

# 인공지능기술 윤리성 인식 척도개발 연구

김도연<sup>1\*</sup>, 고영화<sup>2</sup>

<sup>1</sup>서울대학교 사범대학 협동과정 환경교육전공 박사과정,  
<sup>2</sup>연세대학교 일반대학원 융합기술경영공학과 박사

## Development and Validation of Ethical Awareness Scale for AI Technology

Doeyon Kim<sup>1\*</sup>, Younghwa Ko<sup>2</sup>

<sup>1</sup>Ph D candidate, Graduate School of Education, Environmental Education, Seoul National University

<sup>2</sup>Ph. D, Graduate School of Convergence Technology Management Engineering, Yonsei University

요 약 본 연구의 목적은 인공지능 기술 또는 서비스를 수용하는 사용자의 윤리성 인식을 측정하기 위한 척도 개발 및 타당화에 있다. 이를 위해 인공지능 윤리성 관련 문헌 분석을 통해 구성개념 및 속성을 확인하였다. 전국의 10대–70대 남녀 133명(개방형 설문:1차 문항), 273명(예비조사:2차 문항), 500명(본조사:최종 문항)을 대상으로 실시한 온라인 설문조사 결과를 확인적 요인분석에 의해 정제하여 최종적으로 인공지능기술 윤리성 척도를 개발하였다. 인공지능기술 윤리성 인식 척도는 총 4개 요인(투명성, 안전성, 공정성, 책임성) 16개 문항으로 개발하여 일반적인 인공지능기술 관련 윤리성 인식을 세부 요인별로 측정할 수 있도록 하였다. 개발된 척도를 활용하여 다양한 분야의 측정 변인들과의 관련성을 밝힐 수 있을 것이며, 인공지능기술 발전의 초기 단계에서 윤리성 인식을 높이기 위한 기초 데이터를 제공하는 데 중요한 역할을 할 것으로 기대한다.

주제어 : 인공지능, 윤리성 인식, 투명성, 안전성, 공정성, 책임성

Abstract The purpose of this study is to develop and validate a scale to measure the ethical awareness of users who accept artificial intelligence technology or service. To this end, the constructs and properties of AI ethics were identified through literature analysis on AI ethics. Reliability and validity were assessed through a preliminary survey(N=273), after conducting an open-type survey to men and women(N=133) in 10s to 70s nationwide, extracting the first questions, and reviewing them by experts. The results of an online survey conducted on men and women(N=500) were refined by confirmatory factor analysis. Finally, an AI technology ethics scale was developed. The AI technology ethics awareness scale was developed with 16 questions in total of 4 factors (transparency, safety, fairness, accountability) so that general awareness of ethics related to AI technology can be measured by detailed factors. In addition, through follow-up research, it will be possible to reveal the relationship with measurement variables in various fields by using the ethical awareness scale of artificial intelligence technology.

Key Words : Artificial Intelligence, Ethical Awareness, Transparency, Fairness, Safety, Responsibility

\*Corresponding Author : Doeyon Kim(gould55@snu.ac.kr)

Received November 16, 2021

Accepted January 20, 2022

Revised January 7, 2022

Published January 28, 2022

## 1. 서론

4차 산업혁명의 가속화와 함께 인공지능(Artificial Intelligence: AI) 기술의 적용 범위가 빠르게 확대되고 있다. 인공지능기술 발전에 따라 다양한 인공지능제품과 서비스 이용에 수반되는 윤리성에 관한 관심도 점차 높아지고 있다. 인공지능기술 개발이 시작되던 2010년대 후반 무렵에는 주로 인공지능으로 인해 발생할 수 있는 윤리적 문제의 책임이 인공지능에게 있는 것인가의 질문이 인공지능기술을 개발하고 적용하는 개발자에게 윤리적 책임을 묻는 방향으로 옮겨가기 시작했다. 따라서 Microsoft와 같은 기업에서 2017년 무렵 인공지능 개발 인력이 준수해야 할 AI 가이드라인을 제시했다[1].

그러나 이보다 앞서, 2007년 한국에서 ‘로봇윤리현장’의 초안을 만들어 인공지능에 관련된 일반적 원칙을 제시했으며 로봇산업진흥원에서 ‘인공지능로봇에 대한 윤리 가이드라인’을 만들었다. 기업에서는 2018년 카카오가 ‘알고리즘 윤리현장’을 만들기도 했다[2].

