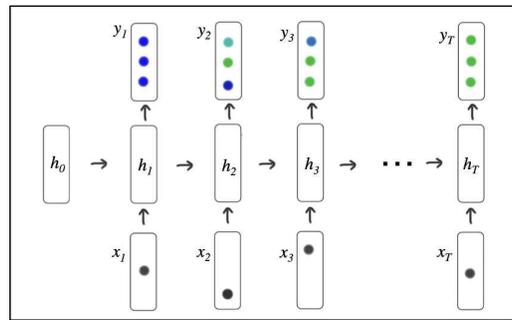


Figure 8. The connection between variables in a simple recurrent neural network[4]

이 모델은 학습자의 지식수준과 학습 진행 상황을 추적하고 분석하기 위해 인공지능 기술을 활용한 가장 대표적인 사례로 인정받고 있으며, 베이저안 지식 추적(Bayesian Knowledge Tracing, BKT)과 같은 통계 기반의 전통적인 지식 추적 알고리즘을 뛰어넘는 성능을 보여주었다[10].

심층 지식 추적 모델의 다른 장점 중 하나는 확장성이다. 이 모델이 처음 소개되었을 때부터 후속 연구들은 주로 수학 교과와 지식 추적을 우선 실험했지만 영어나 다른 교과들로도 확장할 수 있을 것으로 기대하고 있다. 반면 심층 지식 추적 모델은 RNN 구조가 가지고 있는 단점으로 인해 문항에 대한 정보를 담고 있는 시퀀스의 길이가 길어질수록 정답 예측 성능이 저하되는 한계도 가지고 있다[11].

3.3.2 지식 공간

학습자의 지식을 진단하고 추적하는 모델은 인공지

생성하는 문항은 다양한 수학 및 과학 기호를 표현할 수 있고, 이미지와 동영상도 포함할 수 있는 멀티모달 AI가 생성하는 문항의 형태가 되어야 한다. 4.2에서 그 방법과 주요내용을 제안한다.

두 단계의 기본적인 요건이 갖추어지면 이제 동적인, 다시 말해 다수의 문항이 포함된 사전에 제작된 시험지가 아니라 실시간으로 문항별로 호출되거나 문항이 생성되는 동적인 평가환경에서 학습자가 평가를 수행할 수 있도록 시스템을 설계하고 효과를 실증하는 과정으로 이어져야 한다. 4.3에서 그 방법과 주요 내용을 제안한다.

4.1 교육과정 및 역량체계 기반 지식 체계 설계

심층 지식 추적[10]과 지식 공간 기술[13]의 공통적인 목표는 현재 학습자의 지식 상태를 측정하고 취약 지식이 발견되면 원인을 찾아 처방을 하는 것이다. 부족한 취약 지식이 발견되면 처방적 학습 경로는 뒤쪽을 향하게 되며, 취약 지식 노드들을 중심으로 보충학습 문항을 풀거나 내용을 학습하도록 유도하게 된다. 반면 지금까지 필요한 지식을 잘 축적한 경우라면 앞쪽을 향해서 추천 학습 경로가 제시된다. 이와 같이 학습자별로 학습 경로를 수준 또는 상태에 따라 생성하는 것은 확정적인 경우의 수를 다루는 문제가 아니라 정해지지 않은 불확실한 상황을 다루는 문제이므로 인공지능을 활용해야 한다[1].

인공지능이 학습자의 지식 상태를 측정하고 원인을 추적하기 위해서는 탐색의 기준 정보가 되는 지식 체계가 필요하다. 3.3.2 지식 공간에서 소개한 ALEKS 사례처럼 지식 체계는 우선 교과별 내용 전문가들이 구성을 해야 하는데, 일반적으로는 교육과정 문서에 있는 내용 요소와 성취기준을 기본 정보로 구조화한다. 그 다음에 더 이상 쪼갤 수 없는 수준의 개념 수준까지 세분화하는 순서로 지식 체계를 구조적으로 완성해 나갈 수 있다[15]. 과정 중심 평가와 같이 변동성이 큰 상황에서 교과별 확장성을 높이려면 ALEKS와는 다른 접근법이 필요한데, ALEKS는 지식 공간을 수학적 집합으로 표현하고, 지식 공간에 알고리즘까지 적용한 통합 모델이기 때문에 확장성이 떨어진다.

