Journal of The Korean Association of Information Education Vol. 25, No. 5, October 2021, pp. 803-812

http://dx.doi.org/10.14352/jkaie.2021.25.5.803

초등 인공지능 교육을 위한 설명 가능한 인공지능의 교육적 의미 연구

박다빈 · 신승기
서울교육대학교 교육전문대학원 인공지능교육전공

요약

본 연구는 문헌 연구 통해 설명 가능한 인공지능의 개념과 문제해결과정을 탐구하였다. 본 연구를 통하여 설명 가능한 인공지능의 교육적 의미와 적용 방안을 제시하였다. 설명 가능한 인공지능 교육이란 인간과 관련된 인공지능 문제를 다루는 사람 중심의 인공지능 교육으로 학생들은 문제 해결 능력을 함양할 수 있다. 그리고, 알고리즘 교육을 통해 인공지능의 원리를 이해하고 실생활 문제 상황과 관련된 인공지능 모델을 설명하며 인공지능의 활용 분야까지 확장할 수 있다. 이러한 설명 가능한 인공지능 교육이 초등학교에서 적용되기 위해서는 실제 삶과 관련된 예를 사용해야 하며 알고리즘 자체가 해석력을 지닌 것을 활용하는 것이 좋다. 또한, 이해가 설명으로 나아가기 위해 다양한 교수학습방법 및 도구를 활용해야 한다. 2022년 개정 교육과정에서 인공지능 도입을 앞두고 본 연구가 실제 수업을 위한 기반으로써 의미 있게 활용되기를 바란다.

키워드 : 인공지능 교육, 설명 가능한 인공지능, 인공지능 알고리즘 교육, 인공지능 모델, 문제해결력

A Study on the Educational Meaning of eXplainable Artificial Intelligence for Elementary Artificial Intelligence Education

Dabin Park · Seungki Shin
Artificial Intelligence Education, Graduate School of Education,
Seoul National University of Education

Abstract

This study explored the concept of artificial intelligence and the problem-solving process that can be explained through literature research. Through this study, the educational meaning and application plan of artificial intelligence that can be explained were presented. XAI education is a human-centered artificial intelligence education that deals with human-related artificial intelligence problems, and students can cultivate problem-solving skills. In addition, through algorithmic education, it is possible to understand the principles of artificial intelligence, explain artificial intelligence models related to real-life problem situations, and expand to the field of application of artificial intelligence. In order for such XAI education to be applied in elementary schools, examples related to real world must be used, and it is recommended to utilize those that the algorithm itself has interpretability. In addition, various teaching and learning methods and tools should be used for understanding to move toward explanation. Ahead of the introduction of artificial intelligence in the revised curriculum in 2022, we hope that this study will be meaningfully used as the basis for actual classes.

Keywords : Artificial Intelligence education, eXplainable Artificial Intelligence, Artificial Intelligence algorithm education, artificial intelligence model, human-centered artificial intelligence

교신저자 : 신승기(서울교육대학교 컴퓨터교육과)
논문투고 : 2021-10-07
논문심사 : 2021-10-13
심사완료 : 2021-10-14

1. 서론

대통령 직속 4차산업혁명위원회의 '인공지능의 대중화'를 위한 대국민 인식조사 결과 많은 국민들이 인공지능(Artificial Intelligence, AI)시대의 도래를 공감하고 인공지능 대중화를 요구하고 있었다. 그 중, 인공지능의 서비스 경험과 난이도, 인지도 측면에서는 긍정적인 답변을 하였으나 교육경험, 활용수준, 신뢰도 부분은 향상이 필요하다고 답하였다[20]. 인간의 삶에서 인공지능의 사용이 증가함에 따라 인공지능 교육에 대한 요구와 더불어 인공지능의 신뢰도 향상에 대한 윤리적인 요구 또한 증가하고 있는 대중의 인식을 반영한다. 유럽에서는 인공지능의 의사결정에 관한 신뢰도를 높이기 위하여 설명 가능한 인공지능(eXplainable AI: XAI)에 관해 지속적으로 강조하고 있다. 이는, 유럽인들의 삶에 영향을 미치는 누군가의 의사결정에 대하여 설명해야 한다는 유럽일반정보보호규정(General Data Protection Regulation)에 부합하는 개념이다[21]. 미국의 경우에는 인공지능 기술이 도입된 주택대출, 신용카드 발급 등의 금융 분야에서 인공지능의 의사결정에 이유를 제공하도록 법적으로 의무화하고 있다[22]. 인공지능은 점점 인간의 삶에서 중요한 부분에서 의사결정을 내리고 있으며 그 결정에 대한 이해 및 신뢰도를 높이기 위해서는 사용자에게 인공지능 의사결정의 이해를 알기 쉽게 설명해주고 의사결정을 보장해주는 법적인 제도와 윤리적인 체계가 요구된다.

교육부는 인공지능시대 교육정책방향과 핵심과제(2020)에서 인공지능교육 영역으로서 총 4가지(프로그래밍, AI 기초원리, AI 활용, AI 윤리)를 제시하였다. 그리고, AI 기초원리 영역에 속하는 인공지능 알고리즘과 AI 알고리즘 편향과 사회문제를 예로 제시하며 각 내용 간의 연계성을 강조하였다[1][24].

