

## A Study on ways to strengthen AI literacy capabilities for high school students through explainable AI education

Eun-Sook Jung\*, Han-Bin Lee\*\*, Kwi-hoon Kim\*\*\*

\*Master of Artificial Intelligence Convergence Education, Graduate School of Education,  
Korea National University of Education, Chung-buk, Korea

\*\*Ph.D. Candidate, Dept. of Elementary Computer Education, Korea National University of Education,  
Chung-buk, Korea

\*\*\*Professor, Dept. of Computer Education, Korea National University of Education, Chung-buk, Korea

### [Abstract]

This study focused on developing an artificial intelligence education program for high school students and verifying its effectiveness by reflecting changes in the curriculum and differentiation by school level at the present time when the importance of artificial intelligence is increasing. Education was organized to understand core principles and develop an ethical perspective using explainable artificial intelligence (XAI), and a teaching and learning model centered on CT (Computational Thinking) elements was developed based on the ADDIE model. As a result of application, the average score of all AI literacy sub-factors improved, especially in the area of 'understanding artificial intelligence'. This shows that XAI-based education is effective in cultivating artificial intelligence literacy among high school.

▶ **Key words:** AI, AI literacy, Explainable Artificial Intelligence, ADDIE

### [요 약]

본 연구는 인공지능의 중요성이 높아지는 현시점에서 교육과정 변화와 학교급별 차별성을 반영하여 고등학생 대상 인공지능 교육 프로그램을 개발하고 효과를 검증하는 데 초점을 두었다. 설명 가능한 인공지능(XAI: eXplainable Artificial Intelligence)을 활용해 핵심 원리를 이해하고 윤리적 시각을 기를 수 있도록 교육을 구성했으며, CT(Computational Thinking) 요소 중심의 교수학습 모형을 ADDIE 모형 기반으로 개발하였다. 적용 결과, 모든 인공지능 리터러시 하위 요소의 평균 점수가 향상되었고 특히 '인공지능의 이해' 영역에서 가장 큰 폭의 성장이 나타났다. 이는 XAI 기반 교육이 고등학생의 인공지능 리터러시 함양에 효과적임을 보여준다.

▶ **주제어:** 인공지능, 인공지능 리터러시, 설명 가능한 인공지능, ADDIE 교수학습 모형

- 
- First Author: Eun-Sook Jung, Corresponding Author: Han-Bin Lee, Kwi-hoon Kim
  - \*Eun-Sook Jung (dooyee@korea.kr), Graduate School of Education, Korea National University of Education
  - \*\*Han-Bin Lee (lhb718@korea.kr), Dept. of Elementary Computer Education, Korea National University of Education
  - \*\*\*Kwi-hoon Kim (kimkh@knue.ac.kr), Dept. of Computer Education, Korea National University of Education
  - Received: 2025. 06. 13, Revised: 2025. 07. 18, Accepted: 2025. 08. 14.

## I. Introduction

앞으로 인공지능과 함께 살아가야 할 미래 세대의 학생들에게 인공지능에 대한 학습은 더 이상 선택이 아닌 필수적인 요소가 되었다. 교육부(2022)는 미래 사회의 변화에 효과적으로 대응하기 위한 교육과정 혁신을 핵심 과제로 제시하면서, 언어·수리·디지털 소양을 기본 역량으로 강조하고 있다. 특히 디지털 소양 강화를 위한 방안으로, 학교급별 발달 수준에 맞춘 내용 기준을 마련하고, 전 교과에 걸쳐 디지털 역량을 강화하는 방향으로 교육과정을 개편하고자 하고 있다[1].

또한, 모든 교과에서 디지털 기초 소양을 함양하고 디지털 융합 수업이 이루어지도록 하며, 정보 교과에서는 기초 코딩과 컴퓨팅 도구 활용을 통해 정보 처리 능력과 AI·SW 관련 개념 및 원리 학습이 가능하도록 하고 있다. 이를 바탕으로 심화 학습을 통해 전문 역량을 강화하며, 정보 과목은 인공지능의 개념과 원리를 포함하도록 교과 내용을 재구성하고 '인공지능에 대한 학습(learning about AI)'을 강화하고 있다[2].

교육 분야에서 인공지능을 다루는 방식은 크게 두 가지로 나눌 수 있다. 하나는 인공지능 기술을 교육 도구로 활용하여 학습을 지원하는 '인공지능과 함께하는 학습(Learning with AI)'이며, 다른 하나는 인공지능 기술의 개념과 활용 방법 자체를 배우는 '인공지능에 대한 학습(Learning about AI)'이다[3]. 인공지능에 대한 학습은 인공지능의 개념과 원리를 교육함으로써 미래 사회에서 요구되는 핵심 역량을 기르는 것을 목적으로 한다. 그러나 사회 변화에 대한 관심이 높아지는 상황에도 불구하고, 2021년 이전까지의 연구 동향을 살펴보면 초·중등 교육 현장에서 인공지능의 개념이나 원리에 대한 소개는 거의 이루어지지 않은 실정이었다[4].

최근 사회적 요구와 교육과정 변화에 따라 초·중등 교육에서 인공지능 개념과 원리를 다루는 연구가 진행되고 있지만, 대부분 노코드 기반 도구를 활용한 모델 제작과 활용에 그치고 있다. 개념 및 원리 중심의 연구도 증가하고 있으나, 주로 언플러그드 활동에 의존하고 있으며, 학교급간 교육 내용과 방법의 구분이 뚜렷하지 않은 실정이다.

설명 가능한 인공지능 교육은 인공지능의 작동 원리뿐 아니라 활용 방식과 윤리적 고려까지 포괄하기 때문에, 인공지능 교육에서 핵심적인 요소로 다루어져야 한다는 연구가 존재한다[5]. 하지만 설명 가능한 인공지능 교육은 비교적 최근에 등장한 기술이기 때문에 이를 교육에 적용한 연구 사례는 아직 많지 않다. 특히, 학교급에 따른 차별성

을 고려하면서 고등학생의 수준에 적합한 인공지능 개념 및 원리 학습에 관한 연구가 요구된다.

이에 본 연구는 설명 가능한 인공지능을 주제로 교육 프로그램을 개발하고 실제 수업에 적용함으로써, 고등학생의 인공지능 리터러시 향상에 어떤 영향을 미치는지 탐색하는 데 목적이 있다. 이러한 목적을 바탕으로 본 연구에서는 다음과 같은 연구 문제를 설정하였다.

첫째, 고등학생의 인공지능 리터러시 향상을 위한 설명 가능한 인공지능 교육 프로그램은 어떻게 구성되어야 하는가?

