

텍스트 마이닝으로 살펴본 대학생들의 인공지능 윤리 인식 연구

유수진* · 장운재**
한양대학교* · 경기대학교**

요약

본 연구는 대학생의 인공지능 윤리 인식을 파악하여 대학 교양 인공지능 윤리 교육의 방향성을 탐색하고자 한다. 대학생 83명이 총 5개의 인공지능 윤리 토론 주제에 대한 의견을 작성하고, 작성된 텍스트를 기반으로 텍스트 마이닝 중 언어 네트워크를 이용하여 분석하였다. 분석 결과, 첫째, 인공지능 사회의 미래에 대해서 62.5%의 학생이 긍정적으로 바라보고 있었다. 둘째, 자율주행 자동차 사고 발생 시 39.2%의 학생이 현재 자율주행 수준으로는 차량 소유자의 책임으로 생각하였다. 셋째, 인공지능 발전의 역기능으로는 사생활 침해와 기술 악용, 정보편식을 꼽았다. 역기능 최소화 방안으로 인공지능 사용자와 개발자 모두의 윤리 교육이 필요하며 제도적인 준비도 병행되어야 한다고 언급하였다. 넷째, 얼굴 인식 기술이 보편화된 사회에 대해 19.2%의 학생만이 긍정적인 의견을 나타내었다. 마지막으로 데이터 수집 시 개인정보 이용 동의를 얻은 부분만 활용해야 할 뿐 아니라 도덕적인 기준이 없는 인공지능 활용 방안에 대해 사용자와 개발자 모두의 윤리적 소양을 강조하였다. 본 연구는 대학 교양 수준의 인공지능 윤리 교육을 설계할 때 시사점을 제공한다는 점에서 의의가 있다.

키워드 : 인공지능 윤리, 교양교육, 온라인 토론, 텍스트마이닝, 언어 네트워크 분석

A Study on Artificial Intelligence Ethics Perceptions of University Students by Text Mining

Sujin Yoo* · YunJae Jang**
Hanyang University* · Kyonggi University*

Abstract

In this study, we examine the AI ethics perception of university students to explore the direction of AI ethics education. For this, 83 students wrote their thoughts about 5 discussion topics on online bulletin board. We analyzed it using language networks, one of the text mining techniques. As a result, 62.5% of students spoke the future of the AI society positively. Second, if there is a self-driving car accident, 39.2% of students thought it is the vehicle owner's responsibility at the current level of autonomous driving. Third, invasion of privacy, abuse of technology, and unbalanced information acquisition were cited as dysfunctions of the development of AI. It was mentioned that ethical education for both AI users and developers is required as a way to minimize malfunctions, and institutional preparations should be carried out in parallel. Fourth, only 19.2% of students showed a positive opinion about a society where face recognition technology is universal. Finally, there was a common opinion that when collecting data including personal information, only the part with the consent should be used. Regarding the use of AI without moral standards, they emphasized the ethical literacy of both users and developers. This study is meaningful in that it provides information necessary to design the contents of artificial intelligence ethics education in liberal arts education.

Keywords : Artificial Intelligence Ethics, Liberal Arts Education, Online Discussion, Text Mining, Language Network Analysis

교신저자 : 장운재(경기대학교 SW중심대학사업단)

논문투고 : 2021-10-25

논문심사 : 2021-11-02

심사완료 : 2021-11-20

1. 서론

인공지능이란 무엇일까. Alan Turing(1950)에 따르면 '생각하는 기계'이며[33], John McCarthy(1956)는 닥터 머스 회의에서 '기계가 사람 행동의 지식처럼 행동하게 하는 것'이라고 정의하였다[22]. Stuart Russel & Peter Norvig(1994)은 합리적 사고, 합리적 행동, 인간적 사고, 인간적 행동으로 나누어 정의하였다[30].

1950년대 인공지능이 처음 등장한 이래로 현재 우리는 3차 인공지능 붐의 시대를 살아가고 있다[34]. 인공지능 기술을 사용하여 제조업 분야에서 생산라인을 자동으로 분류하거나, 의학 분야에서 종양을 진단하기도 한다. 또한 언론 분야에서 뉴스 자동 생성, 금융 분야에서 수익 예측 및 패턴 파악을 통한 이상 감지 등 다양한 분야에서 이미 인공지능 기술을 활용하고 있다. 이처럼 인공지능 기술을 통해 다양한 분야에서 비약적 발전을 이루기도 하였다. 3차 붐은 컴퓨팅 파워 증가, 빅데이터와 인공지능 기반의 기계학습, 딥러닝이라는 큰 파도가 만나 일어나게 되었으며 현재 많은 사람들이 접하게 되는 인공지능 기술은 딥러닝의 산물이다[20].

Google DeepMind가 만든 AlphaGo는 인공지능을 일반 대중에게도 알렸다. 많은 사람들은 인공지능이 Alan Turing의 정의처럼 '생각하는 기계'인 것처럼 말한다. 하지만 사람들이 생각하는 인공지능은 인공 일반지능(AGI) 혹은 강인공지능으로 아직 성취되지 않았으며 현재 우리가 접하는 인공지능 기술들은 일부 기능이 특화된 약인공지능의 형태로 발전되고 있다. 이러한 기능들은 사람의 이미지 인식과 같은 컴퓨터 비전 분야, 음성 신호 처리와 음성 인식, 언어 처리 등에서 좋은 성능을 보인다. 하지만 딥러닝의 내부 연산 과정은 사람들이 보고 들을 때 생각하는 과정과는 다르다. 즉, 인공지능의 정의처럼 기계가 정말 생각해서 행동하는 사람의 고차원적인 이성적 사고과정이 아닌 감각 인식 작용에 국한되고 있다.

이러한 특성은 인공지능으로 하여금 사람처럼 어떤 데이터를 보고 이성적, 윤리적으로 판단하여 선별해서 받아들이는 것이 아닌 잘못된 데이터를 주면 잘못된 결과를 낳게 만든다. 이러한 이유로 많은 인공지능 기술이 윤리 문제를 발생시키기도 한다. 기업에서 개발한 챗봇이 비윤리적인 대화를 학습하여 성차별, 인종차별적 발

언을 공개적으로 하였고, 뿐만 아니라 인공지능 학습을 위한 데이터 수집 단계에서도 법적인 문제를 야기하기도 하였다[19][21][28].