대안으로 제시하는 접근법은 표준화된 방식으로 교육과정의 구성요소(내용 요소, 성취기준, 개념 등)를 알고리즘과 독립적으로 표현하여 마치 지도 데이터와 같은 역할을 하고, 학습자의 지식수준 또는 상태는 인공지능이 지식체계와 독립적인 알고리즘으로 추적할 수

있도록 역할을 분리하는 것이다. 지식 표현 부분과 추적 알고리즘 부분을 분리하면 교과별 확장성을 높일 수 있을 뿐만 아니라 교과별 알고리즘 개발도 병렬로 추진할 수 있어서 문제를 단순화 할 수 있을 것으로 기대한다.

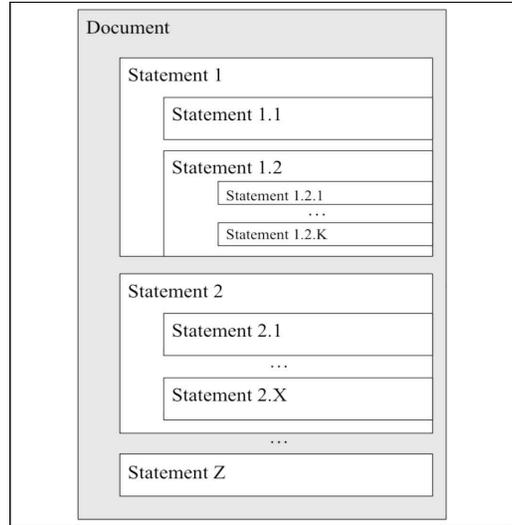


Figure 10. Logical structure for curriculum standards and statements using the CASE standard

이 연구에서는 교육과정과 역량체계를 표준화된 형식으로 표현하기 위해 관련 국제 표준인 CASE(Competency and Academic Standards Exchange)를 활용한다. CASE 표준을 적용하면 그림 10과 같이 교육과정과 성취기준의 구성 요소들은 논리적인 구조에 따라 집합적으로 표현될 수 있는 패키지 구조로 결합된다. 또한 CASE 표준은 계층적으로 구성된 교육과정의 구성 요소를 표현할 수 있는 데이터 모델을 정의하고 있으며, 각 계층 구조는 재귀적으로 연결될 수 있도록 관계를 표현함으로써 개념을 제한 없이 세분화할 수 있다[16].

CASE 표준은 미국에서 초중등 및 고등교육 분야에서 폭넓게 활용되고 있는 기술이지만, 국내 교육과정도 잘 표현할 수 있는지 확인하기 위해 그림 11과 같이 수학과 교육과정의 내용 요소와 성취기준 정보를 CASE 정보모델로 매핑해 보았다[3].

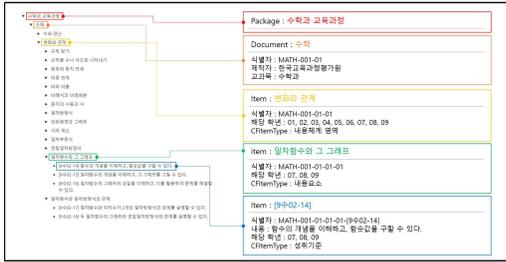


Figure 11. Mapping case between curriculum standards and CASE information model

교육과정은 자주 변경되는 정보는 아니지만 교육과정 문서에서 다루는 정보의 단위가 크고 추상적인 내용들이어서 실제 교육 서비스로 구현하기에는 한계가 있다. 일반적으로 교과서 발행사나 민간 에듀테크 기업들이 자체적으로 차시 단위까지 개념의 단위를 세분화하는 것도 그러한 한계가 있기 때문이다. 이 문제를 지식 체계 관점에서 보면 두 단계로 나누어서 접근할 수 있다. 먼저 교육과정 문서는 국가 수준에서 정해지는 것이므로 변경할 수 없는 고정된 지식 체계이다[3]. 반면 실제 서비스로 구현하는 지식 체계는 민간에서 자체적으로 확장하고 다시 재구조화할 수 있는 자유도가 높은 영역이 된다. 따라서 자유도가 높은 지식 체계 부분은 처음에는 내용 전문가가 구성을 하겠지만, 일단 구축된 이후에는 생성 AI가 수정하거나 추가하거나 재구성할 수도 있을 것이다.