둘째, 설명 가능한 인공지능 교육 프로그램은 고등학생의 인공지능 리터러시를 향상시키는가?

## II. Theoretical Background

### 1. Artificial intelligence literacy

인공지능을 활용하는 시대적인 변화에 맞추어 인공지능 기술을 활용하고 이해하는 역량을 의미하는 인공지능 리터러시는 인공지능 기술의 개념과 작동 원리를 이해하고 인공지능 기술을 적극적으로 활용하거나 인공지능과 적극적으로 협력하여 실생활 문제를 효율적으로 해결할 수 있는 능력을 의미한다[6]. 한국과학창의재단(2020)에서는 '인공지능 교육 길라잡이'에서 인공지능 기술을 활용해 창의적인 결과물을 만들어 낼 수 있는 역량을 인공지능 리터러시라고 보았다. 한편으로는 우리 주변과 다양한 산업에서 활용되는 인공지능 서비스와 제품을 알아보고 건강하게 활용하여 해당 분야의 작업 효율성을 향상시킬 수 있는 역량을 인공지능 리터러시라고 일컫는 사람도 있다. 이윤희 외(2021)는 인공지능 기술 매체를 이해하고 활용하는 능력과 이를 활용해 사회와 소통하는 능력, 인공지능 사회에 대한 비판적 생각을 할 수 있는 능력을 인공지능 리터러시라고 정의하였다[7]. 인공지능 리터러시를 다양하게 정의하고 있는 것과 마찬가지로 인공지능 리터러시를 구성하는 하위 요인 또한 여러 가지로 제시되고 있다.

황현정(2022)은 인공지능 리터러시를 상위 차원과 하위 차원으로 나누어 분석하면서 지식적 차원에서 AI 이해와 AI 식별, 기술적 차원에서 AI 분석, AI 활용, AI 창조, 태도적 차원에서 AI 이용 윤리성과 AI 비판적 평가의 7가지로 구체적으로 제시하였다[8].

## 2. Explainable Artificial Intelligence

안재현(2020)은 설명 가능한 인공지능은 인공지능이 결과를 도출하기까지 어떤 과정과 처리를 거쳤는지, 결과의 근거가 무엇인지를 사람이 이해할 수 있는 형태로 설명해주는 기능을 포함한 인공지능 시스템 또는 인공지능의 결과에 대해 설명을 제공하는 기법이라고 하였다.

설명 가능한 인공지능(XAI: eXplainable Artificial Intelligence)은 높은 수준의 학습 능력과 정확도를 유지하면서 사용자가 협업자로서 인공지능을 이해하고 적절하게 활용하여 효과적으로 관리하는 것을 목적으로 개발하고 있다. 설명 가능한 인공지능은 각 단계의 결정에 대한 판단 근거를 설명하고 전반적인 약점과 강점을 이해시키며 그 시스템이 미래 어떻게 작동할지를 알린다. 그리고 시스템 오류 수정 방법을 알릴 수 있다.

설명 가능한 인공지능을 구현하는 기법은 크게 두 가지로 나누어 볼 수 있다. 첫 번째는 투명한 모델을 사용하는 방법이다. 투명한 모델은 모델 자체가 스스로의 동작을 설명할 수 있는 특성이 있다. 기계학습 알고리즘들은 데이터가 주어졌을 때 알고리즘에 따라 처리하여 결론을 출력하며, 별도의 설명을 추가하지 않더라도 인공지능의 처리 과정을 사람이 이해할 수 있는 형태로 동작한다. SVM이나 k-NN, k-Means, 의사결정트리, 나이브 베이즈, 선형회귀, 로지스틱 회귀 등 대부분의 기계학습 알고리즘이 투명한 모델로서, 동작 과정을 사람이 이해할 수 있다.

두 번째는 블랙박스 인공지능 모델에 대하여 모델이 만들어진 이후에 설명을 추가하는 사후 설명(post-hoc) 방식이다. 기계학습 알고리즘과 달리 딥러닝 알고리즘들은 기계학습 알고리즘들에 비해 훨씬 더 높은 성능(정확도)을 보이지만 인공지능의 동작 과정을 사람이 이해하기 어렵거나 불가능하다. 이런 경우 인공지능이 내놓은 결과에 대해 다양한 방법으로 사람이 이해할 수 있는 형태의 설명을 덧붙여서 사용자를 이해시키고 인공지능의 결과를 활용하게 된다.

## 3. DPAAP CT Element-Focused Learning Model

DPAAP CT 요소 중심모델은 SW 교육을 쉽고 효과적으로 할 수 있도록 문제해결 학습법에 기반을 두고 컴퓨팅 사고력을 기르고 SW 교육의 특징을 잘 살릴 수 있도록 개발되었다. 문제해결 학습법의 절차는 문제 확인, 문제 해결 방법 도출, 문제해결, 일반화의 네 단계 교수학습 절차를 따른다.

Table 1. Overview of the DPAAP Model

Decomposition	• Break down problems into computer-resolvable units
Pattern Recognition	• Exploring recurring constant trends and rules
Abstraction	• Simplifying problems, formulating principles discovered by pattern recognition
Algorithm	• Procedural construction of abstracted core principles
Programming	• Run in a language that your computer can understand

김진숙(2015)은 CT 요소 중심모델은 각각의 단계를 가지고 있는 절차식 모델로 볼 수도 있고, 다른 한편으로는 다양한 요소와 절차로 결합하는 모듈식 모델로 볼 수 있다. 필요한 경우 자료를 수집·분석하고, 시뮬레이션, 자동화 등의 여러 전략을 추가할 수 있다고 했고, 본 연구에서는 Table 1에 제시된 DAPPA 모델에 컴퓨터과학 원리를 지루하지 않고 쉽게 전달하는 데에 효과적인 언플러그드 활동을 추가하여 수업 모델을 구성하였다[9].

## III. Methods

### 1. Research Procedures

본 연구는 일반고등학교 2학년 학생 20명을 대상으로 ADDIE 교수 설계 모형에 따라 CT 요소 중심모델을 활용한 4차시의 설명 가능한 인공지능 교육 프로그램을 개발하였다. ADDIE 교수학습 프로그램 개발 모형은 Table 2과 같다.