인공지능기술은 1990년대에 주로 온라인 쇼핑 플랫폼 중심의 기술 적용이 이루어졌다. 1994년 아마존의 온라인쇼핑몰 사업이 시작되었고, 2000년대에 들어와 검색과 소셜네트워크 분야로 인공지능기술의 확장이 이루어졌다. 2005년에는 구글맵(Google Maps)서비스가 시작되면서 온라인 네비게이션이 대중화되기 시작했다. 2008년에 페이스북(Facebook)의 소셜네트워크서비스(SNS: Social Network Services)가 시작되었다. 2010년대에는 넷플릭스와 같은 비디오 스트리밍(video streaming) 서비스를 통해서 이전의 CD나 DVD대여를 대체하는 양상을 보인다. 2014년에 인공지능 기술의 본격적 적용 사례로 볼 수 있는 아마존 에코(Echo)와 알렉사(Alexa)와 같은 음성보조 서비스가 시작되었다. 2016년에는 스포티파이(Spotify)와 같은 음원 스트리밍 서비스가 시작되었고 구글이 인공지능 언어번역 서비스를 시작하였다. 2016년 삼성의 갤럭시 노트7 휴대폰에 안면 인식 기술 탑재와 함께 인공지능 기술이 사용자가 일상에서 직접 체험하고 느낄 수 있는 범위에 등장하였다. 2020년대의 인공지능기술은 더욱 현실 적용이 강화되는 양상을 보이는데, 아마존은 스마트홈 기술(smart home tech)을 도입하여 초인종 사용자의 비디오 검토 인력을 감축하였으며, 틱톡(TikTok) 앱이 상용화되고, 셀프카메라(Selfie photo)앱 기술로 진화되었다[3,4].

이와 같이 인공지능기술은 가정용 전자기기, 자율주행 자동차, 휴대용 기기, 언어관련 서비스, 의료서비스, 쇼핑

등 일상적인 생활 환경 내의 다양한 영역에서 활용되고 있다. 이렇게 인공지능 기술이 빠르게 발전하고 확산함에 따라 인공지능 기술의 편리하고 긍정적인 측면 뿐만 아니라 그로 인해 발생할 수 있는 사회적 문제와 윤리적 측면을 간과할 우려에 대한 문제가 제기되고 있다. 근본적으로 인공지능기술의 불완전성과 축적된 데이터의 부족, 또는 예측하지 못한 학습의 오류로부터 발생할 수 있는 문제점으로부터 인공지능의 윤리성 문제는 출발한다. 동시에, 인공지능기술에 의한 결과물이 왜곡과 편향을 내포할 수 있다는 점 또한 윤리성 이슈를 불러일으키는 요인이다. 요컨대, 인공지능기술을 개발하고 관리하며 사용하는 사람이 간과할 수 있는 부분과 인공지능기술 자체가 항상 옳은 판단을 하는 것은 아니라는 부분이 인공지능 기술에 대한 윤리성 고려 요구를 높이는 근본 원인이 될 수 있다[5].

인공지능기술의 윤리성 고려에 대한 필요성은 높아지고 있으나, 현재 인공지능 윤리성에 대한 연구는 매우 초기 단계이므로 연구가 거의 이루어지지 않은 상황이다[15]. 정부 및 기업 차원의 ‘인공지능 윤리 가이드라인’이 존재한다고 해도 인공지능기술이 적용된 제품과 서비스를 이용하는 사용자가 인공지능 윤리에 대한 개념과 인식을 가지고 있지 않다면 윤리성에 대한 문제제기가 이루어지기 어렵기 때문에 인공지능 윤리성 문제가 근본적으로 해결되기 어려울 것이다. 따라서 사용자가 인공지능 윤리성 부분에 대한 인식 정도와 인공지능 기술이 적용된 제품과 서비스의 이용 여부 결정 시에 윤리성을 판단의 기준에 포함시키는지를 측정하여 인공지능 연구와 기술개발에 반영하는 것은 매우 중요한 과정이 될 수 있다.

이에 본 연구에서는 인공지능기술의 윤리적 사용에 대한 인식을 판단할 수 있는 근거로서 양적 측정도구인 인공지능기술 윤리성 인식 척도를 개발하여 인공지능기술이 윤리적 측면에서 올바른 방향으로 사용될 수 있는 토대를 마련하고자 한다.