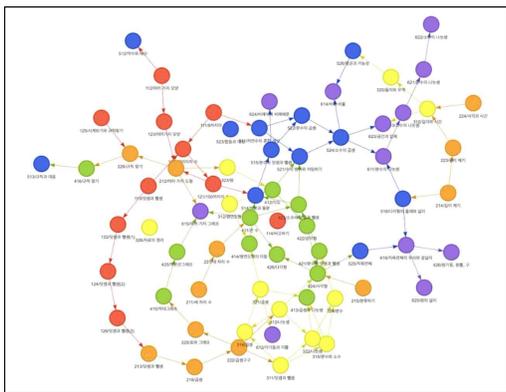


Figure 12. Visualization of knowledge map for Math curriculum generated by AI

그림 12는 인공지능이 생성한 지식맵의 한 부분을 시각화한 것이다. 예를 들어, 내용전문가들은 보통 기

하 영역의 지식을 구성할 때 수와 연산에 해당되는 내용을 연결하지 않지만, AI는 문항의 지문, 보기, 해설에 나와 있는 내용, 즉 맥락을 참조해서 기하 영역에 분수의 연산과 같은 개념을 스스로 연결할 수 있다[4]. 이와 같은 조정 작업은 실제 문항들이 시스템에서 운영되고 학습자들의 반응에 따라 생성된 데이터를 참조하여 수행되므로 계속 변하는 지식 체계로 다루어지게 된다.

4.2 멀티모달 생성 AI를 활용한 자동 문항 생성 워크스트림 설계

더 이상 쪼갤 수 없는 수준으로 세밀한 지식 체계가 확립이 되면 생성 AI를 활용해서, 구체적으로는 멀티모달 생성 AI를 활용해서 문항을 자동으로 생성할 수 있다. 모달리티(modality)는 어떤 것이 존재하거나 경험되거나 표현되는 방식을 의미하는데, 인공지능 분야에서 모달리티는 데이터 유형을 말한다. 멀티모달 AI란 텍스트, 음성, 이미지 또는 영상과 같은 여러 가지 형태의 데이터, 즉 모달리티를 동시에 처리하는 인공지능을 말한다. 예를 들어 문항 생성에 필요한 교육과정의 내용 요소와 성취기준을 조건으로 설정하고, 측정하고 싶은 내용을 프롬프트라고 하는 텍스트로 인공지능에게 전달하면, 생성 AI가 이미지가 포함된 텍스트 형태의 멀티미디어 문항을 생성하면서 동시에 정답과 힌트 등 하나의 문항을 구성하는 모든 메타데이터까지 자동으로 생성하는 과정이 멀티모달 AI를 활용하는 시나리오가 될 수 있다.

그림 13은 문항을 자동으로 생성하는 워크스트림을 설명한 것이다. 3.2 자동 문항 생성에서 설명한 한계를 극복하기 위해 문항 생성에는 멀티모달 생성 AI가 필요하다. 위에서 예로 설명한 바와 같이 문항의 지시문이나 지문에 포함된 맥락을 시각적으로 표현하거나 문항의 보기 내용에 시각 자료를 제시해야 할 경우 멀티모달 생성 AI를 활용하면 문항을 자동으로 생성할 수 있다.

그러나 멀티모달 생성 AI도 거대 언어 모델(Large Language Model, LLM)을 기반으로 한 범용적인 인공지능 모델로 공개될 것이므로 이 모델을 베이스 모델로 활용하고, 교육과정 및 성취기준 정보와 학습자의 지식을 측정하기 위한 대표적인 예시 문항들을 멀티모달 AI가 참조할 수 있도록 미세 조정(fine-tuning)을 시행해야 한다. 그림 13의 중간 부분과 아래 부분이 이 과정을 설명한 것이다. 미세 조정이란 범용적인 인공지능 모델이 특수한 목적을 정확하게 수행할 수 있도록

추가적인 데이터로 학습을 시키거나 예시를 학습시켜서 학습된 예시를 따라서 행동하도록 조정하는 과정을 말한다. 이 연구에서는 지식체계를 따라서 해당 주제나 성취기준 정보에 해당하는 문항의 내용을 참조하는 것과 대표문항을 따라서 유사한 문항들을 생성하도록 미세 조정하는 과정이 수행된다.

ChatGPT에 적용된 거대 언어 모델인 GPT 시리즈 외에도 그림 2에서 표현한 다양한 오픈소스 거대 언어 모델들을 이 연구에서 활용할 수 있으므로 기본적으로는 거대 언어 모델에 미세 조정을 해서 문항을 자동으로 생성하도록 프로토타입이 설계된다. 현재 연구 단계에서는 거대 언어 모델을 사용한 텍스트 형태의 문항만 생성할 수 있지만, 앞으로 1년 내 멀티모달 생성 AI도 오픈소스 모델 또는 실용적인 사용이 가능한 수준이 될 것으로 예상된다. 앞서 설명한 바와 같이 멀티모달 AI도 거대 언어 모델을 기반으로 확장된 인공지능 모델이므로 기본적으로 프롬프트를 활용해서 인공지능과 상호작용하는 방식이 동일하고 미세조정하는 과정도 유사하므로 인공지능 모델을 교체하더라도 연구 과정의 연속성을 담보할 수 있다.