Table 2. Instructional Program Development Procedures and Contents Using ADDIE Model

Step	Details
Analysis	• Analysis of learner characteristics • Artificial Intelligence Education Program Content Analysis
Design	• Setting the class objectives of the AI education program
Development	• Draft AI Education Program Classes • Development of educational content and learning materials • Program modification and development through expert group review
Implementation	• Application of Artificial Intelligence Education Program • Modify and improve based on feedback
Evaluation	• Evaluation and data analysis to validate effectiveness • Analysis of the Effectiveness of Artificial Intelligence Education Programs

### 2. Research Subjects and Research Design

개발한 수업 모형이 인공지능 리터러시 향상에 미치는 영향을 측정하기 위하여 단일집단에 대해 사전/사후 검사 결과의 대응표본 t-검정을 실시하도록 설계하였다. 실험 설계의 내용을 요약하면 Table 3과 같다. 본 연구는 실험 집단(20명)에 대해 사전조치 8차시와 설명 가능한 인공지능 교육 프로그램을 사용한 수업 4차시, 사전 설문조사 및 사후 설문조사로 이루어졌다.

본 연구는 대조집단이 없는 단일집단 사전-사후 설계로 진행되었기 때문에, 시간 경과나 반복 학습 등 외생변인에 의한 영향을 완전히 배제하기 어렵다. 이러한 설계상의 제약은 내적 타당도 확보에 한계를 줄 수 있으며, 향후 연구에서는 통제집단을 포함한 실험 설계를 통해 결과 해석의 신뢰성을 높일 필요가 있다.

Table 3. Experimental design

G1	A1	O1	X1	O2
G1: Experimental group (1 class, 20 people)				
A1: Pre-action (the 8th session of the existing artificial intelligence class)				
O1: Artificial Intelligence Literacy Pre-Test				
X1: Explainable AI training program (4th session)				
O2: Artificial Intelligence Literacy Post-Test				

또한 연구 참여자는 모두 일반계 고등학교 2학년으로 Python 프로그래밍 기초 경험이 있고 ‘인공지능 기초’ 과목을 수강 중인 학습자이다. 이들은 비교적 학습 동기와 사전 준비도가 높은 집단으로 연구 결과를 모든 고등학생 집단에 일반화하는 데에는 한계가 있다.

### 3. Inspection Tools

본 연구에서는 고등학생의 인공지능 리터러시를 측정하기 위해 김성원·이영준(2021)이 개발한 ‘중학생의 인공지능 리터러시 검사 도구’를 기반으로 하였다[10]. 기존 도구의 전반적인 문항과 측정 영역은 유지하되, 연구 대상이 고등학생으로 변경함에 따라 학교급만 수정하여 적용하였다. 전문가 3인이 검토를 수행하였으며, 이를 통해 타당도를 확보하였다. 또한, 본 연구에서 실시한 사전-사후 검사 자료를 토대로 산출한 Cronbach’s  $\alpha$  계수는 .85로, 측정 도구의 내적 일관성이 양호한 수준임을 확인하였다.

다만, 본 도구는 고등학생을 대상으로 한 별도의 표준화 절차가 충분히 이루어지지 않았기 때문에, 검사 결과를 해석할 때 이러한 제한점을 고려할 필요가 있다. 검사 도구는 Table 4와 같이 인공지능 리터러시의 6개 하위 영역으로 구분된 총 30문항으로 구성되었으며, 각 문항은 5점 리커트 척도(Likert scale)로 응답하도록 설계되었다.

Table 4. Artificial Intelligence Literacy Inspection Tool Sub-elements

Sub-elements	Question number	type
AI Social Impact	1~8	Likert a five-point scale
AI Action Plan	9~13	
Solving AI Problems	14~18	
Understanding AI	19~24	
Data Literacy	25~28	
AI ethics	29~30	

### 4. Development of Learning Program

ADDIE 교수설계 모형에 따라 CT 요소 중심모형을 활용한 4차시의 설명 가능한 인공지능 교육 프로그램을 개발하였다. 프로그램을 적용하기 전 사전 조치로 인공지능의 개념과 발전 과정, 인공지능의 학습, 기계학습의 종류와 특징, 인공지능 분류 모델을 활용한 인공지능시스템 실습을 포함한 인공지능 학습을 8차시에 걸쳐서 실시하였다.

#### 4.1 Analysis

분석 단계에서는 학습자 특성, 설명 가능한 인공지능 교육 프로그램 콘텐츠, 2022 개정 교육과정 ‘인공지능 기초’ 과목의 성취기준을 분석하였다.

대상 학급은 일반고등학교 2학년 남녀 혼합반으로 지난 학년도에 ‘정보’ 과목을 수강하여 파이썬 프로그래밍에 대한 기초 지식과 프로그래밍 경험이 있는 학생들이며, 현재 진로선택으로 ‘인공지능 기초’ 과목을 선택하여 수강하고 있는 학생 20명으로 설정하였다.

다음으로 프로그램의 콘텐츠를 분석하였다. 김수환(2020)의 연구에 따르면 많은 인공지능 교육과정은 대부분 알고리즘의 이해보다는 ‘활용과 응용’을 다루고 있고, 기계학습을 이용한 ‘음성, 이미지, 텍스트 인식’을 주로 실생활 사례와 연계하여 다루고 있다[11]. 그 방법으로 기존의 코딩 도구에서 음성, 이미지, 텍스트 인식과 같은 인공지능 모듈을 블랙박스 형태로 제공하고 Fig. 1과 같이 API를 통해 불러와서 사용할 수 있도록 하고 있다.

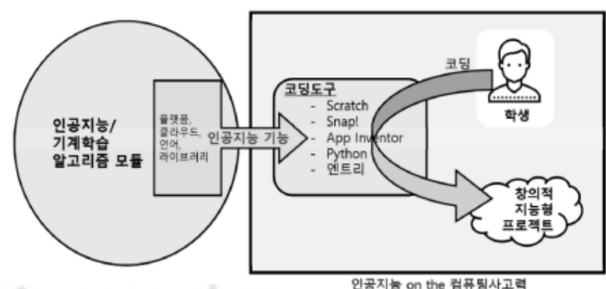


Fig. 1. API usage tool case

같은 연구에서 소개한 Fig. 2의 웹 기반 플랫폼 사례를 통한 학습에서는 학습자가 자신의 데이터를 활용하여 의도하는 인공지능 모델을 직접 설계·구현해 볼 수 있다. 이러한 접근은 학습자 주도의 탐구 활동을 가능하게 하지만, 여전히 인공지능의 작동 원리보다는 인공지능의 활용 측면에 더 초점이 맞추어져 있다.