본 연구에서는 인공지능의 대중화에 대한 국민들의 인식, 세계 강국들의 인공지능 관련 신뢰성을 높이기 위한 연구 및 제도화, 교육부의 인공지능 교육 방향을 바탕으로 하여 설명 가능한 인공지능의 교육적 도입이 필요하다고 판단하였다. 설명 가능한 인공지능에 대하여 여러 분야에서 연구되어지고 있으나 명확한 정의와 방법, 절차적인 과정의 성립이 부족하다[2]. 이와 더불어, 교육계에서 설명 가능한 인공지능의 연구는 거의 찾아보기 어렵다. 따라서, 본 연구에서는 다양한 분야에서 연구된 설명 가능한 인공지능의 정의와 문제해결과정을 탐구하며 그 속에 담긴 교육적 의미를 찾고자 하였다.

2. 이론적 배경

2.1. 초등 인공지능 교육

인공지능의 기계학습에서 모델이란 데이터로 학습이 완료된 기계학습 알고리즘이라고 할 수 있다. 즉, 인공지능 모델은 실제 데이터와 인공지능 알고리즘의 합이라고 볼 수 있다[18]. 따라서, 인공지능 모델을 이해하고 설명하기 위해서는 인공지능 알고리즘 교육이 선행되어야 한다.

교육부(2020)에서 제시한 인공지능 교육 요소는 AI 기초원리, AI 활용, AI 윤리 세 가지 요소이며 각 내용 간의 연계될 수 있도록 언급하였다[1]. 구체적인 내용은 <Table 1>과 같다.

<Table 1> The role of AI education elements and contents[1]

Element	Content
Programming	· Recognizing and solving problems · Computational thinking through objective language use, conceptualization, logical order etc
AI basic principles	· Communication and collaboration education · Understand my memories, learning, reasoning through AI principle education
Using AI	
AI ethics	· Critical thinking through AI prejudice, error education · Practicing choosing moral standards

유정수 외(2020)는 국내외 인공지능 교육과 관련된 문헌 조사를 통해 핵심 내용 요소를 찾고, 2015 개정교육과정의 내용과 수준을 고려하여 한국형 인공지능 내용요소를 제시하였다. 이 중, <Table 2>의 초등 인공지능 교육 내용요소는 국가교육과정 교육목표에 따른 저학년, 고학년의 수준별 과목 및 내용요소이다.

<Table 2> The contents of AI education in elementary school[17]

Element	Elementary school grade		
	1~2	3~4	5~6
Understanding AI	· AI story	· Narrow AI · General AI	· AI vs human · Moravec's paradox · Turing test
AI & Data	· Various forms of data	· Guess numbers by hints	· Predicting with data · Make new data
AI algorithm	· Classifying · Finding something in common	· Reaction to the condition and situation	· Classifying things with decision trees
Application of AI	· Making AI robot	· Machine learning	· Making AI
Social impact	· Changing in daily life	· Common and difference of humans and AI	· The 4th industrial revolution · AI ethics

인공지능 교육 영역은 5가지로 구분되어있으며 주로 초등학교급에서 제시된 내용으로는 인공지능을 체험하거나 간단한 개념과 원리를 이해하는 내용으로 구성된다. 그러나, 각 내용 요소와 체계 간에 연계성이 부족하고 더 구체적이고 다양한 내용들이 제시되어야 한다. 또한, 인공지능의 알고리즘 영역에서 초등학생들이 학습할 수 있는 구체적인 알고리즘 범위와 종류들이 제시되어야 할 필요가 있다.

2.2. 설명 가능한 인공지능의 필요성과 교육에서의 적용

설명 가능한 인공지능이 인공지능 분야에서 필요한 이유를 탐구함으로써 교육에서도 설명 가능한 인공지능 교육이 필요한 이유를 생각해볼 수 있다.

A. Adadi, M. Berrada(2018)은 설명 가능한 인공지능의 필요한 이유를 크게 4가지로 제시하였다. 첫 번째 이유는 인공지능의 결정에 대해 정당하게 설명하기 위한 근거가 된다(Explain to justify). 특히, 예기치 못한 인공지능의 결정에 대한 정당한 설명을 제공한다. 두 번째 이유는 인공지능의 잘못된 것을 바로잡을 수 있다(Explain to control). 중요한 결정 상황에서 심각한 오류 등을 방지할 수 있다. 세 번째 이유는 인공지능을 지속적으로 개선하기 위함이다(Explain to improve). 많은 사용자들이 이해하고 설명할 수 있는 모델은 더욱 쉽게 개선될 수 있다. 마지막으로는, 새로운 지식을 얻을 수 있다(Explain to discover).누군가에게 설명을 요청하는 것은 지식을 얻기 위한 새로운 사실을 배우는 데 도움이 되는 수단이다. 설명 가능한 인공지능 모델은 다른 분야에서의 통찰을 제공해줄 수 있다[2]. Alejandro Barredo Arrieta et al(2020)은 인공지능과 관련된 대상에 따라서 왜 XAI가 필요한지 달라진다고 하였다. 예를 들어, 데이터 과학자, 개발자 등은 상품이나 제품의 효과성, 연구, 새로운 기능들을 확인하고 개선시키기 위해 XAI가 필요하다면 인공지능 모델의 의사결정에 의해 영향을 받는 사용자의 경우에는 그러한 결정이 공정한지 입증하기 위해서 설명 가능한 인공지능이 필요하다고 보았다[6].