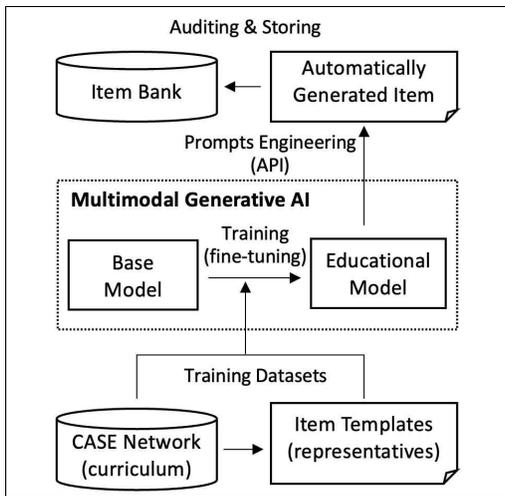


Figure 13. Workstreams for automatic item generation using multimodal generative AI

교육용 멀티모달 생성 AI는 인공지능에게 최적의 지식과 효과적인 답변을 유도할 수 있는 프롬프트 엔지니어링으로 설계된 API의 파라미터 조건에 따라 문항을 자동으로 생성하고, 생성된 문항은 내용전문가들의 검토와 확정을 거쳐서 문제은행에 저장된다. 인공지능이 생성하는 문항들의 신뢰성과 정확도가 사람이 제작

하는 문항의 수준과 동일한 수준에 도달하더라도 당분간은 문항의 품질관리 측면에서 전문가의 검토는 필수적으로 수행되어야 한다. 이 목적을 위해 실시간 문항 생성과 평가는 지양하고, 검증과 실증 기간에는 문제은행이 당분간 활용될 예정이다.

중장기적으로는 전문가의 문항 검토 방식과 기준을 학습한 인공지능 모델이 이 과정을 대신 수행하게 될 것으로 예상된다. 마치 2개의 인공지능 모델이 협력하는 방식으로 수행을 하게 될 텐데, 문항을 생성하는 AI와 문항을 검토해서 확정하는 분류 AI가 서로 독립적으로 동작하는 방식으로 연구를 진행할 예정이다.

4.3 동적인 맞춤형 진단 평가 환경 및 지식 추적 모델 설계

동적인 평가 환경은 기본적으로 두 가지 부분으로 구성되는데, 문항을 시각적으로 표현하고 전달하는 평가 엔진(test delivery engine)과 실시간으로 정오답을 판별하여 다음 문항을 결정하는 인공지능 엔진(adaptive engine)이 서로 통신을 하면서 동작을 하게 된다. 이러한 구조를 표준화 한 것이 CAT(Computer Adaptive Test) 표준이다[17]. 그림 14에서 알 수 있듯이 평가 엔진과 인공지능 엔진은 서로 독립적이고, 블랙박스처럼 동작하기 때문에 통신 방식은 API로 정해진 규칙에 따라 운영된다.

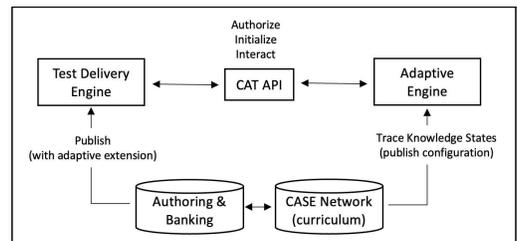


Figure 14. Workstreams to deliver assessment items using adaptive testing engine

그림 14는 CAT 표준에서 정의한 구조에 교육과정 및 역량체계 정보를 CASE Network로 표시하고, 이 지식체계를 기반으로 지식을 추적할 수 있도록 일부 변형을 한 것이다. 또한 인공지능 엔진은 객관식 문항, 단답형 문항, 서술형 문항과 같은 다양한 형태의 문항을 처리하고 채점할 수 있도록 언어 모델 기반의 생성 AI를 적용하여 실험을 진행하고 있다. 그러나 4.3에서 설명한 것처럼 손글씨 답안이나 이미지가 포함된 문항

등 다양한 포맷의 문항을 처리하기 위해서는 멀티모달 생성 AI로 대체되어야 한다.

그림 14의 인공지능 엔진(adaptive engine)은 지식체계(Case Network)와 독립적으로 동작하는데, 지식체계가 지도 데이터와 같은 학습 경로 맵 기능을 수행하고, 인공지능 엔진은 지식 체계의 맵 데이터를 참조하여 문항별 정오답 결과에 따라 다음 평가 문항 및 학습 경로를 탐색하고 조정하는 기능을 수행하게 된다. 이와 같이 지식체계와 알고리즘을 분리하여 평가환경을 구축함으로써 교과별 확장성뿐만 아니라 개발 기간 및 효율성을 개선하고 있다.