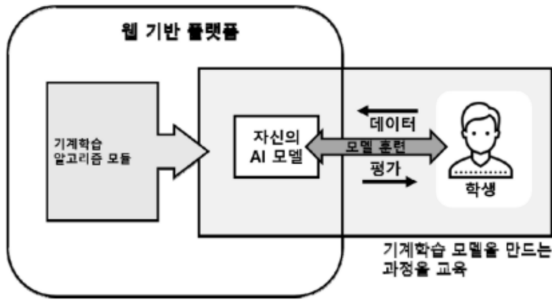


Fig. 2. Web-based platform example

2022 개정 교육과정의 ‘인공지능 기초’ 과목의 내용은 Table 5와 같으며, 1단원인 ‘인공지능의 이해’ 영역에 인공지능의 원리가 내용 요소로 포함되어 있다. 그러나 성취 기준 해설에 따르면, 학습자는 인공지능의 개념과 특성을 이해하고, 튜링 테스트를 활용해 인공지능의 지능적 판단에 대해 사고하며, 인공지능이 적용된 최신 사례를 분석하여 그 활용 범위와 중요성을 설명하는 데 중점을 둔다.

즉, 현행 교육과정에서는 인공지능 원리 학습이 일부 포함되어 있으나, 실제 수업에서는 다양한 활용 사례 탐구와의 파악이 주된 학습 활동으로 구성되는 경향이 있다.

Table 5. 2022 Revised Curriculum ‘Artificial Intelligence Basic’ Content System

Learning unit	Content Elements (Knowledge and Understanding Categories)
Understanding AI	<ul style="list-style-type: none"> <li>Principles of Artificial Intelligence</li> <li>Artificial Intelligence and Exploration</li> <li>the expression and reasoning of knowledge</li> </ul>
AI and Learning	<ul style="list-style-type: none"> <li>Machine Learning and Data</li> <li>machine learning algorithm</li> <li>Artificial neural networks and deep learning</li> </ul>
The Social Impact of AI	<ul style="list-style-type: none"> <li>Development of Artificial Intelligence and Social Change</li> <li>Artificial Intelligence and Careers</li> <li>Artificial Intelligence and Ethics</li> </ul>
AI project	<ul style="list-style-type: none"> <li>AI and Sustainable Development Goals</li> <li>Artificial Intelligence Problem Solving Procedures</li> </ul>

인공지능의 원리 이해를 돕고 인공지능 윤리에 대한 인식을 높일 수 있도록 설명 가능한 인공지능 교육 프로그램의 교육 내용은 Table 6과 같이 정리하였다.

Table 6. Explained Artificial Intelligence (XAI) Training Program Learning Content

Classification	Learning Content
Requirements for XAI	<ul style="list-style-type: none"> <li>Consider the position of users, operations managers, and developers</li> </ul>
The concept and necessity of XAI	<ul style="list-style-type: none"> <li>Artificial intelligence that can explain AI’s judgment in a form that can be understood by humans by presenting rational reasons and grounds</li> </ul>
Technology for XAI	<ul style="list-style-type: none"> <li>Transparent Model, Post-Description XAI</li> </ul>
The principle of AI using transparent models	<ul style="list-style-type: none"> <li>Centering on k-NN</li> </ul>
Post-Description XAI Methods and Cases	<ul style="list-style-type: none"> <li>LIME, Deconvolution, CAM, Grad-CAM Case Introduction</li> </ul>
Responsible AI	<ul style="list-style-type: none"> <li>The process of developing and distributing AI technology is ethical, reliable, and respects users’ personal information and social values</li> </ul>

#### 4.2 Design

분석 결과를 바탕으로, 본 연구에서는 총 4차시로 구성된 설명 가능한 인공지능 교육 프로그램을 설계하였다. 교수·학습 모형은 CT 요소 중심 모델 중 하나인 DPAAP 모델을 적용하였으며, 이를 통해 문제를 세분화하고, 반복되는 규칙과 특성을 식별하며, 본질적인 개념으로 단순화하는 과정을 거친 뒤 절차적 해결 방안을 설계하고 프로그래밍으로 구현하는 전 과정을 포함하도록 구성하였다. 특히 문제 분해와 패턴 인식 단계에서는 언플러그드 활동을 도입하여 학습자가 컴퓨터 없이도 인공지능의 작동 원리와 결론 도출 과정을 직관적으로 이해할 수 있도록 하였다.

설명 가능한 인공지능 교육 프로그램의 전체 구조는 Table 7에 제시하였으며, 수업의 흐름과 운영 방식은 Fig. 3과 같이 도식화하였다. 각 차시는 핵심 개념의 이해와 실습 활동 간의 균형을 유지하도록 설계되었으며, 복잡한 알고리즘 구현과 프로그래밍 실습은 학습 집중도를 고려하여 2~3차시에 배치하였다.

마지막 4차시에서는 이전 차시에서 다룬 학습 내용을 종합·정리하고, 인공지능 윤리와 사회적 함의에 관한 토론 활동을 통해 학습자의 인지적 부담을 완화하는 동시에 비판적 사고력과 책임 있는 활용 태도를 함양하도록 하였다.

이러한 차시 구성은 학습자의 개념적 이해, 실습을 통한 절차적 지식 습득, 그리고 태도 및 가치관 형성을 균형 있게 지원하기 위한 의도적 설계에 해당한다.

Table 7. Explainable Artificial Intelligence (XAI) Training Program Overview

Subject	<ul style="list-style-type: none"> <li>Explainable Artificial Intelligence (XAI)</li> </ul>
Classes Intention of the plan	<ul style="list-style-type: none"> <li>It enhances understanding of artificial intelligence by deviating from classes centered on using artificial intelligence and examining how artificial intelligence works.</li> <li>Understanding XAI (explainable artificial intelligence) as a reliable artificial intelligence and having a positive perspective on artificial intelligence have an attitude to actively prepare for the artificial intelligence society.</li> </ul>
Class flow	<ul style="list-style-type: none"> <li>Pre-Action: The 8th session of the basic understanding process of artificial intelligence before learning about XAI</li> <li>1st lesson: the need for explainable artificial intelligence (XAI), the concepts and types of explainable artificial intelligence, and how to explain artificial intelligence</li> <li>2~3rd session: Explaining artificial intelligence using transparent models, overview of k-NN algorithms, understanding how they work (unplugged activities), writing and using k-NN algorithms as programs (understanding algorithms, programming)</li> <li>4th session: Introduce how to add descriptions and examples of complex AI models, and learn about proper AI awareness and development directions</li> </ul>



Fig. 3. Class flow and teaching style

4.3 Development

4차시의 설명 가능한 인공지능 교육 프로그램의 초안을 작성하고 교수학습 지도안과 활동지를 개발, 제작하였다. 제작한 자료와 설명 가능한 인공지능 교육 프로그램의 초안을 컴퓨터교육 관련 전공자인 현직 교사 15인으로 구성된 전문가 집단의 검토를 받아 타당성을 검증받고, 피드백 내용을 기반으로 수정 보완하였다.