국내외 설명 가능한 인공지능을 주제로 한 교육 사례를 살펴보면 국내에서는 교육부와 한국과학창의재단(2020)이 중학생을 위한 수업 자료를 배포하였다[23]. 본 교재에서 제시된 프로그램은 먼저 설명 가능한 인공지능의 개념과 필요성을 알아보았다. 그 후에 Machine learning 4 kids로 불쾌지수 판별 모델을 만들고, 의사결정트리를 기반으로 인공지능 모델의 판단 과정을 설명할 수 있도록 수업을 구성하였다. 국외에서는 Jose M. Alonso(2020)가 고등학생을 대상으로 하는 'XAI4TEENS'라는 XAI 교육 프로그램을 개발하였다. 농구 선수 데이터를 가지고 스크래치로 코딩하고 의사결정트리 분류기를 활용하여 선수 데이터를 분류한다. 그리고, 이것을 시각화하여 그래프로 결과를 나타내본다. 학생들은 신체적인 특성과 과거의 점수 데이터만을 활용하여 성별과 피부색에 관계없이 선수를 선발할 수 있는 설명 가능한 인공지능을 만들고 체험할 수 있다[21].

설명 가능한 인공지능이라는 개념은 교육에서 연구가 거의 진행되지 않았으며 초등학생을 대상으로 하는 국내외 연구도 찾아보기 어렵다. 따라서, 본 연구에서는 설명 가능한 인공지능이 교육에서 가지는 의미와 구체적인 교육 방안에 대하여 탐구하고자 한다.

3. 연구 방법

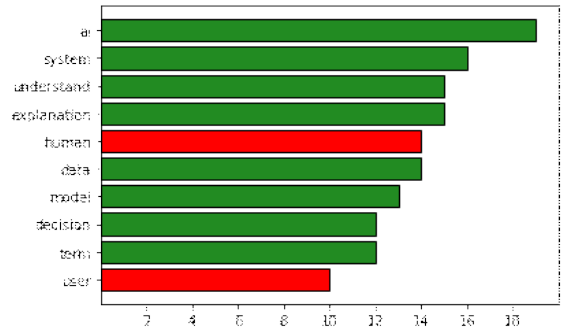
3.1. 빈도 분석을 통한 설명 가능한 인공지능의 개념 탐구

설명 가능한 인공지능(XAI)은 Van Lent et al(2004)에 의하여 처음 사용되어 그 의미는 여러 분야에서 다양한 의미로 해석하여 연구되고 있다. <Table 3>는 설명 가능한 인공지능의 의미 분석을 담은 해외 연구 사례들이다. 본 연구에서는 아래 해외 연구들에서 설명 가능한 인공지능의 의미를 담은 문장들을 추출하였다. 추출된 문단 속 문장들은 자연어 처리 및 분석, 텍스트 마이닝 등에 활용되는 NLTK(Natural Language Tool Kit) Python 패키지를 활용하여 품사를 분리하고 빈도 분석을 통해 가장 많이 활용된 핵심어를 찾고자 하였다.

<Table 3> Research of eXplainable AI

Researcher	Meaning of XAI
Arrieta, A. B. et al(2020)	*Given a certain audience, explainability refers to the details and reasons a model gives to make its functioning clear or easy to understand[6].*
Gunning, D., & Aha, D. (2019)	*Providing users with explanations that enable them to understand the system's overall strengths and weaknesses, convey an understanding of how it will behave in future or different situations, and perhaps permit users to correct the system's mistakes[4].*
Derek doran et al(2017)	*Combining their results with work investigating factors influencing the human comprehensibility of representation formats and reasoning approaches[5].*
Gunning, D. et al(2019)	*To make its behavior more intelligible to humans by providing explanations[9].*
Guidotti, R. et al(2018)	*Interpretability is defined as the ability to explain or to provide the meaning in understandable terms to a human[7].*
Dino Pedreschi et al(2019)	*Given a dataset of training decision records, how to develop a machine learning decision model together with its explanation. Given the decision records produced by an obscure black box decision model, how to reconstruct an explanation for it[8].*
Adadi, A., & Berrada, M. (2018)	*XAI is not a monolithic concept, it reflects several distinct related notions[2].*

(Fig. 1) The word cloud of XAI



(Fig. 2) The frequency of XAI keywords

총 1089개의 단어들을 분석하였으며 그 중 401개의 명사, 116개의 형용사, 172개의 동사를 추출하였다. 그리고 나서, 표제어 추출(Lemmatization)을 통해 각 품사의 원형을 찾고 불용어(Stopsords)를 제거하여 텍스트를 전처리하였다. 이 중, 가장 빈도가 높은 상위 20가지의 단어를 (Fig.1)과 같이 워드 클라우드로 표현하였으며, 상위 10가지의 단어를 (Fig.2)와 같이 막대그래프로 표현하였다.

