

인공지능 윤리 딜레마 문항 개발

김은경^o, 이영준^{*}

^o한국교원대학교 컴퓨터교육과,

^{*}한국교원대학교 컴퓨터교육과

e-mail: egkim618@gmail.com^{*}, yjlee@knue.ac.kr^o

Development of AI Ethics Dilemma Questions

Eun-Gyeong Kim^o, Young-Jun Lee^{*}

^oDept. of Computer Education, Korea National University of Education,

^{*}Dept. of Computer Education, Korea National University of Education

● 요약 ●

인공지능 기술이 사회 전반에 확산하며, 인공지능 윤리 문제 또한 큰 이슈가 되었다. 이에 따라 정부, 기업, 연구기관, 국제기구 등 다양한 단체에서 인공지능 기준안을 발표하고 있으나 이러한 인공지능 윤리 기준안을 발표하는 것으로 인공지능 윤리 문제를 해결할 수 없다. 이에 인공지능 윤리교육이 절실히 필요하다. 그러나 학교급에서의 인공지능 윤리교육을 위한 프로그램이 미흡하다. 이러한 상황에서 Moral machine은 인공지능을 위한 윤리교육 도구로 학교급을 막론하고 활발하게 사용되고 있다. 그러나 Moral machine의 딜레마 상황은 교육용으로 다소 부적절한 요소가 있다. 이에 본 연구에서는 인공지능 윤리교육에 적합한 인공지능 윤리 딜레마 문항을 사람이 중심이 되는 인공지능 기준의 3대 원칙을 바탕으로 개발하였다.

키워드: 인공지능 윤리(AI ethics), 딜레마(dilemma)

I. Introduction

인공지능 기술이 우리 삶에 미치는 영향이 커질수록 인공지능 윤리 문제에 대한 중요성 또한 커지고 있다. 이러한 흐름에 따라 국가, 기업, 연구기관, 국제기구 등 다양한 곳에서 인공지능 윤리 기준을 발표하고 있다[1]. 그러나 McNamara의 연구에 따르면 윤리 기준을 제시하고 명시적으로 따르게 한 집단과 그렇게 하지 않은 집단들 사이의 의사결정에 유의한 차이가 없는 것으로 나타났다[2]. 이러한 결과는 인공지능 윤리 기준을 발표하는 것이 인공지능 윤리 문제를 해결하는 대안이 될 수 없음을 이야기한다. 따라서 이렇게 제시된 인공지능 윤리 기준을 바탕으로 한 인공지능 교육이 별도로 이루어져야 한다.

우리나라 공교육에서의 인공지능 교육은 이제 시작 단계이며 특히, 인공지능 윤리와 관련된 교육자료는 부족한 실정이다. 이런 상황에서 Moral machine의 자율주행차 딜레마는 학교급을 가리지 않고 인공지능 윤리교육 소재로 활발하게 활용되고 있다. 그러나 Moral machine은 본래 교육을 위해 개발된 도구가 아니며 자율주행차라는 제한적인 주제만을 다루고 있어 인공지능 윤리교육 도구로 사용하기에는 미흡한 점이 있다.

이에 본 연구에서는 인공지능 윤리교육에 더 적합한, 인공지능 윤리 기준을 바탕으로 한 인공지능 윤리 딜레마 문항을 개발하고자 한다.

II. Preliminaries

1. 사람이 중심이 되는 인공지능 윤리 기준

인공지능이 사회에서 차지하는 비중이 커짐에 따라 인공지능 윤리 문제에 대한 관심이 크게 증가하였고, 이에 따라 다양한 단체에서 인공지능 윤리 기준을 발표하고 있다. 이중 과학기술정보통신부에서 발표한 사람이 중심이 되는 인공지능 윤리 기준에서는 ‘인간의 존엄성’, ‘사회의 공공선’, ‘기술의 합목적성’이라는 3대 기본 원칙과 10대 핵심 요건을 제시하였다[1]. 본 연구에서는 이 기준안의 3대 원칙을 바탕으로 인공지능 윤리 딜레마 문항을 개발하였다.