1차 전문가 집단 검토에서 전문가 15인은 설문을 통해 설명 가능한 인공지능 교육 프로그램을 검증했다. 검증 문항은 Likert 5점 척도를 활용한 문항과 개방형 응답 문항을 활용하였으며, 인공지능 리터러시 교육의 필요성, 교육 내용의 적절성, 수업 차시 및 평가도구, 교수학습 모형의 적절성, 내용 구성의 적절성, 학생 활동지에 대한 적절성에 대해 응

답했다. 검증의 효과를 높이기 위해 설명 가능한 인공지능 교육 프로그램의 초안 및 교수학습 지도안과 활동지, 사전-사후 인공지능 리터러시 검사 도구를 제공하였다.

4.4 Implementation

실험 집단에 대해 8차시의 인공지능 교육을 먼저 실시하였다. 사전학습은 인공지능의 개념과 발전 과정, 인공지능의 종류와 특징 및 인공지능의 학습, 기계학습의 종류와 특징, 인공지능 분류 모델을 활용한 인공지능시스템 실습을 포함하는 내용이다. 구성된 사전 교육에 포함된 인공지능시스템 실습은 인공지능 모델을 활용 중심으로 사용하는 형태이며, 모델을 사용하는 실습을 하지만 라이브러리를 통해 제공되는 모듈을 사용하지만, 내부의 동작 과정에 대해서는 이론과 설명을 통한 수업을 넘어서기 어렵다. 해당 사전 학습을 실시한 후에 인공지능 리터러시 검사 도구를 사용하여 사전 검사를 실시하였다.

4.4.1 Class Program for 1st and 2nd sessions

1차시에는 인공지능 면접과 관련한 기사를 제공하고 인공지능의 결론 도출 과정에 대한 설명이 필요함을 인공지능 시스템을 사용하는 사람과 인공지능 시스템의 도입 및 운영 관리하는 사람, 인공지능 시스템을 개발하는 사람의 입장에서 생각해보고 학습지를 작성하도록 하였다. 이를 통해 설명 가능한 인공지능의 필요성에 대해 알아본 후, 설명 가능한 인공지능의 개념과 역할, 인공지능 시스템을 설명하는 방법과 기술에 대하여 소개하였다.

2차시에는 k-NN의 동작 원리를 알아보는 언플러그드 활동을 진행하였다. 실생활에서 쉽게 접할 수 있는 예시 데이터를 사용하여 새로운 입력 데이터를 k-NN으로 유사한 그룹으로 분류하고 그 결과에 따라 메뉴를 추천하도록 하였다. 데이터를 벡터로 나타내고 좌표 평면에 표시하여 유클리드 거리로 유사도를 측정하는 활동으로 구성하였다.

2차시의 언플러그드 활동을 통해 k-NN의 동작 원리를 순서대로 따라가며 인공지능 시스템이 결론을 도출하는 과정을 절차에 따라 진행해볼 수 있도록 하였으며, 이런 단계의 흐름을 바탕으로 이어지는 알고리즘 설계에서 그대로 사용하도록 하였다.

이 활동에서 인공지능 시스템이 결론을 도출한 과정, 다시 말해 해당 메뉴를 왜 추천했는지를 활동지에 작성하면서 설명하도록 하였고, 이를 통해 k-NN의 동작 원리를 다시 한번 이해하고 정리할 수 있도록 하였다.

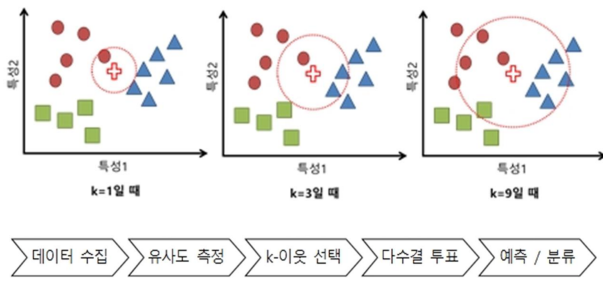


Fig. 4. k-NN classification algorithm

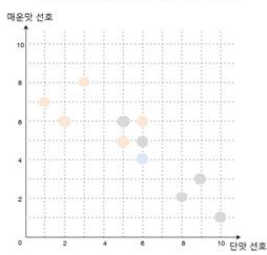
1) 데이터 수집 : 친구들은 단맛 선호(0-10), 매운맛 선호(0-10) 정도에 따라 짜장파와 짬뽕파 나뉜다.

짜장파	친구 ㉠	친구 ㉡	친구 ㉢	친구 ㉣	친구 ㉤
	(8, 2)	(9, 3)	(10, 1)	(6, 5)	(5, 6)
짬뽕파	친구 ㉥	친구 ㉦	친구 ㉧	친구 ㉨	친구 ㉩
	(1, 7)	(2, 6)	(3, 8)	(5, 5)	(6, 6)

2) 유사도 측정

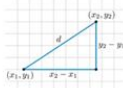
- i. 맛 선호가 비슷한 정보를 쉽게 파악하기 위해 좌표 평면에 친구들을 색깔 점으로 표시한다. (짜장파 - 검정 / 짬뽕파 - 빨강 / 길동이가 - 파랑)
- ii. 길동이와 각 친구들의 위치에 대해 좌표평면 상의 거리를 구한다.

< 단맛/매운맛 선호에 따른 메뉴 분포 >



친구	그룹	거리
친구 ㉠	짜장파	$\sqrt{16+9}$
친구 ㉡	짜장파	$\sqrt{25+4}$
친구 ㉢	짜장파	$\sqrt{36+16}$
친구 ㉣	짜장파	2
친구 ㉤	짜장파	$\sqrt{1+1}$
친구 ㉥	짬뽕파	$\sqrt{9+1}$
친구 ㉦	짬뽕파	$\sqrt{4+1}$
친구 ㉧	짬뽕파	$\sqrt{1+9}$
친구 ㉨	짬뽕파	1
친구 ㉩	짬뽕파	$\sqrt{4+1}$

※ 유사도는 유클리드 거리(Euclidean distance)를 사용하여 구한다.