데이터로 인한 이슈와 별개로 이전에는 발생하지 않았던 새로운 윤리적인 문제들이 등장하고 있다. MIT에서 개발한 Moral Machine은 자율주행 자동차의 교통사고 장면에서 탑승자 또는 보행자 중 어느 쪽을 살릴지 결정하는 딜레마 문제를 제시하고, 사람들의 응답 결과를 분석하였다. 전 세계 233개국에서 230만 명이 참여한 3,961만 개의 판단 결과를 분석한 연구에서는 전 세계 문화권별, 국가별로 판단의 기준이 다르다는 것을 발견하였다[1][25].

인공지능 윤리는 인공지능 기술과 함께 다뤄져야 하며, 인공지능 기술을 사용하는 사람의 다양한 역할, 즉 개발자, 이용자, 공급자의 측면에서 지켜야 할 윤리 규범과 원칙을 제시하고 있다[2][26]. 또한 인공지능 윤리와 관련된 이슈, 규범과 원칙, 딜레마 등을 활용한 교육을 학교 교육에 반드시 포함하도록 권고하고 있다[27]. 초중등학교에서 인공지능 시민성, 역기능 대처 등의 인공지능 윤리 교육이 필요하며, 대학에서도 인공지능의 사회적 영향력과 윤리적 문제를 교육할 필요가 있다[4][29].

대학 교양 수준의 인공지능 교육은 SW·AI 중심대학을 중심으로 교양 교과목을 개설하여 운영되고 있으나 이제 시작하는 단계로 교육 사례가 부족하며, 특히 인공지능 윤리 교육과 관련된 교육 방향과 내용 구성에 대한 체계성이 아직 부족한 실정이다. 따라서 본 연구에서는 대학생들의 인공지능 윤리 문제에 대한 인식을 파악하여 대학 교양 수준의 인공지능 윤리 교육의 방향성을 탐색하고자 한다.

본 연구의 구체적인 연구 문제는 다음과 같다. 첫째, 인공지능 윤리 문제에 대해 대학생들은 어떻게 생각하고 있는가? 둘째, 대학 교양 수준의 인공지능 윤리 교육은 어떤 방향성을 갖춰야 하는가?

인공지능 윤리 문제에 대한 대학생들의 인식을 분석하기 위하여, 온라인 토론 문제를 이용하여 학생들의 토론 내용을 텍스트 마이닝 기법을 활용하여 분석한다. 분석 결과를 토대로 인공지능 교육의 방향성을 제시하고자 한다.

2. 이론적 배경

2.1. 인공지능 윤리 연구

1950년대부터 인공지능 연구가 시작되고 그 뒤를 이어 1960년대에 인공지능 윤리 연구가 시작되었다[31]. 그 이후 인공지능 기술이 최근 몇 년 사이에 급성장하면서 인공지능 기술의 활용에 따른 현실적 문제를 논의하기 위한 인공지능 윤리 연구 역시 2016년을 기점으로 점차 확대되어 논의되고 있다[11]. 교육 분야에서 인공지능 윤리에 대한 논의는 로봇 공학 윤리와 정보 윤리에서 논의되고 있다.

로봇 윤리에서 출발하는 인공지능 윤리는 지능을 가진 로봇이 인간과 어떻게 상호작용해야 하는가를 묻고 있다. 로봇 윤리의 시초는 Isaac Asimov의 소설에서 등장하는 로봇공학 3원칙으로부터 시작하여 이탈리아 로봇공학자인 Veruggio(2008)에 의해 로봇 윤리가 용어로 언급되기 시작하였다[3][8][18]. 로봇 공학 3원칙은 인간과 로봇의 안전에 초점을 맞추고 있다. 최근에 지능 로봇이 상용화되기 시작하면서, 지능 로봇이 도출한 의사결정이 인간과 사회에 어떤 영향을 미치는지 고려해야 함을 지적하고 있다. 정채연(2018)은 로봇 윤리 담론에서 인공지능 또는 그것이 포함된 지능 로봇의 도입되면서 변화되는 사회적 영향과 윤리적 평가를 로봇 개발자와 사용자의 윤리적 지침 그리고 로봇 스스로 윤리적 판단의 주체로서 논의가 확장되고 있다고 언급하였다[12]. 최현철과 변순용(2019)은 인공지능 로봇이 인공지능 윤리 주체(Artificial Moral Agent, AMA)로서의 지위를 갖게 된다는 점을 지적하면서 인공지능의 결정이 도덕적으로 어떤 의미를 갖는지 생각해 봐야 한다고 언급하였다[5]. 김효은(2020)은 의사결정 자동화에 대한 대응으로서의 인공지능 윤리 교육을 언급하면서 인공지능 윤리는 기존의 공학 윤리와 차이가 있다고 하였다[15]. 공학윤리의 경우 결정 주체가 사람으로 정해져 있는 반면 인공지능 윤리는 자율지능시스템의 자율성으로 인해 시스템 설계자의 예상과 다른 결과가 나올 수 있다는 점을 지적한다. 따라서 인공지능 윤리는 ‘사람’이 갖춰야 하는 윤리와 ‘인공지능’이 갖춰야 하는 윤리로 구분해야 한다는 것이다.

정보윤리는 컴퓨터로 대표되는 디지털 정보 기기에서

다뤄지는 정보와 관련하여 발생할 수 있는 윤리적 문제를 다루는 분야이다. 정보윤리는 Moor(1985)의 컴퓨터 윤리란 무엇인가? 라는 질문에서 논의가 시작되었다[24]. Moor는 컴퓨터 기술이 새롭게 등장하면서 인간과 사회에 새로운 영향을 미칠 것이며 이에 대한 정책의 공백 상태가 존재하기 때문에 컴퓨터 윤리가 논의되어야 한다고 주장하였다. Holmes et al(2018)는 인공지능을 통한 교육에서 Moor가 제기했던 문제가 여전히 있다고 주장하였다[10]. 인공지능을 통한 교육의 윤리를 다룬 연구자, 가이드라인, 정책개발, 규범이 부족하며, 학교 교육에서 인공지능 활용에 따른 부정적인 영향력에 대한 도덕적 공백(moral vacuum) 여전히 존재함을 지적하였다. 김한성(2019)은 정보 윤리에서 다루지는 윤리적 이슈를 분석하여 다양한 윤리적 이슈를 제시하였다[16]. 인공지능 윤리에 대한 이슈로 ‘인공지능의 사회적 영향’을 제시하였다. 인공지능이 우리 삶에 어떤 영향을 미치는지, 또한 인공지능의 도입으로 인해 이전에 경험하지 못했던 어떤 윤리적인 문제가 있는지 등이 논의되는 내용이다.