2. Moral machine

Moral machine은 고전 트롤리 딜레마에 자율주행차를 접목한 인공지능 윤리 사고 실험 플랫폼이다. Moral machine을 통해 수집된 데이터를 분석한 논문이 Nature에 게재된 이후 Moral machine은 인공지능 윤리 논의를 주도하게 되었고 국내에서도 인공지능 윤리교육 도구로 활발하게 사용되고 있다[4]. 그러나 발생 가능성이 극히 드문 극단적인 사례에 초점을 두어 대중을 겁먹게 할 수 있다는 지적과 인간 존엄성의 기본 원칙을 위반한다는 비판을 받고 있어 교육 도구로 활용하는 것에 숙고가 필요하다[3, 4]. Moral machine을 활용한 인공지능 윤리교육이 초등학교의 인공지능에 대한 태도에 부정적이라는 연구 결과 또한 이러한 주장을 뒷받침해준다[5]. 그러나 이러한 지적에도 Moral machine의 딜레마 사례는 인공지능 윤리교육에 대한 학습자들의 몰입도를 높여주는 흥미로운 소재임이 분명하다.

III. The Proposed Scheme

본 연구는 인공지능 윤리교육에 활용할 수 있는 다양한 인공지능 윤리 딜레마 문항을 개발하는 것을 목적으로 한다. 문항은 ‘인간 존엄성’과 ‘사회 공공선’, ‘사회 공공선’과 ‘기술 합목적성’, ‘기술 합목적성’과 ‘인간 존엄성’의 가치가 상충하는 3개 카테고리로 나뉘며 카테고리마다 3문항씩 총 9개의 문항을 만들었다. 각 문항의 우리 생활에서 이미 문제가 되고 있거나 곧 문제가 될 수 있는 생활 밀접형 문제들로 몰입도를 높였다. 또한 인공지능 기술에 대한 이해는 인공지능 윤리교육에 밑거름이 되어야 하기 때문에 다양한 인공지능 기술이 반영된 사례를 제시하여 문항을 통해 인공지능 기술에 대한 학습도 함께 이뤄질 수 있도록 하였다. 학습자는 딜레마 문항에서 본인이 생각하는 최선의 선택을 하기 위해 문항에 대해 더 이해하려고 할 것이고, 그 속에 녹아있는 가치에 대해 진지하게 고민할 것이다. 이러한 과정은 인공지능 윤리 기준의 내면화와 기술에 대한 이해를 도울 것이다.

IV. Conclusions

본 연구에서는 사람이 중심이 되는 인공지능 윤리 기준을 바탕으로 한 인공지능 윤리 딜레마 문항을 개발하였다. 개발된 문항은 총 9개 문항으로 인간 존엄성, 사회 공공선, 기술 합목적성 사이의 딜레마를 다양한 인공지능 기술이 적용된 상황으로 표현하였다.

개발된 문항은 Moral machine의 딜레마 상황에 비해 다양한 인공지능 윤리 문제 상황을 담고 있으며 인공지능 기술로 인한 문제뿐 아니라 이로온 점도 함께 제시하고 있어 인공지능 윤리교육을 위한 자료로 활용되기에 더 적합하다. 다만 문항에 대한 접근성이 웹 기반의 Moral machine에 비해 많이 부족한 편으로 이는 교육적 활용에 장애요인이 된다. 이에 후속 연구에서는 본 연구에서 개발된 문항을 담은 웹 기반의 인공지능 윤리 도구를 개발이 요구된다.

REFERENCES

- [1] Ministry of Science and ICT. (2020). AI ethical standards
- [2] McNamara, A., Smith, J., & Murphy-Hill, E. (2018, October). Does ACM’s code of ethics change ethical decision making in software development?. In Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering (pp. 729-733).
- [3] Kochupillai, M., Lütge, C., & Poszler, F. (2020). Programming away human rights and responsibilities? “The Moral Machine Experiment” and the need for a more “humane” AV future. NanoEthics, 14(3), 285-299.
- [4] Jaques, A. E. (2019). Why the moral machine is a monster. University of Miami School of Law, 10, 1-10.
- [5] Kim, E & LEE, Y. (2022). The influence of artificial intelligence ethics education using Moral machine on elementary school students’ perception of artificial intelligence. Korean Association of Artificial Intelligence Education,, 25(3), 1-8.