(Fig.2)를 보면 Human은 14회, User은 10회가 사용되었다. 이 경우, User도 인공지능을 사용하는 사람이라는 맥락에서 볼 때, 설명 가능한 인공지능에서 결국 가장 중요한 단어는 '인공지능을 사용하는 사람'이라고 할 수 있다. AI라는 단어를 제외하고 system, understand, explanation이 다음으로 가장 많이 사용되는 단어였다. 분석한 단어의 빈도수를 바탕으로 설명 가능한 인공지능의 의미를 여러 문장으로 조합해보면,

· 설명가능한 인공지능(Explainable AI) 시스템(system)은 데이터(data)를 기반으로 결정(decision)한 것을 사람

(human)이 이해할 수 있는(understand) 모델(model)을 사용자(user)에게 설명해주는 용어(term)이다.

- 설명 가능한 인공지능(Explainable AI) 용어(term)는 인간(human)이 의사결정(decision)할 수 있도록 데이터(data)를 바탕으로 한 모델(model)을 사용자(user)가 이해할 수 있도록(understand) 설명해주는 시스템(system)이다.
- 설명 가능한 인공지능(Explainable AI)은 인공지능이 결정한(decision) 내용을 인간(human, user)이 이해할 수 있는 데이터(data)를 기반으로 한 모델(model)을 바탕으로 설명하는 것을 의미하는 시스템(system)이다.

위에서 조합한 문장들을 살펴볼 때, 설명 가능한 인공지능의 몇 가지 의미를 도출해볼 수 있다. 첫째, 설명 가능한 인공지능은 결국 사람이 중심이며 사람의 이해를 위한 시스템이라는 것이다. 둘째, 설명 가능한 인공지능은 인공지능의 의사결정 또는 인간의 의사결정과 관계가 있다. 셋째, 설명 가능한 인공지능은 사용자가 이해할 수 있는 모델이 중요하다는 것이다. 위의 세 가지 의미를 바탕으로 설명 가능한 인공지능의 문제해결과정을 탐구해 보고자 한다.

3.2. 설명가능한 인공지능의 문제해결과정

위의 연구 중 블랙박스의 인공지능의 문제를 설명 가능한 인공지능을 통해 해결하는 과정을 명확하게 제시한 몇 가지의 논문을 분석하였다. <Table4>는 문제 해결 과정을 정리한 표이다.

Samek, W *et al*(2017)은 인공지능 모델을 설명하는 두 가지 해석 기법을 사용하는데, SA(Sensitivity Analysis)기법(입력 변화에 대한 예측 결과의 변화량을 정량화하여 어떤 부분이 결과 도출에 큰 영향을 끼쳤는지 설명하는 방법), LRP(Layer-Wise Relevance Propagation)기법(딥러닝 모델에서 각 레이어별 기여도를 측정하는 방법)을 활용하여 모델을 설명한다. 이때, 이 모델을 이해시키기 위한 도구로서 히트맵(heat map)을 사용한다. 이러한 시각화 도구는 어떤 모델이 적절한지, 어떻게 인공지능이 구현되었는지를 설명하는 중요한 장치가 된다[3].

Gunning, D., & Aha, D.(2019)는 DARPA(미 국방성 연구기관)에서 XAI를 연구하는 11개팀에 대하여 분석하였다. 11개의 XAI 연구팀은 효과적인 설명을 통해 인간이 이해할 수 있는 AI 시스템을 만드는 것에 집중하였다. 각 팀은 인공지능으로 해결할 수 있는 문제 상황과 그에 맞는 적절한 설명할 수 있는 모델을 선정하였다. 이때, 인간의 이해와 모델을 연결할 수 있는 방법을 'Explanation interface'라는 하였으며 그것을 위한 기법으로 대부분의 연구팀에서 그래프, 그림 등으로 시각화(visualization)기법을 활용하였다[4]. 이 문제 해결 방식에 주목할 점은 인공지능 사용자의 심리적인 모델(Mental model)을 만들어 이해의 평가도구로 활용하였다는 점이다.