2.2. 인공지능 윤리 원칙 탐색

인공지능 기술 발전에 따른 사회적 문제를 해결하기 위해 전 세계적으로 인공지능 윤리 연구가 진행되고 있다. OECD, EU와 같은 국제기구에서 인공지능 윤리 가이드라인을 제시하였으며, 인공지능 관련 기업, 개발자, 연구자들이 모여 인공지능 윤리 원칙을 제시하였다[23].

각 국가별, 기업별로 제시한 인공지능 윤리 기준 및 활용 원칙은 <Table 1>과 같다[9][13].

<Table 1> AI Ethics Principles

Institution	Content
EU(2018)	Ethics guidelines for trustworthy AI
OECD(2019)	OECD Principles on AI Reflected it to G20 Leaders' Declaration
UNESCO(2019)	Recommendation on the ethics of artificial intelligence
Japan(2019)	Social Principles of Human-Centric AI
Korea(2020)	Artificial Intelligence (AI) Ethics Standards
Google(2018)	Google's principles for Artificial Intelligence
Kakao(2021)	Algorithmic Ethics Charter

Jobin et al(2019)은 다양한 민간 부문이나 기관이 정의하는 인공지능 윤리 분야 원칙 및 윤리적인 인공지능 실현에 필요한 요구사항들이 수렴하는지 혹은 차이가 있다면 조정 가능한지 확인하고자 전 세계 주요 84개의 윤리 가이드라인을 분석하였다[11]. 분석 결과, 11개의 핵심 요소를 제시하였으며, 각 핵심 요소는 투명성, 정의, 해악금지, 책임, 사생활보호, 혜택, 자유, 신뢰, 지속가능성, 존엄성, 연대성이다.

Fjeld et al(2020)는 다양한 AI 원칙의 동향 파악, 맥락 비교를 위해 36개의 저명한 인공지능 원칙 문서 내용을 비교 분석하여, 8가지 핵심 원칙을 도출하였다[7]. 8가지 핵심 원칙은 사생활 보호, 책무성, 안전과 보안, 투명성과 설명가능성, 공정과 차별금지, 인간에 의한 기술 통제, 전문적 책임, 인간 가치의 증진이다.

본 연구에서는 인공지능 기술의 등장으로 발생하는 인간과 사회의 변화를 중심으로 기존의 로봇 공학 윤리와 정보 윤리가 확장되는 인공지능 윤리에 초점을 맞추고자 한다. 또한 인공지능 윤리에서 지적하는 윤리적 판단 주체로서의 인공지능 기술의 중요성을 토론 주제로 포함하였다. 또한 다양한 국가와 기업에서 제시한 인공지능 윤리 핵심 원칙을 바탕으로 토론 주제를 선정하였다.

3. 연구방법

3.1. 연구대상

연구 대상은 서울 H대학교에서 2021년 여름 계절학기에 개설된 ‘인공지능과 기계학습’ 교양 과목의 수강생 83명을 대상으로 하였다. 수강생의 기본 정보는 <Table 2> 와 같다.

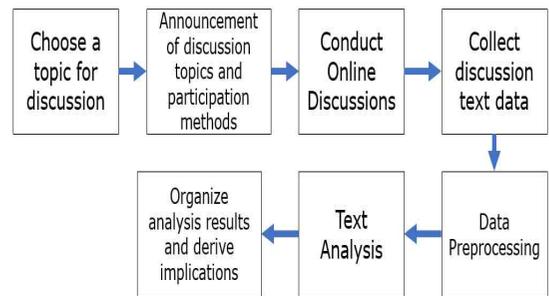
<Table 2> Students' Information (N=83)

Category		n
School year	First Grader	3
	Second Grader	7
	Third Grader	31
	Fourth Grader	42
Major Field	Liberal Arts	7
	Social Sciences	22
	Education	7
	Engineering	25

	Natural Science	12
	Arts, Music and Physical Education	9
	Credit Exchange	1
Gender	Male	46
	Female	37

3.2. 연구절차

본 연구는 대학생들의 인공지능 윤리 인식을 파악하기 위하여 AI 윤리 문제에 대한 온라인 토론 내용을 분석하였다. 연구 절차는 (Fig. 1)과 같다.



(Fig. 1) Research Process

먼저 선행연구를 바탕으로 토론 주제를 선정했다. 이후 수업에 참여하는 학생들에게 토론 주제 및 방법을 안내하였다. 학생들은 온라인 토론 게시판을 통해 토론을 진행하였다. 온라인에 게시된 토론 내용을 텍스트 데이터로 수집하였다. 수집한 데이터는 전처리하고, 분석하였다. 마지막으로 분석 결과를 정리하여 시사점을 도출하였다.

3.3. 토론 주제

인공지능 윤리 토론 주제는 대학생들이 교양 수준에서 이해할 수 있으면서 인공지능 윤리 원칙에 대해 생각해 볼 수 있는 주제로 선정하였다. 연구자가 선행연구에서 제시된 인공지능 윤리 원칙 중 학생들의 수준과 관심 및 최근 사례를 바탕으로 7개의 이슈를 선정하였다. 7개의 인공지능 윤리 이슈와 조작적 정의는 <Table 3>과 같다.

<Table 3> AI Ethics Issues

AI Ethics Issue	Definition
Sustainability	Sustainable development of AI and AI for sustainability
Fairness	AI for all without bias or distinction
Privacy	Using authorized data only and abuse prevention
Transparency	Disclosure of AI's decision process
Responsibility	Assigning the morally responsible entity of AI development and utilization to minimize damage
Safety	Defining and resolving risks that may arise due to technical limitations
AMA's Moral Status	The need to think about the moral implications of AI decisions

인공지능 윤리 이슈를 중심으로 최근 뉴스를 통해 대학생들이 한번쯤 접해봤을 만한 소재를 선정하여 5개의 토론 주제를 선정하였다. 각 토론 주제와 관련된 인공지능 윤리 이슈는 <Table 4>와 같다.