$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

(피타고라스의 정리 이용)

Fig. 5. k-NN Algorithm Unplugged

#### 4.4.2 Class Program for 3rd and 4th sessions

3차시에는 전시에 정리한 k-NN의 단계별 동작을 순서도로 표현하면서 구체적인 처리 절차를 정리하고 프로그램을 작성하기 위한 기초 작업을 하도록 했다. 데이터 수집 단계는 주어진 샘플 데이터를 자료구조로 정리하여 표현하는 과정으로 하였고, 유사도 측정 단계에서는 리스트에 입력 데이터와 학습 데이터 사이의 유클리드 거리를 계산하여 저장하도록 하였다. 3차시 활동 마지막에는 작성한 프로그램을 이용하여 k-NN의 k값을 변경하면서 추천 메뉴가 달라지는 것을 확인하였다. 이를 통해 데이터, 알고리즘, 하이퍼파라미터의 설정에 따라 인공지능 시스템이 다른 결론을 도출할 수 있음을 알 수 있도록 하였다. 더 나아가 이러한 실습을 통해 학습자는 인공지능 모델이 결정하는 과정의 민감성과 결과 해석의 중요성을 직접 체감하게 된다.

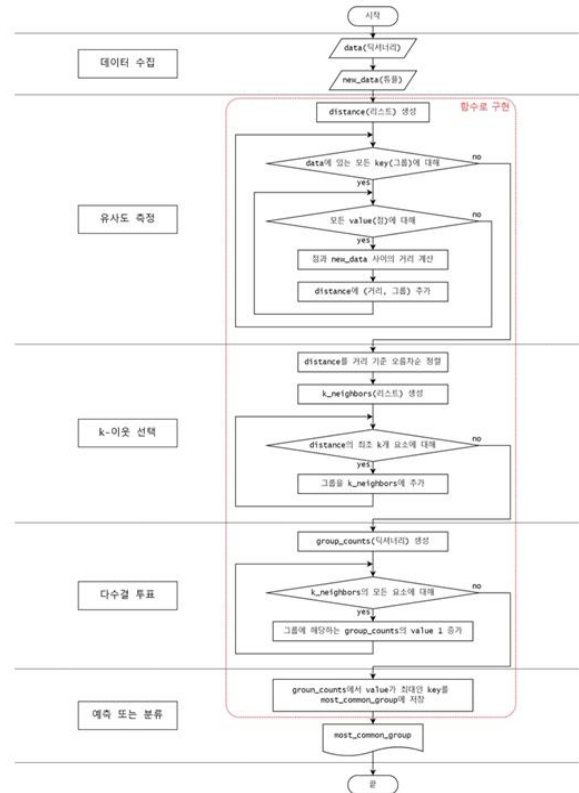


Fig. 6. Algorithm Design

4차시에서는 사후 설명 방식의 설명 가능한 인공지능 기법인 LIME, CAM, Grad-CAM, Deconvolution을 소개하고, 각 기법이 인공지능의 결론을 어떻게 설명하는지 설명하였다. 복잡한 내부 처리 과정을 가진 블랙박스 모델의 특성상 자세한 설명은 어려웠으나, 기존 결과를 시각화한 간단한 예제를 통해 설명 방식을 제시하였다.

마지막으로, 인공지능의 높은 성능뿐 아니라 윤리적 책임 요구가 커지고 있음을 언급하며, 이에 대응하기 위한 책임 있는 인공지능(Responsible AI) 개념과 국가 및 기업 차원의 윤리 가이드라인 마련 노력이 지속되고 있음을 소개하였다.

#### 4.5 Evaluation

4차시에 걸친 실험이 종료된 후, 사후 검사를 실시하여 설명 가능한 인공지능 교육 프로그램이 인공지능 리터러시에 미치는 영향을 평가하고 개발된 프로그램의 효과성을 확인하였다. 사후 검사는 사전에 사용한 동일한 인공지능 리터러시 평가도구를 활용하였으며, 사전-사후 검사 결과를 대응표본 t-검정을 통해 분석하였다. 또한 연구의 효과를 정성적으로 파악하기 위해 학생들이 작성한 설문 내용을 참고하였다.

## IV. Results

### 1. Quantitative analysis

검사 도구에서는 인공지능 리터러시를 인공지능의 사회적 영향, 인공지능 실행 계획, 인공지능 문제해결, 인공지능의 이해, 데이터 리터러시, 인공지능 윤리의 6개 하위 요소로 분류하고 있다. 이에 검사 도구의 30문항에 대해서 하위요소별로 각각 검증을 진행했다. 분석을 위한 결과는 Table 8과 같다.

Table 8. Results of paired-sample t-test for awareness of love for country

$p < .1$ ,  $p^{**} < .05$ ,  $p^{***} < .001$

Subelements	Average		t	P
	pre	post		
AI Social Impact	30.25	32.55	-2.544	.008**
AI Action Plan	9.6	11.2	-2.338	.012**
Solving AI Problems	14.05	15.09	-2.152	.019**
Understanding AI	14.95	18.4	-3.834	.000***
Data Literacy	10.35	11.75	-1.642	.054*
AI ethics	6.35	7.15	-1.904	.032**
The entire (AI Literacy)	85.55	96.95	-4.180	.000***

‘인공지능의 사회적 영향’ 요소에서 t값이 -2.544로 사전 검사가 사후 검사보다 낮으며, P값이 .008로 <.05 유이 수준으로 설명 가능한 인공지능 교육 프로그램이 인공지능 리터러시의 ‘인공지능의 사회적 영향’ 하위 요소의 점수 증가에 유의미한 수준으로 긍정적인 영향을 미친 것으로 해석할 수 있다. ‘인공지능 실행 계획’ 요소에 대해서도 <.05 유의수준으로 교육 프로그램의 적용이 해당 요소의 역량 증가에 긍정적인 영향을 미친 것으로 해석할 수 있다. 같은 방식으로 다른 모든 하위 요소에 대해서 설명 가능한 인공지능 프로그램이 학생들의 인공지능 리터러시의 각 요소 역량 강화에 긍정적인 영향을 미친 것으로 나타났다. 특히 ‘인공지능의 이해’ 하위 요소에 대해서는 <.001의 매우 높은 수준으로 프로그램 적용에 의한 학생들의 리터러시에 변화가 있었음을 알 수 있다.

그러나 설명 가능한 인공지능 교육 프로그램에서 k-NN의 동작을 알아보는 언플러그드 활동과 프로그래밍 활동에서 데이터는 10개의 샘플데이터로만 제공하며, 별도의 처리 과정이 포함되어 있지 않아 전처리나 변환, 출력 등에 대한 학습이 이루어질 기회가 없었다. 이 때문에 인공지능 리터러시 검사도구 중 ‘데이터 리터러시’ 하위 요소의 경우 사전/사후 검사의 결과에서 큰 차이가 나타나지 않은 것으로 해석할 수 있다.