Derek Doran *et al*(2017)은 현재 대부분의 설명 가능한 인공지능의 문제들은 해석 가능하고 이해하기 쉬운 모델에 관한 연구가 대부분이라고 말한다. 하지만, 본 연구에

<Table 4> XAI Problem solving process[3][4][5]

Researcher	XAI Problem solving process
Samek, W et al(2017)	· AI decision · Choose Explain methods(models) : SA, LRP · Visualization : heatmap · Understand
Gunning, D., & Aha, D(2019)	· AI Recommendation, Decision or Action · Set up Explainable model · Explanation interface (Visualization, language understanding and generation etc) · User Decision · Measure of Explanation Effectivenss(Mental model)
Derek Doran et al(2017)]	· AI decision · Set up Comprehensible and Interpretable model · Formulating Reasoning engine · Combining symbol by a comprehensible machine and concepts by symbol

서는 그런 모델들은 설명을 가능하게 하지만, 설명 자체를 제시하지 않는 점이 한계라고 말한다. 따라서, 설명 가능한 인공지능을 구현하는 과정에서 가장 중요한 것은 명시적인 자동 추론 기능을 공식화하는 것을 강조한다. 즉, 기계에서 표현되는 기호와 그 기호로 설명되는 개념을 연결시키는 것이 추론 엔진이며 이것이 설명 가능한 인공지능의 핵심이라고 주장한다[5].

설명 가능한 인공지능의 일반적인 문제 해결 과정은 **'AI가 인간과 관련된 결정을 내리는 문제 상황 -> 설명 가능한 인공지능 모델 추출 및 이해 -> 모델 설명 -> 사용자 이해'**의 과정을 따른다. 위 과정에서는 문제 상황에 적절한 사용자가 이해할 수 있는 인공지능 모델이 핵심이라는 것을 알 수 있다. 이를 교육에서 활용하기 위해서는 학생 수준에서 이해할 수 있는 인공지능 알고리즘을 교육해야 한다. 또한, 알고리즘으로 구현된 모델과 사용자를 연결할 수 있는 전략은 다양하며 주로 문제 상황, 모델에 따라 전략이 달라진다.

3.3. 초등학교에서 활용할 수 있는 설명 가능한 인공지능 알고리즘

기계학습에서 모델은 데이터로 학습이 완료된 기계학습 알고리즘을 의미한다[18]. 초등학교에서 설명 가능한 인공지능의 모델을 설명하기 위해서는 학생 수준에서 이해할 수 있는 인공지능을 알고리즘을 선정하는 것이 중요하다. 기존에 진행된 인공지능 알고리즘 관련 연구에서 초등학생 학생들에게 유의미한 결과를 도출하였던 프로그램에 적용된 인공지능 알고리즘을 <Table 5>에 정리하였다.

초등 인공지능 알고리즘 교육 연구는 활발하게 이루어지고 있으며 인공지능 교육 프로그램에서 주로 사용된 인공지능 알고리즘은 위의 표와 같다.

Arrieta, A. B. et al(2020)은 설명 가능한 인공지능에 사용할 수 있는 인공지능 모델을 일정 분류기준에 따라 분류하고 그에 따른 설명 전략도 제시하였다[6]. <Table 6>는 Arrieta, A. B. et al(2020)의 연구에서 설명 가능한 인공지능 모델에 적합한 해석 기법을 적용하기 위한 두 가지 분류 기준이다.

<Table 5> AI algorithm in elementary school[10][11][12][13][14][15][16]

Researcher	Decision Tree	K-means	Linear Regression	K-Nearest Neighbors	Neural Network model
Jang(2019)	O		O		O
Ryu&Han (2019)					O
Jang(2020)	O				
Kim&Moon (2021)	O	O			
Lee&Moon (2021)		O			
Choi&Park (2021)				O	
Sim(2021)				O	

<Table 6> Criteria for classifying XAI models[6]

Criteria	Meaning of criteria
Transparent models	· These models mean that have some interpretability by themselves.
Post-hoc explainability	· Post-hoc explainability is a method for models that cannot be easily interpreted. · Text explanation, Visual explanation, Local explanation, Example explanation, Simplification explanation etc.

투명한 모델(Transparent models)이란 모델 그 자체가 어느 정도의 해석력을 지니는 것을 의미한다. 모델을 기반으로 인공지능의 결정에 대한 해석이 구조적으로 가능하다. 사후 설명가능성(Post-hoc explainability)는 쉽게 해석할 수 없는 모델을 위해 고안된 것이다. 이 방법은 외부 기법에 의해 설명되는 것인데 주로 사용되는 기법은 텍스트 설명, 시각적 설명, 부분 설명, 예시 설명, 단순화 설명 등이 있다. 텍스트 설명(Text explanation)이란, 모델의 기능을 심볼(symbol)로 생성하는 것을 의미하는데 인간의 언어나 수식 등을 활용한다. 시각적 설명(Visual explanation)이란, 모델의 행동을 시각적으로 설명하는 것으로 가장 많은 방법들과 함께 사용되어 인간의 이해도를 높인다. 부분 설명(Local explanation)은 전체 모델의 일부만을 집중하여 설명하는 것이다. 예시 설명(Example explanation)은 특정 모델로 인해 생성된 결과와 관련된 예시를 추출하는 방식이다. 단순화 설명(Simplification explanation)은 훈련된 모델에 기초하여

복잡한 것을 줄이며 설명을 위한 새로운 시스템을 만드는 방법이다[6]. XAI의 분류기준은 일반적인 것이며 각 관점 중 하나에 귀속시키는 것도 아니다. 어떤 상황의 경우 두 가지 분류기준에 모두 속하기도 한다[2][6].