<Table 4> AI Ethics Discussion Topics

No.	Discussion Topics	AI Ethics Issue
1	The future of the AI society	Sustainability
2	Responsibility for self-driving car accidents	Responsibility, The problems of AMA's Moral status
3	Adverse Functions and Minimization Plans for Artificial Intelligence Development	Fairness, Privacy
4	What is the future for facial recognition?	Privacy, Safety
5	What are the necessary regulations for data collection and ethical use of artificial intelligence?	Sustainability, Fairness, Privacy, Transparency, Responsibility. Safety, AMA's Moral status

첫 번째 주제는 인공지능이 발전된 미래 사회가 우리 생활에 미치는 영향이 긍정적일지 부정적일지 생각해 보는 것이었다. 이 주제를 통해 인공지능의 미래 소양인 지속가능성을 반영하였다. 두 번째는 트롤리 딜레마로 알려진 자율주행 자동차 간의 사고 시 누구의 책임인지 토론해보았다. 이 주제에는 AMA의 도덕적 지위 문제, 책임성이 반영되었다. 세 번째 주제는 인공지능 발전의 역기능에 대해 생각해보고 최소화 방안을 생각해 보는 것이었다. 공정성과 사생활 보호 원칙이 반영되었다. 네 번째로는 얼굴인식 기술이 보편화된 사회가 온다면 어떨지 함께 토의해보았다. 사생활 보호와 안전성 원칙이 반영되었다. 마지막 주제로 데이터 수집 시 필요한 규제가 무엇이며 데이터의 질을 어떻게 보장할 수 있을지, 그리고 도덕적 기준이 없는 인공지능을 어떻게 올바르게 활용할 수 있을지 생각해보았다. 이 주제는 모든 원칙이 반영되었다.

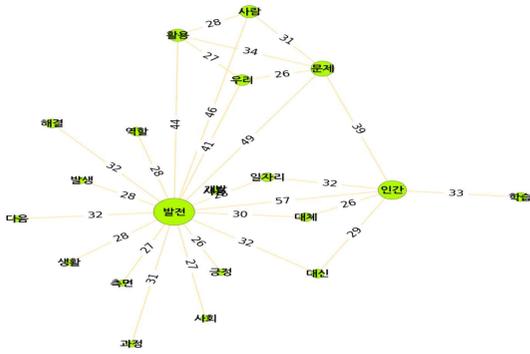
3.4. 분석방법

인공지능 윤리 토론 주제에 대한 학생들의 의견을 수집하여 언어 네트워크 모델을 이용하여 분석하였다.

데이터 전처리 과정으로 각 주제에 대한 학생들의 의견을 긍정, 부정, 중립 견해로 연구자가 구분하고, 구분별로 분석을 실시하였다. 데이터 분석 단계에서는 텍스트의 명사를 추출하여 빈도수를 기준으로 키워드를 선정하고, 키워드간의 관계 구조를 찾기 위하여 언어 네트워크 분석을 실시하였다. 언어 네트워크는 망을 사용하여 용어 간의 동시 출현 정도를 표현함으로써 관계를 도식화하여 보여주는 기법이며[17]. 두 키워드가 동시에 나타나는 경우를 뽑아서 네트워크로 표현한다. 본 연구에서는 키워드 간의 관계를 개개인의 토론 의견 내에서 동시에 발생한 단어쌍을 찾아 동시 발생 빈도 행렬을 만들고, 해당 행렬을 바탕으로 언어 네트워크로 그려서 표현하였다. 언어 네트워크 구성시 가중치는 분석할 토론 주제별 데이터 수를 기반으로 네트워크가 적절하게 구성되도록 연구자가 설정하였다.

분석 도구는 파이썬과 텍스트 분석 라이브러리를 사용하여 분석하였다. Pandas를 이용하여 CSV 파일을 처리하고, 텍스트 분석에는 파이썬의 KoNLPy와 OKT, Counter, NetworkX, tqdm, Numpy를 이용하였다.

마지막으로 중립적인 의견은 (Fig. 4)와 같다. 긍정적인 의견과의 차이점은 ‘발전’에 초점을 맞추었다는 것이다. 인공지능의 발전을 막을 수 없으므로 발생가능한 부작용이 무엇일지 고려하여 제도적 장치 마련이 우선이라는 의견이었다.



(Fig. 4) The future of the AI Society - Neutral

4.3. ‘자율주행 자동차 사고 책임’ 분석 결과

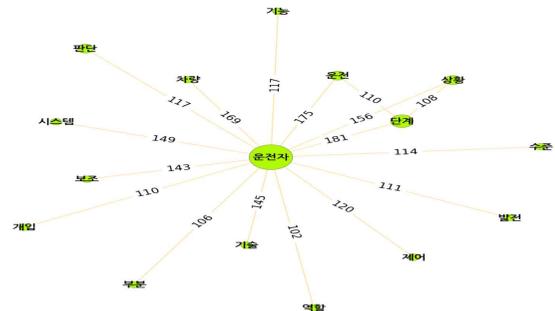
‘자율주행 자동차 사고 책임’에 대한 토론 주제에 대한 학생들의 의견은 운전자 책임, 제조사 책임, 둘 다 책임으로 구분하였다. 분석된 의견 수와 비율은 <Table 7>과 같다.

<Table 7> Who’s responsibility of self-driving car accident?

	Driver	Manufacturer	Both	Total
The number of opinions(n)	31	14	34	79
Rate(%)	39.2	17.8	43.0	100.0

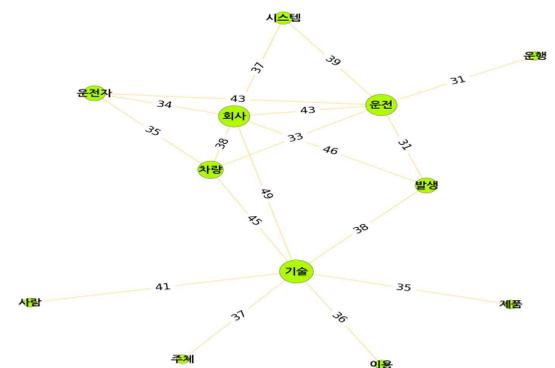
먼저 운전자에게 책임이 있다는 의견에 대한 언어 네트워크 결과는 (Fig. 5)와 같다. 현재의 자율주행자동차 ‘단계’(혹은 ‘수준’)는 사람이 ‘개입’할 수밖에 없으며 자율주행은 ‘보조’기능을 할 뿐이므로 최종 ‘판단’은 사람의 몫이라는 의견이었다. AMA는 인공지능 윤리 주체 혹은 인공적 도덕 행위자라고도 일컬어지는 인공지능에 대한 도덕적 주체성을 의미한다[6]. 도덕 행위자란 인격을 가진 자율적 존재로서 자기 행동을 책임질 수 있는 존재를 의미한다[32]. 자율주행자동차의 AMA 지위에

대해서는 사회적 합의가 필요한 부분임을 확인할 수 있었다.