## 2. 이론적 고찰

### 2.1 인공지능윤리

윤리는 타자를 향해서 자신이 어떻게 행동해야 하는가를 알려주는 규범적 실행원칙이다[6]. 즉 윤리는 개인과 사회에서 일어나는 모든 행동에 대한 기본 가치영역이며 무엇이 옳고 그른지에 대한 판단에 기준을 제시하는 원칙 또는 지침이다[22]. 기본적인 윤리 원칙이 인공지능

기술에 응용되어 나타난 것이 인공지능 윤리로서 인공지능 개발자의 책임임과 동시에 인공지능이 수행하는 작업 자체의 윤리적 의무로 개념이 확대되고 있다. 인공지능기술의 발전과 함께 이러한 윤리적 원칙을 적용하는 대상은 인간뿐만 아니라 동물, 과학, 기업, 로봇 등 그 범위가 매우 다양해지고 있다[7,22].

인공지능 윤리는 감정적 측면, 법적 측면, 공정성 측면으로 윤리적 기준의 틀을 나누거나, 정의성 측면, 실용주의적 측면, 공공선 측면, 도덕적 측면 등으로 분류하기도 한다[6,7]. 특화된 인공지능 적용 분야로 분류하여 로봇 윤리(robo-ethics), 기계윤리(machine ethics)로 나누기도 한다. 로봇윤리는 인간이 인공지능이 적용된 기계에 대해 적용하는 윤리적 의무의 개념이며, 기계윤리는 인공지능이 적용된 기계나 서비스가 지켜야 하는 윤리성의 개념으로 알려져 있다. 기계윤리는 인공지능기술을 적용한 기계가 작업을 윤리적으로 수행하면서 윤리적 가치를 실현해야 한다는 의미를 내포한다. 다시 말해서, 인공지능기술이 적용된 로봇이나 기계는 인간에 해를 입혀서는 안되며, 이러한 윤리를 기본으로 하여 인간의 명령에 따라야 하고, 다음으로는 로봇이나 기계 자신의 존재를 보호해야 한다는 뜻이다[7]. 요컨대, 인공지능 윤리는 인공지능 기술의 설계와 사용에 있어서 인간의 행동 방향을 제시하고 제어하기 위한 최소한의 장치로 정의할 수 있다[22].

이와 같이, 인공지능 윤리에 대한 논의가 진행되면서 인공지능기술을 개발하고 제품과 서비스에 적용하고 있는 기업과 그 기업들을 총괄하는 국제기구 및 각국 정부는 인공지능 개발에 대한 가이드라인을 제시하고 있다. 2019년 EU는 '신뢰할 수 있는 인공지능 윤리 가이드라인'을 발표했고, OECD 역시 'OECD Digital Economy Outlook 2017' 보고서를 발표하여 인공지능 기술에 대한 대응 규범과 정책 방향을 제시하였으며 2019년에 인공지능 기술개발에 대한 기본 권고안을 발표했다. 위에서 언급한 지점에서 공통적으로 나타나는 윤리성 관련 내용은, 인공지능에 의한 의사결정 과정의 모니터링과 투명성의 제고, 차별적 대응을 방지하는 인공지능 기술의 공정성과 인공지능기술에 의해 발생할 수 있는 결과들에 대한 책임성 및 안전성에 관련된 윤리적 기준을 제시하고 있다[1,14].

인공지능에 대한 윤리적 기준은 사용되는 인공지능 알고리즘의 종류에 따라 달라질 수 있다. 예를 들어, 의사결정나무(decision tree)에 의해 문제 해결에 도달하는 알고리즘은 그 과정이 투명하다. 그리고 그것을 사용하는

사람이 최종적으로 무엇을 선택하느냐에 따라 책임이 부과되며, 인공지능 개발자는 결정과정을 설명할 수 있으므로 기술 개발에 대한 책임을 질 수 있다[17]. 즉 인공지능 윤리의 적용 기준과 경계가 명확하다. 반면에 신경망(neural networks) 알고리즘을 사용한 머신러닝 또는 딥러닝을 수행할 경우, 이에 의한 의사결정 과정을 설명할 수가 없기 때문에[19] 투명하지 않으며 누구의 책임인지가 불분명해진다. 인공지능이 무엇을 수행하는지는 알 수 있지만 최종적인 결정사항이 어떤 과정에 의해 이루어졌는지 설명하기가 어려워 인공지능 윤리의 적용 기준과 경계가 모호해진다[17,18]. 인공지능의 이러한 특성 때문에 인공지능 윤리성에 대한 논의가 반드시 필요한 것이다.