<Table 7>는 <Table 4>의 제시된 기준을 바탕으로 초등학교 인공지능 프로그램에서 자주 사용되는 인공지능 모델을 추출하여 설명할 수 있는 전략을 정리하였다.

<Table 7> Explainability methods by AI algorithm[6]

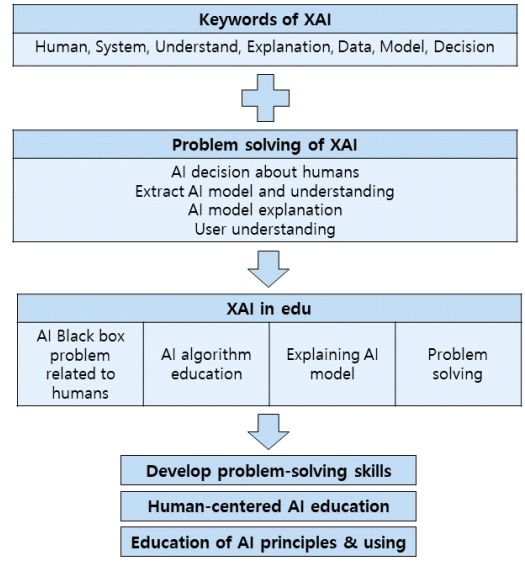
AI algorithm	Explainability methods
Decision Tree	· Transparent model to fulfill every constraint of transparency.
Linear Regression	· Transparent model to takes the assumptions of linear dependence between the predictors and the predicted variables. · Post-hoc explainability techniques(mainly, visualization) also demanded.
K-NN (K-Nearest Neighbors)	· Transparent model relying on the notion of distance and similarity of data
CNN (Convolutional Nerual Network)	· Post-hoc explainability : Visualization that human cognitive skills favors the understanding of visual data.

의사결정트리(Decision Tree)는 모델 자체가 가장 뚜렷한 해석력을 가지고 있으며 예측 결과를 해석하는 과정이 직관적이다. 훈련데이터를 의사결정트리 그래프로 구현하여 해석력을 확보할 수 있다[2][6]. 선형 회귀(Linear Regression)는 독립 변수가 종속 변수의 관계를 직선으로 표현한 것으로 예측 결과에 영향을 미치는 독립 변수를 추리는 과정만으로 해석력을 지닌다[2][12]. 이는 투명한 모델로 분류되지만 비전문가 사람들에게 설명할 때는 사후 설명 가능한 기법(주로, 시각화 기법)가 병행되어야 효과적이다[6]. K-NN(K-Nearest Neighbors)은 데이터를 가장 가까운 유사 속성에 따라 분류하여 데이터를 분류하는 기법으로 마찬가지로 투명한 모델이다[14][6]. 모델의 해석 가능성 관점에서 K-NN모델은 데이터들의 거리와 유사성의 개념에 의존한다는 것을 관찰하는 것이 해석에서 핵심이다. 사후 설명 가능한 기법은 주로 딥러닝 모델에서 활용된다[6]. 딥러닝 모델 중, 초등학교에서 자주 활용되는 것이 이미지 인식 모델이며 CNN(Convolutional Neral Network)은 데이터를 통해 이미지를 직접 학습하고 패턴을 분석해 이미지를 분류하고 개체를 감지하는 딥러닝 모델이다. 인간의 인지 능력이 시각 데이터 이해를 선호하므로 시각화기법을 활용하여 CNN모델의 경우 다른 딥러닝 모델에 비해 설명하기 수월하다[6].

4. 연구결과

4.1. 설명 가능한 인공지능의 교육적 의미

빈도 분석을 통한 설명 가능한 인공지능의 개념 탐구, 설명 가능한 인공지능의 문제 해결 과정 및 전략 탐구를 통해 도출된 교육적 의미를 (Fig. 3)와 같이 나타내었다.



(Fig. 3) Educational meaning of XAI

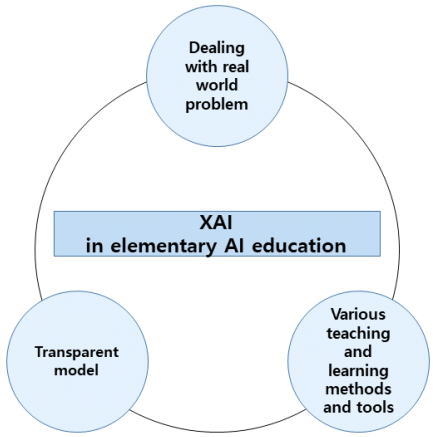
설명 가능한 인공지능의 빈도 분석을 통하여 '인간', '시스템', '이해', '설명', '데이터', '모델', '의사결정'과 같은 핵심어를 추출하였다. 또한, 각 연구들의 공통점을 바탕으로 설명 가능한 인공지능의 문제 해결 과정을 탐구하였다. 일반적인 문제 해결 과정은 '인간 문제와 관련된 인공지능의 의사결정 -> 인공지능 모델 추출 및 이해 -> 인공지능 모델 설명 -> 사용자 이해'의 과정을 따른다. 위의 연구를 바탕으로 한 설명 가능한 인공지능의 교육적 의미는 다음과 같다. 설명 가능한 인공지능 교육은 인공지능의 블랙박스 문제를 해결하는 것을 목표로 한다. 그리고, 인공지능의 블랙박스 문제는 인공지