생성 AI가 적용된 동적인 평가 환경에도 거대 언어 모델이 적용되는데, 그림 14의 인공지능 엔진이 그 부분이다. 4.2에서 설명한 방식과 유사하게 프롬프트 엔지니어링 기법으로 인공지능 엔진에 설정된 기능으로 동작을 하게 된다. 한 가지 예시 상황을 그림 15로 설명할 수 있다. 그림 15의 'system' 블록은 생성 AI가 수행할 작업의 단계를 설정한 부분이다. 이 예시에서는 학생의 답안을 평가하기 전에 먼저 인공지능이 스스로 문제를 풀어보도록 설정을 했다. 인공지능이 잘못된 채점을 하지 않도록 일종의 생각할 시간을 먼저 주는 과정에 비유할 수 있다. 그 다음 단계는 학생의 정답과 인공지능이 계산한 답을 비교해보도록 지시를 했다. 세 번째 단계는 학생이 틀린 답을 제시했을 때 바로 정답을 보여주지 말고 어떤 힌트를 줄 것인지를 결정하라는 지시이다. 네 번째 단계는 실제 힌트를 학생에게 제시하는 것이다.

<p>SYSTEM</p> <p>Follow these steps to answer the user queries.</p> <p>Step 1 - First work out your own solution to the problem. Don't rely on the student's solution since it may be incorrect. Enclose all your work for this step within triple quotes ("").</p> <p>Step 2 - Compare your solution to the student's solution and evaluate if the student's solution is correct or not. Enclose all your work for this step within triple quotes ("").</p> <p>Step 3 - If the student made a mistake, determine what hint you could give the student without giving away the answer. Enclose all your work for this step within triple quotes ("").</p> <p>Step 4 - If the student made a mistake, provide the hint from the previous step to the student (outside of triple quotes). Instead of writing "Step 4 - ..." write "Hint:"</p>	<p>USER</p> <p>Problem Statement: A box contains 4 red balls and 6 blue balls. Two balls are drawn from the box without replacement. What is the probability that one of the drawn balls is red and the other one is blue?</p> <p>Student Solution: $(4/10) \times (6/9) = 24/90 = 4/15$</p>
<p>ASSISTANT</p> <p>Step 1 - To find the probability that one ball is red and the other is blue, we need to consider two cases:</p> <p>1) The first ball drawn is red and the second ball drawn is blue.</p> <p>2) The first ball drawn is blue and the second ball drawn is red.</p> <p>For the first case, the probability of drawing a red ball on the first draw is $4/10$. After the first ball is drawn, there are 9 balls left in the box, with 6 of them being blue. So the probability of drawing a blue ball on the second draw is $6/9$.</p> <p>For the second case, the probability of drawing a blue ball on the first draw is $6/10$. After the first ball is drawn, there are 9 balls left in the box, with 4 of them being red. So the probability of drawing a red ball on the second draw is $4/9$.</p> <p>To find the total probability, we need to add the</p>	

Figure 15. Use inner monologue or a sequence of queries to hide the model's reasoning process[19]

그림 15에서 모든 단계의 인공지능이 생성한 답변, 즉 프롬프트는 모두 독백 처리를 해서 사용자에게는

보여 지지 않는다. 필요한 상황에서만 생성한 답변이 보이게 프롬프트 엔지니어링을 함으로써 학습자와의 상호작용이 담보된 동적인 평가 환경을 구성한 예시이다. 그림 15의 'User' 부분은 문항과 학생이 제출한 답안이 있는 부분이고, 'Assistant' 부분은 인공지능이 설정된 단계에 따라 기능을 수행한 결과가 표시되어 있다.

그림 15의 예시 상황과 같이 동작하기 위해서는 4.1에서 제안한 지식 체계를 기반으로 노드라고 부르는 구체적인 내용 요소와 성취기준의 어느 한 부분을 인공지능이 참조할 수 있어야 한다. 바로 이 노드가 하나의 정보의 단위가 되어 인공지능이 학습자의 지식을 추적하면서 움직이는 기준 정보가 되는 것이다. 그 다음에 4.2에서 설명한 미세 조정된 인공지능 모델이 학습자의 수준에 따라 문항을 생성해서 전달하고, 4.3에서 설명한 예시처럼 학습자가 제시한 답안에 따라 힌트를 제시하거나 또는 정답과 함께 해설을 설명하는 방식으로 동적인 평가 체계가 설계되어야 한다. 4.3에서 소개한 예시와 같은 상황을 위해서도 상당히 다양한 상황과 예시들이 인공지능 모델에 미세 조정되어야 하는데, 교육현장의 실제 데이터들을 수집해서 미세 조정을 위한 데이터 형식으로 가공하는 일들이 수반되어야 한다. 4.1에서 제시한 지식 체계가 매우 정교하게 설계되었다고 4.2와 4.3에서 설명한 미세 조정이 교육현장의 실제 데이터로 정밀하게 수행되지 않으면 효과를 담보할 수 없기 때문이다.