(Fig. 5) Responsibility for self-driving car accidents - Driver

자동차 제조사에게 책임이 있다는 의견에 대한 언어 네트워크 결과는 (Fig. 6)과 같다. 사고 발생 시, 자동차 결함이 원인이 될 가능성이 높고 그렇다면 회사는 안전한 서비스를 보장할 의무를 이행하지 못한 것이므로 책임이 있다는 것이다. 만일 제조사에게 책임을 묻지 않는다면 검증이 제대로 되지 않은 제품을 판매할 수 있다고 주장했다. 실제로 2016년 테슬라의 자율주행자동차 탑승자 사망 사고가 있었던 것을 언급하기도 했다. 또한 차량 탑승자는 자율주행기능을 사용여부를 결정할 뿐, 내부 알고리즘을 통제할 힘이 없다는 의견도 있었다. 이 경우는 ‘기술’ 노드가 중요한 위치를 차지하고 있는데, 자율주행 기술을 구현하는 회사의 책임이라는 의견으로 해석된다.



(Fig. 6) Responsibility for self-driving car accidents - Automobile Company

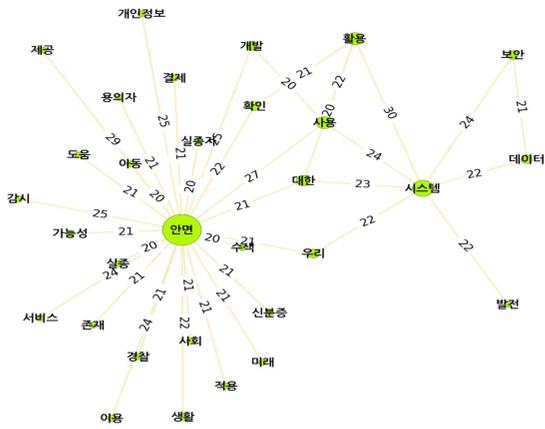
4.4. ‘얼굴 인식이 보편화된 미래’ 분석 결과

‘얼굴 인식이 보편화된 미래’에 대한 토론 주제에 대한 학생들의 의견은 긍정, 부정, 중립으로 구분하였다. 분석된 의견 수와 비율은 <Table 8>과 같다.

<Table 8> The Number of Data Used for Analysis

	Positive	Negative	Neutral	Total
The number of opinions(n)	15	19	44	78
Rate(%)	19.2	24.4	56.4	100.0

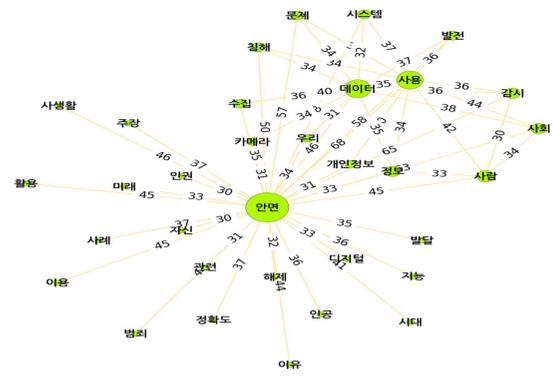
긍정적인 의견에 대한 언어 네트워크 결과는 (Fig. 10)과 같다. 안면 인식을 통해 실종 아동을 쉽게 찾는 사례, 범죄 용의자를 5만 명 규모의 콘서트에서 체포할 수 있었던 사례가 다루어졌다. 또한 얼굴 자체가 ‘신분증’이므로 결제가 편리해질 수 있다는 이야기가 있었다.



(Fig. 10) What is the future of facial recognition technology? - Positive

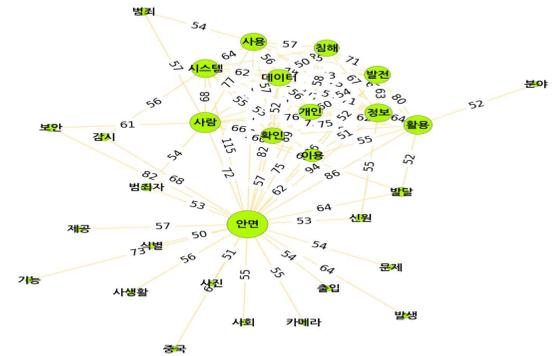
부정적인 의견에 대한 언어 네트워크 결과는 (Fig. 11)과 같다. 얼굴은 쉽게 노출될 수 있는 개인정보이기 때문에 무고한 사람이 얼굴 정보를 도용당해서 범죄 용의자로 지목될 가능성이 있는 것이 그 이유였다. 또한 딥페이크 기술로 인해 지인 능욕과 같은 사례가 더욱 빈번하게 발생할 수 있다고 주장했다. 현재의 안면인식 기술은 백인 남성의 인식 정확도만 99%로 가장 정확하고, 흑인이고 여성일수록 정확도가 더 떨어진다는 근거

도 있었다. 긍정적인 의견에 비해 ‘감시’, ‘사생활’, ‘인권’, ‘침해’, ‘범죄’와 같은 부정적인 단어들도 등장하는 것을 볼 수 있었다.



(Fig. 11) What is the future of facial recognition technology? - Negative

중립적인 의견에 대한 언어 네트워크 결과는 (Fig. 12)와 같다. 중립적인 의견을 보인 학생들의 네트워크에서는 ‘활동’과 ‘이용’, ‘식별’, ‘신원’노드가 긍정적 의견과 마찬가지로 큰 부분을 차지하고 있음을 볼 수 있었다. 범죄자 검거에 큰 도움이 된다는 의견도 많았지만 ‘침해’, ‘범죄’, ‘감시’노드가 긍정적 네트워크 대비 새로 생겼거나 크기가 커진 것을 볼 수 있었다. 즉, 실생활 활용을 위해서는 부정적인 측면도 함께 고려하여야 하며 문제를 최소화할 수 있는 제도적 준비가 우선되어야 함을 보여준다.