인공지능의 안전성과 보안에 대한 부분은 그 모호함과 위험성이 가시적이다. 즉 프로그램된 절차대로만 학습하는 인공지능이 인간을 포함한 생명체의 가치와 자유를 지키기 위해[21] 상황을 정확하게 인지하고 그에 대처할 수 있는가의 문제[17]와 관련되며 이 또한 책임의 문제를 불러일으킬 수 있다. 보안과 관련한 안전성 문제는 더욱 큰 이슈에 속한다. 인공지능 기술이 적용된 제품이나 서비스의 코드가 유출되어 하드웨어나 소프트웨어에 문제를 일으키는 경우[17] 그 결과는 상상을 초월할 수 있기 때문이다. 네트워크로 연결되는 사회에서 인공지능을 이용한 모든 장치와 소프트웨어는 유출 위험에 노출되어 있으며 악의적으로 조작될 수 있다는 점에서 인공지능 윤리의 안전성과 책임성이 강조되는 부분이다[17,20].

인공지능이 인간 존엄성과 어떤 관련을 맺는지는 윤리적 공정성의 영역에 해당한다. 인공지능의 공정성은 특정 목적을 위하여 인간을 수단으로 여기지 않으며 인간의 존엄에 대해 충분히 존중함으로써 충족된다. 또한 지구상의 어떤 인간집단에 대해서도 차별없이 상호 간에 평등한 대우를 받도록 배려하는 태도가 포함된다[22].

위와 같이, 인공지능 윤리에 대한 전반적인 문제점을 기반으로 문헌 검토를 한 결과, 인공지능 윤리성 범주를 언급한 문헌에서 제시된 상위 네 개의 범주와 최다빈도로 언급된 윤리성 범주인 투명성, 공정성, 안전성, 책임성에 대하여 구체적인 내용을 살펴보았다.

### 2.1.1 투명성(transparency)

인공지능기술의 투명성은 시스템이 결정한 사항 자체와 그러한 결정의 이유에 대한 추적과 설명이 가능하며 인공지능 시스템 전 과정의 내용을 검증하고 해석할 수 있어야 한다는 원칙이다[1]. 투명성은 인공지능 가이드라

인에서 제시하는 가장 상식적이고 일반적인 원칙이며 투명성의 세부 영역과 내용의 해석 및 적용 영역은 다양하게 나타나고 있다. 현재까지 제시된 인공지능 가이드라인 상의 투명성 관련 세부 내용은 의사소통(communication), 공개성(disclosure), 외현성(explicability), 설명 용이성(explainability), 이해 용이성(understandability), 해석 용이성(interpretability) 등이 대표적이다.

기본적으로 투명성은 인공지능의 기능을 긍정적인 방향으로 개선함으로써 부정적 측면을 감소시키는 역할을 할 수 있으며, 인공지능기술에 대한 신뢰 증진과 법적 요건을 충족시켜주는 방법이기도 하다. 또한 투명성은 사용자와 개발자 간의 소통(대화, 참여)을 중시하므로 기본적으로 민주주의 원칙과 관련되기도 한다[4,8].

### 2.1.2 공정성(fairness)

인공지능기술에 있어서 공정성은 정의와 평등 개념을 내포하고 있다. 다시 말해서, 의도하지 않은 왜곡(unwanted bias) 및 차별(discrimination)의 회피, 위험의 예방(prevention), 모니터링(monitoring) 개념을 포함하며 다양성(diversity)에 대한 존중과 회복(remedy)의 의미도 가진다[8]. 인공지능기술의 공정성은 인간존엄을 존중하며 인공지능이 가진 효율성을 극대화하는 것을 목표로 하며 다양한 형태의 왜곡으로부터 사용자를 보호해야 함을 의미한다. 동시에 불공정한 상황을 방관하거나 강요하는 상황은 개선되어야 한다[1]. 실제로 인공지능 서비스에 있어서 공정성의 적용은 성별, 연령, 인종, 언어 등의 차이에 대한 식별이나 고려가 이루어질 때 확보될 수 있다[4].