능과 함께 살아가는 사람의 삶과 연관되어 있다. 그 문제를 해결하기 위해서는 인공지능 알고리즘 교육과 실생활 문제와 관련된 인공지능 모델을 이해하는 것을 넘어 언어와 그림으로 설명할 수 있는 교육이 동시에 이루어져야 한다. 이러한 문제 해결 과정을 통해 학생들은 문제해결력을 기르고, 인간과 관련된 인공지능 문제를 해결함으로써 사람 중심의 인공지능 교육을 가능하게 한다. 그리고, 알고리즘 교육을 통해 인공지능의 원리를 이해하고 실생활 문제 상황과 관련된 인공지능 모델을 설명하며 인공지능의 활용 분야까지 확장할 수 있다.

인공지능 교육의 영역이 서로 간 연계가 되어야 한다는 점에서 볼 때[1], 설명 가능한 인공지능 교육은 인공지능의 원리, 활용, 윤리적인 부분까지 다룬다는 점에서 인공지능 교육 요소에 중요한 부분이다.

4.2. 초등학교에서 설명 가능한 인공지능 교육 방안

설명 가능한 인공지능을 교육적 의미를 바탕으로 한 초등학교에서 설명 가능한 인공지능의 구체적 교육 방안은 (Fig. 4)와 같다.



(Fig. 4) XAI methods in elementary school

첫째, 실제 삶과 관련된 문제를 설명 가능한 인공지능 교육 전반에 적용될 수 있도록 해야 한다. 예를 들어, 우리가 자주 접하는 다양한 인공지능 추천 시스템 예를 가지고 K-Nearest Neighbors의 이해를 극대화 시킬 수 있다. 그리고, 인공지능 추천 시스템이 어떻게 만들어지고, 나의 삶에 영향을 주는지 이해하면서 인공지능의 블랙박스 문제를 해결하는데 동기부여가 될 수 있다. 둘째, 초등 인공지능 알고리즘 교육에서는 알고리즘 자체가 해석력을 지닌 것을 사용하는 경우가 많다. 초등 인공지능 교육 프로그램 연구에서 가장 많이 사용되었던 알고리즘은 의사결정트리이다. 그 이유는 직관적이고 학생이 이해하기 쉽기 때문이다. 따라서, 인공지능 모델을 설명하는 활동에 자체 해석력을 확보한 모델을 사용하는 것이 좋을 것이다. 셋째, 인공지능 모델의 이해가 설명으로 나아가기 위해서는 다양한 교수학습방법 및 도구를 활용해야 한다. 누군가를 이해시키는데 가장 효과적인 기법 중 하나는 시각화이다. 실제로, 인공지능 모델을 설명하는 기법으로서 시각화가 가장 많이 사용된다. 따라서, 이를 위한 언플러그드 활동, 시각화 기능을 지원해주는 모듈형 도구를 활용하는 방법, 데이터 시각화 플랫폼 등을 다양하게 적용할 수 있다.

5. 결론

본 연구는 문헌 연구를 통한 설명 가능한 인공지능의 교육적 의미를 도출하고자 하였다. 설명 가능한 인공지능 개념을 탐구하기 위하여 자연어 처리 및 분석에 자주 활용되는 NLTK(Natural Language Tool Kit) 패키지를 활용하여 빈도 분석을 실시하였다. 그 결과, '사람(human, user)'이라는 단어가 빈도가 가장 높게 나왔으며 순서대로 '인공지능(AI)', '시스템(system)', '이해하다(Understand)', '설명(Explanation)', '데이터(data)', '모델(model)', '의사결정(decision)' 등의 단어가 상위 빈도에 차지하였다. 이를 통해, 설명 가능한 인공지능은 사람이 중심이며 사람과 관련된 의사결정 문제를 해결하고 사용자가 이해할 수 있도록 인공지능 모델을 설명해주는 시스템이라는 몇 가지 의미를 도출하였다.

또한, 문헌에 나타난 개념과 문제해결과정을 바탕으로 설명 가능한 인공지능의 교육적 의미를 탐구하였다. 교육에서 설명 가능한 인공지능은 인간의 의사결정에 영향을 미치는 인공지능 블랙박스 문제 상황을 제시해야 한다. 또한, 인공지능 원리를 탐구할 수 있는 알고리즘 교육과 이를 인간의 언어, 그림 등으로 설명할 수 있는 교육까지 나아가야 한다. 이를 통해, 학생들의 문제해결력을 기르고 사람 중심의 인공지능 교육, 인공지능 원리, 활용 교육을 동시에 가능하게 한다. 초등학교에서

이를 활용하기 위한 구체적인 방안으로는 실제 삶과 관련된 문제를 설명 가능한 인공지능 교육 전반에 적용되게 해야 한다. 그리고, 알고리즘 자체가 해석력을 지닌 것을 사용하는 것이 좋으며, 모델의 이해가 설명으로 나아가기 위해서는 언플러그드 활동, 시각화를 지원해주는 인공지능 도구, 데이터 시각화 플랫폼 등 다양한 교수학습방법과 도구가 활용될 수 있다.