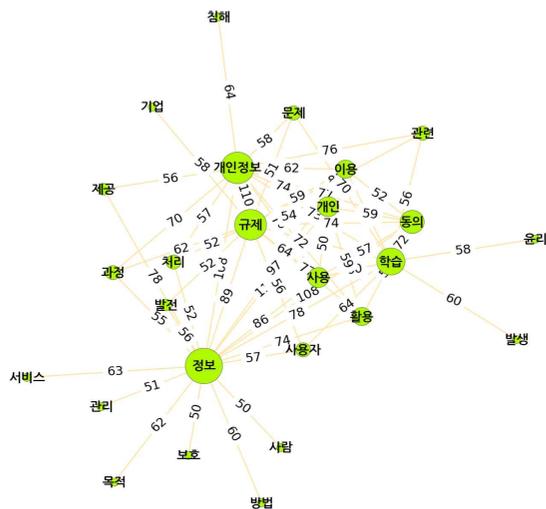


(Fig. 12) What is the future of facial recognition technology? - Neutral

4.5. ‘데이터 수집 규제와 인공지능 기술 활용 방안’ 분석 결과

‘데이터 수집 규제와 인공지능 기술 활용 방안’ 주제는 데이터 수집 과정에서 데이터 질 관리 보장에 대한 문제와 인공지능 기술의 도덕적 활용 방안에 대한 의견을 수집하였다.

데이터 수집에 필요한 규제 및 데이터 질 보장을 위한 방안 의견 분석 결과는 (Fig. 13)과 같다.

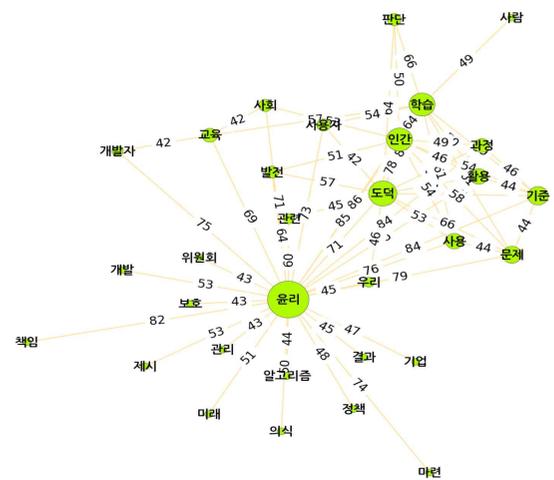


(Fig. 13) What are the regulations necessary for data collection and ethical use of artificial intelligence technology? - Collect good quality data

데이터 수집과 관련된 언어 네트워크에서는 ‘개인정보’, ‘정보’, ‘규제’, ‘학습’, ‘이용’, ‘사용’, ‘개인’, ‘문제’ 등의 노드가 크게 나타나 있었다. 데이터 수집 과정에서 발생할 수 있는 윤리적인 이슈들이 ‘개인정보’와 ‘관련’된 내용이라고 생각함을 알 수 있었다. ‘정보’는 ‘개인정보’를 포함한 상위 개념으로 ‘목적’에 맞는 정보를 수집해야 하며 수집한 정보를 잘 ‘관리’하고 ‘보호’해서 적절한 ‘서비스’를 제공해야 한다는 의견에 등장했다. 개인정보와 정보 모두 ‘규제’와 연결되어 있는데 특히 ‘개인정보’와 ‘규제’ 사이의 가중치가 110으로 데이터 수집 언어 네트워크 중 가장 큰 가중치를 가지고 있음을 확인할 수 있었다. 여기서 규제란 개인정보나 정보를 ‘이용’ 혹

은 ‘사용’할 때 필요한 제재로 ‘동의’를 반드시 받을 것에 대해 언급하는 의견이 많았다. 또한 ‘학습’과 ‘윤리’의 연결을 통해 데이터 학습 과정의 ‘윤리’적 문제 발생 가능성을 학생들이 잘 인식하고 있음을 볼 수 있었다.

도덕적인 기준이 없는 인공지능을 어떻게 활용할 수 있을지에 대한 언어 네트워크 결과는 (Fig. 14)와 같다.



(Fig. 14) What are the regulations necessary for data collection and ethical use of artificial intelligence technology? - AMA

도덕과 유사한 단어인 ‘윤리’가 네트워크에서 큰 비중으로 중요한 위치를 차지하고 있었다. ‘개발자’ (혹은 ‘개발’)와 ‘사용자’ 모두 윤리 ‘교육’이 필요하다는 의견이 높은 비중으로 나타났다. 각각의 노드가 ‘윤리’와 연결된 가중치는 75, 73으로 해당 네트워크의 가장 큰 가중치가 86인 것을 감안하면 상당히 큰 수치이다. 편향되지 않은 데이터를 ‘활용’ 혹은 ‘사용’한 ‘학습’ ‘과정’에서 ‘윤리’적 ‘문제’가 발생할 수 있으므로 인공지능의 행동을 그대로 받아들이지 말고 ‘사람’이 잘 ‘판단’해서 활용해야 한다는 의견도 있었다. 자체적으로 도덕적 기준이 없는 인공지능을 어떻게 활용할 것인지에 대해 ‘책임’, ‘의식’, ‘정책’, ‘관리’, ‘알고리즘’, ‘위원회’ 등 사람 측면에서의 책임 의식, 정부 측면에서의 적절한 정책, 기업 측면에서의 윤리적 알고리즘 개발, 윤리위원회의 설치 등이 대안으로 제시된 것을 볼 수 있었다. 즉, 인공지능의 올바른 사용을 위해서 인공지능 윤리 교육을 모두가 받아야 하

는 것과 동시에 정책적으로, 기업 내부적으로도 노력이 필요하다고 생각하는 것으로 나타났다.

해당 토의에서도 인공지능의 윤리적 지위에 대한 사회적 합의가 아직 이루어지지 않았음을 확인할 수 있었다. 이는 인공지능의 도덕적 판단에 대한 연구가 더욱 필요함을 시사한다.

5. 결론

본 연구는 대학 교양 수준의 인공지능 윤리 교육을 위하여 인공지능 윤리 주제에 대한 토론을 통해 학생들의 윤리적 문제에 대한 인식 수준을 분석하였다. 분석 결과는 다음과 같다.

먼저 인공지능 사회의 미래에 대해서는 62.3%가 긍정적인 의견을 보였다. 인공지능의 발전으로 우리가 그것을 잘 활용하면 더 편리한 사회가 될 것이라는 의견이었다. 중립의 경우 16.9%를 차지했다. 20.8%가 부정적인 의견을 보였는데, 이 경우는 인공지능 수업 초반에 이뤄진 토론으로 배경지식 부족으로 인한 결과라고 분석된다.

자율주행 자동차 간의 사고 책임은 39.2%의 학생이 차량 소유자(탑승자)의 책임이라는 의견이었다. 현재의 자율주행 수준으로는 운전자가 반드시 개입을 해야 하기 때문이라는 것이 주요 골자였다. 43%의 학생이 중립적인 의견을 보였다. 17.8%가 제조사 책임으로 보았는데 다른 주제들에 비해 의견이 많이 갈리는 것을 볼 수 있었고, AMA 지위에 대한 사회적 합의 부족을 확인할 수 있었다.