5. 결론

이 연구에서는 교육 정책과 환경의 변화에 대비하기 위해 생성 AI를 평가 체계에서 활용하기 위한 프레임워크를 설계하였다. 과정 중심 평가와 같은 다양한 미디어 포맷과 평가 형식이 활용되는 정형화 되지 않은 평가 체계에서는 인공지능의 역할이 매우 중요해질 것이다. 인공지능이 안정적으로 신뢰할 수 있는 방식으로 동작할 수 있도록 이 연구에서는 먼저 교육과정과 역량체계를 기반으로 학습자의 현재 지식수준과 학습 후 지식의 상태 변화를 추적할 수 있는 기능을 가장 먼저 구체화하였다. 다양한 학생의 학습 준비도와 지식 상태를 측정하고 처방적 학습 여정을 설계하기 위해 생성 AI를 활용한 자동 문항 생성 기능도 중요한 역할을 하게 된다.

이 연구에서 제안한 평가 체계는 앞으로 현장 실증

과정을 거쳐 인공지능의 진단 및 예측 정확도를 실질적으로 평가하게 될 것이다. 현재 인공지능의 성능은 수학적인 알고리즘의 연산 결과로 판단하기 때문에 학생이나 교사가 체감하는 결과와는 다소 차이가 있는 것이 사실이다. 다시 말해 지식 추적 분야에서 인공지능의 가장 높은 수준의 성능(State Of The Art, SOTA)이라고 할 수 있는 'AUC(Area Under the Curve) > 0.90'이라는 표현이 소비자가 90% 이상 만족했거나 정확했다고 인식하는 것은 아니라는 것이다. 따라서 실증 과정을 통해 교육 현장에서 인공지능의 성능 평가를 보완할 수 있는 전문가 평가와 현장 평가 결과에 주의를 기울일 필요가 있다. 또한 거대 언어 모델과 상호작용하는 유일한 방식인 프롬프트 엔지니어링 기법과 미세 조정을 방법론으로 확립해서 교육 현장으로 빠르게 확산될 수 있도록 준비를 해야 한다.

생성 AI 기반의 평가 체계에 대한 실증에 더해 후속 연구로 추진하게 될 멀티모달 생성 AI의 활용 모델과 과정 중심 평가 결과물에 대한 자동 채점 및 피드백 기능은 더 도전적인 주제가 될 것으로 보인다. 하지만 이러한 일련의 생성 AI를 활용한 평가 체계의 혁신은 학생에게는 보다 정밀한 맞춤형 학습 경험과 함께 자신의 학습 여정을 스스로 설계할 수 있는 기회를 제공하고, 교사에게는 본연의 업무에 더 집중할 수 있도록 AI 보조교사와 같은 기능을 제공하여 학생을 더 잘 이해하고 지도할 수 있도록 도움을 줄 수 있을 것이다.