### 2.1.3 안전성(safety)

인공지능기술에 있어서 안전성은 악의적인 목적의 개발과 이용(non-maleficence)으로부터 사용자를 보호해야 한다는 개념이다. 다시 말해서, 인공지능기술을 개발하거나 적용할 때 사용자에게 대한 일반적인 안전과 보안 요구를 선제적으로 고려해야 한다는 점이 중요하다. 아울러, 특정 위험이나 잠재적인 해악에 대한 예방도 함께 이루어져야 한다. 인공지능의 안전성에 관련된 해악은 개인의 프라이버시 침해, 유해콘텐츠에 의한 정신적 폭력에의 노출 등을 의미한다. 인공지능 안전성은 인공지능 기술의 발전이 법적 규제 장치보다 선행하여 사용자가 보호받지 못할 위험성을 포함하며, 장기적으로는 사회적 건강성과 심리-정서-경제 측면에서도 부정적 영향이 있을 수 있다

는 전제를 포함한다[8]. 동시에 인공지능기술은 내외부적 조작으로부터 안전하게 설계되어 사이버보안의 측면에서 안전성을 확보하는 것이 중요하다[4].

### 2.1.4 책임성(responsibility)

인공지능기술 이용에 대한 가이드라인에서 가장 중요하게 다루어지는 내용이 책임에 관련된 부분이다. 인공지능 기술 개발과 이용의 책임성에 대한 언급은 현재까지 제시된 거의 모든 가이드라인에 나타나 있으나 그 개념과 정의는 통합적이기보다는 다양한 측면으로 제시되어 있다. 인공지능기술의 책임성 측면은 알고리즘과 관련한 법적인 책임과 인공지능 프로세스의 잠재적 위험성에 대한 책임, 인공지능에 대한 지식을 전달하는 교육분야에서 기술, 공학, 과학, 수학 교과 등에 인공지능 책임성을 교육내용으로 포함하는 방안 등이 제시되고 있다. 인공지능 책임성은 기술을 개발하는 인간의 책임과 인공지능기술이 학습에 의해 얻은 결과에 대해 스스로에게 부과되는 책임성으로 크게 나뉜다. 즉 인공지능기술은 운용 범위가 법적인 책임의 범위에서 주체성을 가져야 하며 인공지능기술에 의해 제공되는 서비스가 인간에 의한 것이 아니라는 점을 공개해야 할 책임과 함께, 정보원의 허용 없이는 민감한 정보를 공개하거나 저장할 수 없다는 내용이 인공지능의 책임성에 포함된다[1,4,6,8].

## 3. 연구 방법

### 3.1 문헌 분석

본 연구의 목적에 따라 인공지능 윤리성에 대한 지침과 아웃라인을 다룬 각종 문헌들을 선정하여 검토하였다. 문헌에서 공통적으로 나타나는 인공지능 윤리성의 범주(Table 1)를 도출하고 각각에 대한 특징과 세부 내용들을 면밀히 분석하였다.

### 3.2 개발 절차

본 연구의 절차는 다음과 같은 단계에 의해 수행되었다(Fig. 1).

#### 3.2.1 개방형 설문과 전문가 검토(1단계)

문헌검토 결과 추출된 범주와 주요 단어들을 중심으로 구성된 개방형 질문은 3개 부분으로 구성되었다. 첫 부분은 인공지능에 대해 응답자가 가지고 있는 지식과 생각

을 워밍업할 수 있는 8개의 문항으로 구성하였고, 둘째 부분은 인공지능기술 윤리성 인식 범주로 선정된 투명성, 공정성, 안전성, 책임성에 관한 질문으로 구성하였다. 세 번째 부분은 응답자의 일반사항이다.

문헌 분석 자료를 토대로 개방형 질문(Table 2)을 구성하여 설문조사를 실시하였다. 2021년 1월에 전국의 10대부터 70대까지의 남녀 201명에게 15개문항(일반사항 제외)의 개방형 설문을 실시하여 총 133부의 유효응답을 수집하였다. 수집한 답변은 모두 진술문 텍스트 형태이므로 내용분석을 통해 응답 단어의 빈도를 산출하고 응답 내용의 맥락을 분석하여 전문가 검토(Table 3)를 거친 후 문항으로 구성하였다.

### 3.2.2 예비조사 및 문항 정제(2단계)

위에서 구성된 문항(Table 4)들을 중심으로 설문조사를 실시하였다. 2021년 4월-5월 전국 10대부터 70대까지의 남녀 300명에게 설문을 실시하여 총 273부의 설문응답을 수집하였다. 설문응답결과는 Likert 5점 척도로 측정되었으며 기술통계분석, 탐색적 요인분석, 상관분석을 실시하여 신뢰도와 타당도를 확인하고 요인을 정제(Table 5, Table 6)하였다.