인공지능 발전의 역기능은 사생활 침해, 기술 악용, 개인화 서비스로 인한 디지털 편식을 꼽았다. 디지털 편식의 경우 사용자가 그동안 봐왔던 콘텐츠와 유사한 내용만 계속해서 추천하기 때문에 다른 의견이나 다른 내용에 대해서 접할 기회가 없어져 사고가 좁아져서 일부 내용만으로 사실이라고 결정해버릴 수 있는 위험성이 있다고 지적했다. 역기능 최소화 방안으로는 인공지능을 사용하는 사람들의 윤리 교육이 필요하며 관련된 제도 마련을 촉구하기도 했다.

얼굴 인식이 보편화된 미래에 대해 19.2%가 긍정적인 의견을 보여주었다. 안면 인식을 통해 실종자를 쉽게 찾은 사례, 범죄 용의자를 찾아낸 사례, 본인 인증이나

결제 시 편리할 수 있다는 것이 그 내용이었다. 56.4%의 학생들은 중립적인 의견을 보였는데 좋은 기술인 것은 맞지만 실생활에서 활용하기 위해서는 발생 가능한 부작용을 함께 고려해서 최소화할 수 있는 제도적인 준비가 필요하다고 답했다. 24.4%의 학생이 부정적인 의견을 보였다. 얼굴은 외부에 드러나 있기 때문에 이 정보가 노출되거나 오용되면 심각한 문제가 발생할 수 있다는 의견이었다. 인식 정확도에 대한 우려도 있었는데 백인 남성 99%의 정확도를 보였으나 흑인이고 여성일수록 정확도가 더 떨어진다는 근거를 함께 제시하였다. 기술이 정확히 윤리적으로 작동한다는 가정 하에 긍정적인 의견을 보였던 것으로 보이며, 부정적 의견을 보인 경우는 기술의 부정확성(신뢰성)과 윤리적인 사용이 보장되지 않을 경우임을 볼 수 있었다. 즉, 신뢰할만한 기술인지, 역기능에 대한 방안이 무엇인지 함께 고민해보아야 함을 시사한다.

마지막으로 데이터 수집에 필요한 규제는 무엇이 있으며 데이터의 질은 어떻게 보장할지, 그리고 도덕적인 인공지능 기술의 활용 방안은 무엇이 있을지에 대해서는 상당히 다양한 의견을 볼 수 있었다. 공통적으로 데이터 수집 시 개인정보가 포함될 경우에는 반드시 동의를 얻은 부분만 활용해야 한다고 답했다. 또한 데이터 수집 시 과도한 규제보다 정부가 가이드라인을 주거나 독립 기관을 통해 운영되는 것이 좋다는 의견도 있었다. 반면에 어떤 규제도 크게 효과가 없을 것이라는 내용도 있었다. 데이터의 품질 보장과 관련하여 상당히 다양한 의견이 있었다. 수집부터 데이터를 선별해서 가져오거나, 인공지능이 사용자 반응을 학습하는 특성을 이용한 피드백 시스템을 도입하자는 의견도 있었다. 만약 데이터 수집 과정을 정부가 통제한다면 데이터의 질은 정부가 관리해야 한다는 의견도 있었으며 기업에서 데이터의 질을 검증하는 전담 인력을 고용해야 한다는 의견도 있었다. 다른 의견으로는 인공지능의 개발 단계부터 다양한 국적과 인종, 성별의 개발자들로 팀을 구성하여 비관적으로 데이터를 바라보고 수집해야 한다는 의견도 있었다. 데이터가 많아지면 문제가 되는 데이터가 자연스럽게 걸러져서 데이터의 질 또한 좋아질 것이라는 의견도 있었다. 도덕적인 기준이 없는 인공지능을 어떻게 사용할 수 있을지에 대해서 근본적으로 ‘사용자’가 인공지능 활용에 필요한 윤리적 소양을 갖추는 것이 필요하다는

의견이 있었다. 또 다른 의견으로는 일정 수준의 윤리적 소양을 갖춘 사람만 인공지능을 개발할 수 있도록 하자는 의견도 있었다. 편향되어 있지 않은 데이터만으로 학습을 시키는 과정에서 또 다른 편향이 발생할 수 있으니, 데이터는 그대로 두고 인공지능의 행동을 ‘사람’이 잘 판단하고 활용하면 된다는 의견도 있었다. 공통적으로 인공지능을 사용하는 사람이나 개발하는 사람 모두 인공지능 윤리를 교육받아야 한다는 의견이었다. 사람이 아닌 인공지능의 윤리에 초점을 둔 의견도 있었다. 문화권마다 도덕적 기준이 다르므로 인공지능에게 윤리를 학습시키는 것은 불가능하다는 의견도 있었다. 반대로 인공지능에게 특정 상황에서 어떻게 판단하는 것이 윤리적인지를 가르쳐야 한다는 의견도 있었다. 즉, 데이터의 수집, 품질 관리, 인공지능의 도덕성 부재와 관련해서 사회적인 합의가 필요한 부분임을 보여준다.

대학생들의 인공지능 윤리에 대한 인식 수준을 분석한 결과를 토대로 대학 인공지능 교육에서 고려해야 할 시사점은 다음과 같다. 첫째, 인공지능 사회를 살아가면서 갖추어야 할 기술적, 윤리적 소양을 대학 교양 교육에서 다루어야 한다. 교양 교육에서 대부분의 학생들은 인공지능 비전공자이지만, 사용자 관점뿐만 아니라 개발자 및 공급자 관점에서도 윤리적 문제를 접근할 필요가 있다. 인공지능의 역기능 최소화 방안과 도덕적인 기준이 없는 인공지능의 활용 방안에 대해서는 많은 학생들이 ‘윤리’의 중요성을 강조하였다. 따라서 인공지능 자체에 대한 이해와 윤리 교육이 모두 포함되어야 한다.

둘째, 대학 교양 수준의 인공지능 윤리 교육에서는 딜레마 상황을 제시하여 학생들이 스스로 윤리적 문제를 인식하고 다양한 관점에서 생각해볼 수 있는 기회를 제공해야 한다. 학생들은 하나의 인공지능 윤리 문제에 대해서 기술적 접근과 함께 인문학적 소양을 바탕으로 도덕적 판단의 근거를 제시할 수 있어야 하며, 이런 기회를 통해 새롭게 등장하는 인공지능이나 첨단 기술에 따른 윤리적 문제를 인식하고 판단할 수 있는 능력을 키워야 한다.