참고문헌

- [1] Cho, Y. S. & Gu, D. K. (2022). *Issues and Roles of Digital Transformation in Education domain of Newnormal era*. Daegu. NIA. ISBN 978-89-8483-706-5
- [2] ISTE Live 23. (2023). *Opening Mainstage*. https://conference.iste.org/2023/program/search/detail_session.php?id=116570809
- [3] Ministry of Education, et al. (2023). *Development Guideline for AI Digital Textbook*. ISBN 979-11-6555-369-2
- [4] Cho, Y. S. (2023). *Case study for technical innovation and standardization in Edtech and AI*. https://edu.tta.or.kr/edc/TTAReqstView.do;jsessionid=A5FB46C04E1DD01DF0F4B8FFAF44DDFD?eventId=EVENT_00000000002726
- [5] ASCD, ISTE, et al. (2023). *BRINGING AI TO SCHOOL: TIPS FOR SCHOOL LEADERS*. https://cdn.iste.org/www-root/2023-07/Bringing_AI_to_School-2023_07.pdf
- [6] Zhao, W. X., et al. (2023). *A Survey of Large Language Models*. <https://arxiv.org/abs/2303.18223>. <https://doi.org/10.48550/arXiv.2303.18223>
- [7] Bloom, B., (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*. 13(6)
- [8] Mark J. Gierl, Hollis Lai, et al. (2021). *Advanced Methods in Automatic Item Generation: Theoretical Foundations and Practical Applications*. Routledge. ISBN 978-0367458324
- [9] OpenAI. (2023). *ChatGPT*. <https://chat.openai.com>
- [10] Piech, C., Spencer, J. (2015). *Deep knowledge tracing*. <https://doi.org/10.48550/arXiv.1506.05908>.
- [11] Lee, C. G., et al. (2022). Research on the use of A.I to predicting learner knowledge state. *Proceeding of Korean Association of Computer Education*, 26(1), 181-183.
- [12] McGrawhill-Ed. (2017). *What makes ALEKS Unique*. https://www.aleks.com/about_aleks/tour_ai_intro
- [13] Doignon, JP Falmagne, JC. (2016). *Knowledge spaces and learning spaces*. DOI: 10.1017/9781139245913.006
- [14] Ministry of Education, et al. (2020). *Understanding Student Assessment for the Support of Teaching and Learning*. https://www.gne.go.kr/upload_data/board_data/workroom/164747073332896.pdf
- [15] Cho, Y. S., et al. (2020). *Learner competency measurement data in the field of mathematics*. AIHub. <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=133>
- [16] IEdTech. (2017). *Competency and Academic Standards Exchange*. https://www.imglobal.org/sites/default/files/CASE/casev1p0/information_model/case-servicev1p0_infomodelv1p0.html
- [17] IEdTech. (2020). *Computer Adaptive Test*. <https://www.imglobal.org/cat>
- [18] Cho, Y. S. (2022). *Understanding Edtech standard through Learning Impact 2022 and workstreams*. Daegu. KERIS.
- [19] OpenAI. (2023). *Use inner monologue or a sequence of queries to hide the model's reasoning process*. <https://platform.openai.com/playground/p/default-inner-monologue>



조 용 상

1995년 한림대학교 경영학과(경영학사)
 2001년 성균관대학교 정보통신공학전공 (공학석사)
 2009년 성균관대학교 전기전자및컴퓨터공학과 (공학박사)

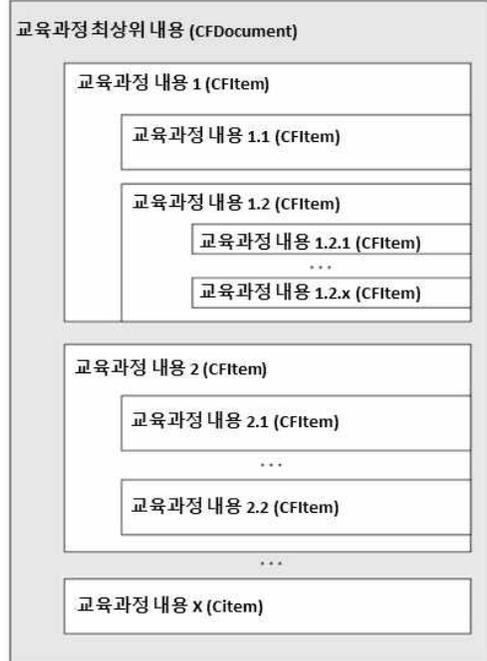
2002년 ~ 2017년 한국교육학술정보원 연구위원
 2017년 ~ 2022년 아이스크림에듀 부사장
 2022년 ~ 현재 한국열린사이버대학교 디지털비즈니스학과장 겸
 인공지능융합학과 교수