본 연구의 한계점으로는 몇 가지 문헌을 중점적으로 연구를 진행하였으므로 문제해결과정 및 교육적 의미를 일반화하기 어렵다는 점이다. 그러나, 기존의 국내 인공지능 교육에서 많이 연구되지 않았던 설명 가능한 인공지능을 바탕으로 교육적 의미를 도출하려 했다는 점과 2022년 개정 교육과정에 인공지능의 도입을 앞두고 새로운 인공지능 교육 개념과 내용을 제시했다는 점에서 본 연구는 의미가 있다.

향후 연구로는 실제 초등학생들이 설명 가능한 인공지능을 쉽게 이해할 수 있는 교육 프로그램의 개발과 설명 가능한 인공지능 교육으로 기를 수 있는 학생 역량 및 프로그램의 효과성 검증에 관한 연구가 필요하다.

참고문헌

[1] Ministry of Education, Korea (2020). Direction of Education Policy and Key Tasks in the Age of Artificial Intelligence.

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

[3] Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.

[4] Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine, 40*(2), 44-58.

[5] Doran, D., Schulz, S., & Besold, T. R. (2017). What does explainable AI really mean? A new conceptualization of perspectives. arXiv preprint arXiv:1710.00794.

[6] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82-115.

[7] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR), 51*(5), 1-42.

[8] Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019, July). Meaningful explanations of black box AI decision systems. *In Proceedings of the AAAI conference on artificial intelligence 33*(1), 9780-9784.

[9] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics, 4*(37).

[10] Kim, J. Moon, S.(2021). Development of an AI Education Program based on Novel Engineering for Elementary School Students, *The Journal of Korea Elementary Education, 32*(1), 425-440.

[11] Lee, S. Moon, G.(2021). Development of AI Education Program for Prediction System Based on Linear Regression for Elementary School Students, *Journal of The Korean Association of Information Education. 12*(2), 51-57.

[12] Jang, Y.(2019). *Development of unplugged education program for elementary school AI classes*. Seoul National University of Education.

[13] Jang, M.(2020). *Unplugged Education Program for Artificial Intelligence Education in Elementary Schools : Focus on 'Constraint satisfaction problem'*. Gyeongin National University of Education.

[14] Choi, E. Park, N.(2021). Application and Development of Machine Learning Training Program based on Understanding K-NN Algorithm. *Journal of The Korean Association of Information Education. 25(1),* 175-184.

[15] Sim, S.(2021). *A Study on the Development and Effectiveness of Python-based Artificial Intelligence Education Program in Elementary School : Focusing on Machine Learning*. Daegu National University of Education.

[16] Ryu, M. Han, S.(2019). AI Education Programs for Deep-Learning Concepts. *Journal of The Korean Association of Information Education. 23(6),* 583-590.

[17] You, J.(2020). Report on Exploratory research issues in the contetns system of AI education in elementary and secondary schools. *The Korea Foundation for Science and Creativity.*

[18] Jason brownlee(2020). Difference Between Algorithm and Model in Machine Learning Retreive from https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/

[19] Shin S(2021). A Constructive Characteristic Analysis Study of Modular Data Analysis Tools. *Collection of academic presentations at the 2021 Summer Conference of the Korean Society of Information Education,* 74-77.

[20] Presidential Committee on the Fourth Industrial Revolution(2021). Retreived from https://www.4th-ir.go.kr/pressRelease/detail/1437?category=report

[21] Alonso, J. M. (2020). Teaching Explainable Artificial Intelligence to High School Students. *International Journal of Computational Intelligence Systems, 13*(1), 974-987.

[22] Choi, J(2019). Recent Advances in Explainable Artificial Intelligence. *Korea Information Science Association, 37*(7), 8-13.

[23] Ministry of Education, Korea et al(2021). Explainable Artificial intelligence for middle school students.

[24] Shin, S. (2021). A Study to Design the Instructional Contents for National Curriculum of Computer Education in Elementary School. *Journal of The Korean Association of Information Education, 25*(1), 13-31.

저자소개

박 다 빈

2017 서울교육대학교 초등교육과 (학사)

2020~현재 서울교육대학교 교육대학원 인공지능교육과(석사)

2018~현재 서울연은초등학교 교사

관심분야: 인공지능 교육, 보편적정보교육, 설명가능한 인공지능

e-mail: dabin2688@gmail.com

신 승 기

2017 University of Georgia(Ph.D.)

2016~2017 미국 칼빈슨 정부연구소 연구원

2019~2020 애리조나주립대학교 컴퓨터교육전공 교수

2020~현재 서울교육대학교 컴퓨터교육과 교수

관심분야: Computational Thinking, 인공지능교육, 보편적정보교육

e-mail: skshin@snue.ac.kr