본 연구의 결과를 토대로 대학 교양 수준의 인공지능 윤리 교육을 설계할 때 현재 대학생들의 인식 수준에 따른 교육을 설계할 수 있을 것으로 기대한다.

참고문헌

- [1] Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J. F. & Rahwan, I. (2018). The Moral Machine experiment, *Nature*, 563(7729), 59-64.
- [2] Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, 1, 316-334.
- [3] Byun, S. (2018). A Study on the Ethical Guideline for artificial intelligence robots - Focusing on 4 principles of artificial intelligence robot ethics -. *Journal of Ethics Education Studies*, 47, 233-252.
- [4] Byun, S. (2020). A Study on the Necessity of AI Ethics Education, *Korean Journal of Elementary Education*, 31(3), 153-164.
- [5] Choi, H. & Byun, S. (2019). A Philosophical Project of Hybrid Approach to Artificial Moral Agents. *Journal of Ethics: The Korean Association of Ethics*, 1(124), 1-16.
- [6] Choi, H. C., Byun, S. & Shin, H. J. (2016). The Ethical Principles for the Development of Artificial Moral Agent - Focusing on the Top-down Approach -, *Journal of Ethics*, 111, 31-53.
- [7] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, (2020-1).
- [8] Gianmarco, V., & Fiorella, O. (2008). Roboethics: Social and ethical implications of robotics. *Springer Handbook of Robotics*, 1499-1524.
- [9] Google AI site by Google. <https://ai.google/principles/>
- [10] Holmes, W., Bektik, D., Whitelock, D. and Woolf, B.P.(2018). Ethics in AIED: Who Cares?. AI in Education *LNAI 10948*. 551-553.
- [11] Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat Mach Intell* 1, 389 - 399.

- [12] Jung, C. (2018). Building an Ethical Intelligent Robot in the Intelligent Information Society: Alternative Approaches towards a Postmodern Reconstruction of Ethics. *Northeast Asian law journal*, 12(1), 87-121.
- [13] Kakao AI ethics site by Kakao. <https://www.kakaocorp.com/page/responsible/detail/algorithm/>
- [14] Kim, H. (2019). AI and Ethics. CommunicationBooks.
- [15] Kim, H. (2020). Education of AI Ethics as a Response to Decision Automation. *Journal of Ethics Education Studies*, 55, 277-308.
- [16] Kim, H. S. (2019). Exploring Information Ethics Issues based on Text Mining using Big Data from Web of Science. *The Journal of Korean association of computer education*, 22(3), 67-78.
- [17] Kim, N., Lee, D., Choi, H., & Wong, W. X. S. (2017). Investigations on techniques and applications of text analytics. *The Journal of Korean Institute of Communications and Information Sciences*, 42(2), 471-492.
- [18] Ko, I. S. (2011). A Reappraisal of Asimov's Three Laws of Robotics as Basic Principles for Roboethics, *Journal of The Society of philosophical studies*, 93, 97-120.
- [19] Lee, H. (2021). AI Iruda, who suffered from sexual harassment, is worried about homophobic learning. <https://www.yna.co.kr/view/AKR20210109055000017>
- [20] Lee, J. (2021). Artificial intelligence technology criticism. CommunicationBooks.
- [21] Lee, J. (2021). 'Iruda Incident' litigation begins... 'Pacific vs Taerim', Tae family fight 'catch public's eyes'. https://biz.chosun.com/topic/s/law_firm/2021/05/09/FGX3WISL7FBENFANY2SXGL0U7Y/
- [22] McCarthy, J., Minsky, M., Shannon, C. E., Rochester, N., & Dartmouth College. (1955). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence.
- [23] Ministry of Science and ICT. (2020). Artificial Intelligence (AI) Ethics Standards, Sejong, Ministry of Science and ICT.
- [24] Moor J. H.(1985). What is Computer Ethics?, *Metaphilosophy*, 16(4), 266-275.
- [25] Moral Machine site by MIT Media Lab. <https://www.moralmachine.net/>
- [26] National Information Society Agency. (2018). Intelligent Information Society Ethics Guidelines, Daegu, National Information Society Agency.
- [27] National Science and Technology Council. (2016). Preparing for the Future of Artificial Intelligence, Washington, D.C., National Science and Technology Council.
- [28] Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics| talking to Bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, 10, 4915-4931.
- [29] Park, J. (2021). Needs Analysis of AI Education in Liberal Arts: Using IPA, The Korean Association Of Computer Education, 23(2), 75-84.
- [30] Russell, S., & Norvig, P. (1994). Artificial intelligence: a modern approach, New Jersey: Prentice Hall.
- [31] Samuel, A. L. (1960). Some moral and technical consequences of automation—a refutation. *Science*, 132(3429), 741-742.
- [32] Shin, S. (2017). Can Artificial Intelligence be an Autonomous Moral Agent?, *Philosophy*, 132, 265-292.
- [33] Turing, A. M., & Haugeland, J. (1950). Computing machinery and intelligence (pp. 29-56). Cambridge, MA: MIT Press.
- [34] Yutaka, M. (2015). Does Artificial Intelligence surpass humans? – Things beyond deep learning, Tokyo: KADOKAWA.

저자소개



유 수 진

2010 고려대학교 사범대학
컴퓨터교육과(이학사)
2012 고려대학교 일반대학원
컴퓨터교육학과(이학석사)
2020 고려대학교 일반대학원
컴퓨터학과(공학박사)
2021~현재 한양대학교 소프트웨
어융합원 SW교육전담교수
관심분야: 인공지능 윤리, 인공지
능 교육과정, 소프트웨어 교
육, 링크드 데이터, 시맨틱웹
e-mail: sujinyoo@hanyang.ac.kr



장 윤 재

2008 고려대학교 사범대학
컴퓨터교육과(이학사)
2012 고려대학교 일반대학원
컴퓨터교육학과(이학석사)
2018 고려대학교 일반대학원
컴퓨터교육학과(이학박사)
2018~2021 고려대학교
정보창의교육연구소 연구교수
2021~현재 경기대학교 SW중심
대학사업단 연구교수
관심분야: 정보·컴퓨팅 교육, 인공
지능 교육, 프로그래밍 교육
e-mail: janggoons@kyonggi.ac.kr