관심분야: 에듀테크, 학습분석, 평가기술, 인공지능, 표준화
 E-Mail: alzsang@ocu.ac.kr

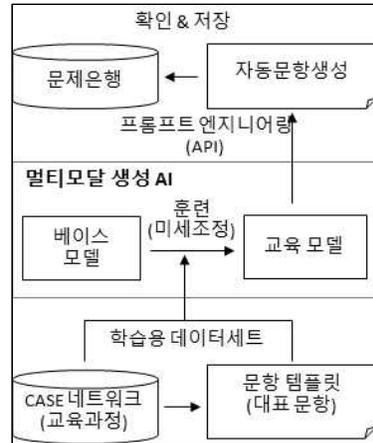
부 록

<표 1> 생성 AI를 적용한 해외 교육 서비스 사례

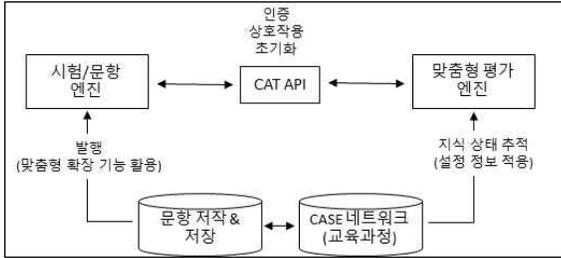
서비스명	설명
Canva	온라인 미디어 제작도구인 Canva는 Image Generator를 활용해서 수업에서 활용할 이미지를 생성하고, Magic Write를 활용해서 텍스트 콘텐츠 또는 아이디어를 생성
QuestionWell	교사의 수업 지원도구. 읽기 자료만 입력하면 AI가 필수 질문, 학습 목표 및 정렬된 객관식 질문을 작성
Curipod	대화형 수업 만들기: 설문 조사, 워드 클라우드, 개방형 질문, 그림, Q&A 등을 사용하여 학생의 반응을 포착. Curipod에서 수업 자료(슬라이드)를 생성하거나 기존 프레젠테이션을 업로드할 수 있음
TLDR This	모든 텍스트를 간결하고 이화하기 쉬운 콘텐츠로 요약하여 정보 과부하에서 벗어날 수 있도록 지원
Recast AI	기사를 소비하는 새로운 방법: Recast는 긴 기사를 재미있고 유익하며 이해하기 쉬운 오디오 대화로 변환하여 긴 기사를 읽는 번거로움을 덜어주는 편의성 제공
ChatGPT	대화 방식으로 상호 작용하는 ChatGPT라는 모델을 훈련. 프롬프트라는 대화 형식을 통해 ChatGPT는 후속 질문에 답하고, 실수를 인정하고, 잘못된 전제에 이의를 제기하고, 부적절한 요청을 거부할 수도 있음
DALL-E	프롬프트를 활용해서 이미지를 생성하는 생성형 AI. 그리고 싶은 그림에 대해 텍스트로 짧게 묘사를 하면 사진 또는 그림의 형태로 생성
poe.com	수업설계서 작성 요청을 하면, 수업설계서 포맷으로 생성을 해준 다음, 필요한 다음 액션을 추천 질의로 생성하고 클릭만 하면 요청에 따라 수업자료까지 이어서 생성을 해주는 편의성 제공



[그림 10] CASE 표준의 교육과정 표준과 성취기준을 표현하는 논리적인 구조 개념



[그림 13] 멀티모달 AI를 활용한 자동문항생성 워크스트림



[그림 14] 평가 문항을 전달하는 맞춤형 평가 엔진 워크스트림

<p>SYSTEM</p> <p>사용자 쿼리에 응답하려면 다음 단계를 따라야 해.</p> <p>1단계 - 먼저 문제에 대한 너의 답안을 만들어, 학생의 답안이 틀릴 수 있으므로 의논하면 안돼. 이 단계의 모든 작업을 세 개의 따옴표(“”)로 묶어.</p> <p>2단계 - 학생의 답안과 너의 답안을 비교하고 학생의 답안이 정확한지 평가해. 이 단계의 모든 작업을 세 개의 따옴표(“”)로 묶어.</p> <p>3단계 - 학생이 실수를 했다면 답을 포기하지 않고 학생에게 어떤 힌트를 줄 수 있는지 결정해. 이 단계의 모든 작업을 세 개의 따옴표(“”)로 묶어.</p> <p>4단계 - 학생이 실수한 경우 이전 단계의 힌트를 학생에게 제공해. (상중따옴표 제외). “4단계 - ...”라고 쓰는 대신 “힌트”라고 써서 제공해.</p>	<p>USER</p> <p>문제: 상자에 빨간색 공 4개와 파란색 공 6개가 들어 있습니다. 공 2개를 교체하지 않고 상자에서 꺼냅니다. 뽑은 공 중 하나는 빨간색이고 다른 하나는 파란색일 확률은 얼마입니까?</p> <p>학생 답안: $(4/10) * (6/9) = 24/90 = 4/15$</p> <p>ASSISTANT</p> <p>1단계 - 너의 답안: 4/15</p> <p>2단계 - 답안 평가: 학생의 답안은 맞습니다. 빨간색 공을 뽑은 후에 파란색 공을 뽑을 확률은 $(4/10) * (6/9)$로 계산할 수 있습니다.</p> <p>3단계 - 힌트: 학생의 답안은 정확합니다. 추가적인 힌트는 필요하지 않습니다.</p> <p>⊙ Add message</p>
--	--

[그림 15] 프롭프트 엔지니어링 기법 중 하나인 내부 독백 기능으로 모델의 추리 과정을 숨기고 필요한 부분만 보여주는 예시[19]