‘AI 윤리’ 하위 요소의 경우에도 역시 사전/사후 검사 결과의 차이가 상대적으로 크지 않은 것으로 나타났다. 이것은 설명 가능한 인공지능 교육 프로그램을 적용하기 전에 사전 학습으로 진행된 인공지능 수업의 영향으로 판단된다. 사전 검사 전에 진행한 8차시의 수업에 이미 인공지능 윤리에 대한 내용이 포함되어 있어서 개발한 4차시 프로그램에 의한 인공지능 윤리의 변화가 크게 나타날 수 없었던 것으로 볼 수 있다. 그럼에도 불구하고 인공지능을 사회적으로 올바르게 활용하기 위해 인공지능의 결과를 무비판적으로 수용하지 않고 신뢰할 수준을 정하는 등의 인식을 더욱 키울 수 있어서 설명 가능한 인공지능 교육 프로그램은 AI 윤리에도 도움이 된다고 해석할 수 있다.

하위 요소뿐만 아니라 검사 전체에 대해서도 설명 가능한 인공지능 교육 프로그램이 인공지능 리터러시를 향상시키는데 통계적으로 높은 유의미 수준으로 긍정적인 영향을 미치고 있음을 사전/사후 검사 결과 분석을 통해 확인하였다.

### 2. Qualitative analysis

수업을 마친 후, ‘인공지능 기초’ 수업 전반에 대한 평가와 ‘설명 가능한 인공지능(XAI)’ 부분의 수업에 대한 평가를 설문으로 진행하였다. 수업에 대한 학습자 인식 변화 설문지의 범주는 Table 9과 같다.

“인공지능 기초 수업 전반에서 의미있거나 인상 깊었던 점을 적어주세요.”라는 문항에 대해 직접 코드를 작성하면서 이해하는 과정이 좋았다는 응답이 있었고, 인공지능의 작동 원리를 이해하는 과정이 있어서 의미있었다는 응답 등이 있었다.

“인공지능 기초 수업 전반에서 어려웠던 점을 적어주세요.”라는 문항에 대해서는 코딩이 어려웠고, 여러 인공지능 알고리즘에서 어려운 부분이 있었다는 응답이 있었다.

“설명 가능한 인공지능(XAI) 수업에서 재미있거나 흥미로웠던 점을 적어주세요.”라는 문항에 대해서는 다양한 응답이 제출되었다. 생소했던 설명 가능한 인공지능에 대해 신기하게 받아들였다는 것을 알 수 있다.

“설명 가능한 인공지능(XAI) 수업에서 어려웠던 점, 힘들었던 점을 적어주세요.”라는 문항에 대해 프로그래밍 과정이 어려웠다는 응답이 있었고, 간단하게 소개만 했던 사후 설명 인공지능 알고리즘이 어려웠다는 학생도 있었다.

“설명 가능한 인공지능(XAI) 수업에서 가장 기억에 남는 것을 적어주세요.”라는 문항에 대해 인공지능 알고리즘의 동작을 직접 프로그래밍까지 하면서 실습한 내용이 기억에 남는다는 응답이 있었고, 인공지능 알고리즘이 생각

보다 간단하고 이해할 수 있는 원리로 결론을 도출하기도 한다는 것이 신기하다는 응답도 있었다.

분석 결과, 학습자들은 설명 가능한 인공지능(XAI) 수업을 통해 AI 작동 원리와 알고리즘 구조에 대한 이해가 심화되었고, 프로그래밍 실습을 통해 학습 흥미와 동기가 강화되었음을 알 수 있었다. 또한 일부 학습자들은 AI의 사회적 영향과 책임 있는 활용의 중요성에 대해 성찰하며 긍정적인 태도 변화를 보였다. 이는 설명 가능한 인공지능(XAI) 기반 교육이 학습자의 인지적 성취뿐만 아니라 태도와 가치관 형성에도 긍정적인 영향을 미칠 수 있음을 알 수 있다.

Table 9. Category of change in learner perception of explainable artificial intelligence (XAI) classes

Top Category	Subcategories
a deepening of understanding	Understanding AI Operation Principles
	Understanding Algorithm Structures
interest and motivation in learning	Fun of hands-on and programming activities
	Awareness of real-life applicability
Change of attitude and perspective	Awareness of AI's Positive Utilization
	a reflection on social and ethical influences

3. Interpret inspection results

지금까지 살펴본 사전 검사와 사후 검사 결과 검증으로 다음과 같이 설명 가능한 인공지능 교육 프로그램을 평가할 수 있다.

첫째, 설명 가능한 인공지능 교육 프로그램은 고등학생의 인공지능 리터러시를 향상시킨다.

둘째, 인공지능 리터러시의 하위 요소 중 특히 ‘인공지능의 이해’ 영역에 대해 학습자 역량 향상에서 큰 성과를 보인다. 설명 가능한 인공지능 교육 프로그램을 적용했을 때 라이브러리 형태로 제공되는 인공지능 모듈을 사용하여 단순히 모듈을 학습시키고 활용하는 교육 프로그램에 비해 인공지능의 작동 원리와 인공지능이 결과를 도출하는 과정에 대하여 효과적으로 학습하여 인공지능 리터러시를 제고할 수 있다는 점을 알 수 있다. 더 나아가 이러한 학습 경험은 학습자가 향후 다양한 인공지능 기술을 접할 때 결과를 비판적으로 해석하고, 책임 있는 활용 방안을 모색하는 데 기초가 된다.

V. Conclusion

본 연구의 목적은 사회의 변화에 따라 인공지능 교육의 중요성이 커지고 정보 교과에서 인공지능 원리 교육이 이루어져야 한다는 요구에 따라 고등학생에게 적합한 수업 모델을 개발하여 인공지능 리터러시를 제고하는 것이다. 특히 CT 요소 중심 모델을 기반으로 하는 설명 가능한 인공지능 교육 프로그램이 기존의 인공지능 모델 활용 중심의 인공지능 교육에 비해 효과적으로 고등학생의 인공지능 리터러시 향상에 기여하는지 확인하고자 하였다. 본 연구의 결론은 다음과 같다.

첫째, 설명 가능한 인공지능 교육 프로그램을 기존의 인공지능 활용 중심 수업에 추가했을 때 고등학생의 인공지능 리터러시가 더욱 향상되는 것으로 나타났다. 이는 설명 가능한 인공지능 교육 프로그램이 단순히 미리 만들어져 있는 인공지능 모듈을 활용만 하는 것에 그치지 않고 알고리즘 설계의 과정을 거쳐 직접 프로그래밍을 통해 구현해보는 활동을 통해 인공지능의 원리를 이해하는 데에 도움을 주었기 때문이다.

둘째, 본 연구를 통해 개발된 고등학생 대상의 설명 가능한 인공지능 교육 프로그램은 4차시로 구성된 교육 프로그램으로, 언플러그드 활동을 포함하고 있고 CT 요소 중심 모델을 기반으로 하여 텍스트 프로그래밍을 통한 인공지능 구현을 포함하고 있다. 이는 주로 언플러그드 활동에 그치고 있는 초등학교나 중학교의 인공지능 원리 이해 수업과 차별화되는 고등학생을 위한 수업 모델을 제시하였고 그 효과를 확인하였다.

셋째, 설명 가능한 인공지능 교육 프로그램은 인공지능 리터러시를 전반적으로 향상시키는 데에 효과적이었다. 그러나 학습자별로 개념과 용어를 어려워하거나 텍스트 프로그래밍에 부담을 가지는 경향이 있었다. 이는 교육 프로그램에서 인공지능과 관련된 개념이나 용어에 대한 추가적인 학습 과정을 제공하거나 협력학습 등 방법의 보완이 필요함을 시사한다.

본 연구를 통해 고등학생을 대상으로 하는 인공지능 교육에 있어 언플러그드 활동과 DPAAAP 모델을 활용한 설명 가능한 인공지능(XAI) 교육 프로그램을 개발하였고, 이를 이용한 인공지능 원리 이해 수업이 인공지능 리터러시 향상에 효과적일 수 있음을 확인하였다. 이 교육 프로그램은 인공지능 리터러시를 향상시키는 데에 효과적인 새로운 교육 모델을 제시하는 기초 자료가 될 수 있을 것이다.

그리고 본 연구에는 몇 가지 제한점이 존재하며, 이를 고려하여 후속 연구를 다음과 같이 제안하고자 한다.

첫째, 본 연구는 설명 가능한 인공지능 중 투명한 모델에 포함되는 k-NN 알고리즘을 중심으로 인공지능과 인공지능의 동작 원리에 대한 이해를 돕는 활동을 기반으로 한다. 추후 다른 기계학습 알고리즘에 대해서도 확대 적용한다면 교육 효과를 높일 수 있을 것이다.

둘째, 1학년 때 정보 과목을 통해 프로그래밍에 대한 이해와 경험을 가진 학생 20명을 대상으로 진행한 소규모 연구이므로 대부분의 공교육에 바로 적용하거나 일반화하는 데에 한계가 있다. 다양한 대상에 대해 적용할 수 있도록 시수나 교육 내용을 유연하게 조정하고, 연구대상의 규모를 키워서 추가 연구를 진행해 볼 수 있다.

셋째, 본 연구에서는 언플러그드 활동을 추가한 DPAAP CT요소 중심모델을 바탕으로 수업을 진행하고 텍스트 프로그래밍을 통한 구현까지는 포함하고 있다. 블록 기반 언어 등 다른 프로그래밍 언어를 사용하는 활동으로 적용을 확대하여 다른 학교급에서도 인공지능 원리 교육에 활용할 수 있도록 후속 연구를 진행할 수 있다.

## REFERENCES

- [1] Ministry of Education, "Ministry of Education Notice No. 2022-33 (2022.12.22) General explanation of the 2022 revised curriculum (high school)"
- [2] Ministry of Education, "[Press release during briefing of Ministry of Education 11-24 (Wednesday)] Announcement of the main draft of the 2022 revised curriculum"
- [3] Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Boston, MA: The Center for Curriculum Redesign.
- [4] Davy Tsz Kit Ng, Min Lee, Roy Jun Yi Tan, Xiao Hu, J. Stephen Downie, Samuel Kai Wah Chu. (2022). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*.
- [5] Da Bin Park et al. (2021). "Studying the Educational Meaning of Explainable Artificial Intelligence for Elementary Artificial Intelligence Education." *Journal of the Society of Information Education* 25 (5), 803-812.
- [6] Lee Da-gyeom, Kim Sung-won, and Lee Young-jun (2021). Analysis of research trends in artificial intelligence literacy education. *Korean Computer Education Association Academic Presentation Conference Paper Collection*, 25 (2), 25-27
- [7] Lee Yu-mi. (2021). A Study on the Characteristics of Literacy in the Age of AI: Focusing on AI Literacy and Relationship Literacy. *Language Studies* 110, 281-302.
- [8] Hyunjung Hwang. (2022). A Study on the derivation of AI literacy compositional concepts: focusing on the conceptualization of detailed competencies by sub-dimension. PhD thesis. Graduate School of Konkuk University.
- [9] Kim Jin-sook et al. (2015). "[Consignment Research CR 2015-35] A Study on the Development of SW Education Teaching and Learning Models". Korea Education Development Institute & Korea Education and Academic Information Service.
- [10] Kim Sung Won et al. (2022). "Artificial Intelligence Literacy Test Tool of Middle School Students". *Journal of the Korean Computer Information Society* 27(3), 225-238.
- [11] Kim Soo-hwan et al. (2020). A review of artificial intelligence education for K-12 students and teachers. *Journal of the Society of Computer Education*, 23 (4), 1-11.

## Authors



Eun-Sook Jung majored in Artificial Intelligence convergence education and received a master's degree from the Graduate School of Education at Korea National University of Education.

She currently teaches students in high school and is interested in Artificial Intelligence convergence education and computer education.



Han-Bin Lee received his master's degree from the Graduate School of Education at Korea National University of Education (KNUE), South Korea. He is currently pursuing his doctoral degree in the

Department of Elementary Computer Education at Korea National University of Education (KNUE), South Korea. he is interested in AI convergence education, computer education, and information education curriculum.



Kwi-hoon Kim received the B.S, M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology(KAIST), Daejeon, South Korea in 1998, 2000 and 2019, respectively.

Kwi-hoon Kim is currently a professor in the Department of Artificial Intelligence Convergence Education, Korea National University of Education (KNUE), South Korea. He worked in LG DACOM 2000~2005. From 2005 to 2020, he was a Principle Researcher with Electronics and Telecommunications Research Institute (ETRI). He is interested in AI convergence education, intelligent edge computing, reinforcement learning and knowledge-converged intelligent service.