### 3.2.3 본조사 및 최종문항 개발(3단계)

예비조사 결과에 의해 선정된 요인과 문항을 중심으로 설문조사를 실시하였다. 2021년 6월-9월 전국 10대부터 70대까지의 남녀 500명의 설문응답을 수집하였다. 응답결과는 Likert 5점 척도로 측정되었으며 확인적 요인분석결과에 의해 요인별 범주를 확인하고 정제한 후 최종적으로 척도(Table 7, Table 8, Table 12)를 개발하였다.

## 3.3 분석방법

### 3.3.1 1단계 분석방법

개방형 설문 결과는 내용분석을 실시하기 위해 Excel과 NVivo 프로그램을 이용하여 텍스트 빈도와 공통성을 중심으로 핵심 단어를 추출하였다. 추출한 핵심 단어들은 전문가 검토를 바탕으로 각 구성 속성별로 예비문항을 개발하였다.

### 3.3.2 2단계 분석방법

설문조사 결과를 중심으로 SPSS Statistics 23.0을 이용해 탐색적 요인분석을 수행하였다. 탐색적 요인분석 결

과에 따라 타당도와 요인의 범주를 확인하고 Cronbach's  $\alpha$ 에 의해 신뢰도를 확인하였다.

### 3.3.3 3단계 분석방법

2단계에서 확인된 범주별 설문문항으로 조사를 실시한 결과에 따라 3단계 분석을 수행하였다. 3단계 분석에는 AMOS 23.0을 이용하여 확인적 요인분석을 수행하여 측정문항의 타당성을 평가하고 최종문항을 추출하였다.

## 3.4 설문조사 참여자의 일반적 특성

### 3.4.1 개방형 설문 참여자 특성

성별은 남자 49명(36.8%), 여자 84명(63.2%)이었고 연령별로 10대 4명(3.0%), 20대 20명(15.0%), 30대 23명(17.3%), 40대 34명(25.6%), 50대 39명(29.3%), 60대 10명(7.5%), 70대 3명(2.3%)이 참여하였다. 직업별로는 관리직 10명(7.5%), 전문직 16명(12.0%), 사무직 13명(9.8%), 서비스직 10명(7.5%), 판매직 9명(6.8%), 농림어업 8명(6.0%), 기능직 9명(6.8%), 장치기계 8명(6.0%), 단순노무 7명(5.3%), 군인 7명(5.3%), 학생 11명(8.3%), 주부 25명(18.8%)이 참여하였다. 응답자의 성별 분포는 여성이 더 많았고, 연령대는 40대와 50대가 가장 많았으며 직업군은 대체로 각 직업군에서 고르게 응답했으나 주부 집단의 숫자가 많았다.

### 3.4.2 예비조사 설문 참여자 특성

성별은 남자 88명(32.2%), 여자 185명(67.8%)이었고 연령별로 10대 14명(5.1%), 20대 45명(16.5%), 30대 29명(10.6%), 40대 79명(28.9%), 50대 80명(29.3%), 60대 26명(9.5%)이 참여하였다. 직업별로는 관리직 19명(7.0%), 전문직 31명(11.4%), 사무직 39명(14.3%), 서비스직 13명(4.8%), 판매직 4명(1.5%), 농림어업 4명(1.5%), 기능직 6명(2.2%), 장치기계 2명(0.7%), 단순노무 4명(1.5%), 학생 47명(17.2%), 주부 109명(39.9%)이 참여하였다. 응답자 성별분포는 여성이 두 배 정도 많았고, 연령대는 40대와 50대가 많았다. 직업군은 전문직, 사무직, 학생이 많았고 가장 많은 직업군은 '주부'였다.

### 3.4.3 본조사 설문 참여자 특성

성별은 남자 213명(42.6%), 여자 287명(57.4%)이었고 연령별로 10대 28명(5.6%), 20대 84명(16.8%), 30대 55명(11.0%), 40대 143명(28.6%), 50대 143명(28.6%), 60대 47명(9.4%)이 참여하였다. 직업별로는 관리직 41

명(8.2%), 전문직 62명(12.4%), 사무직 85명(17.0%), 서비스직 35명(7.0%), 판매직 15명(3.0%), 농림어업 10명(2.0%), 기능직 12명(2.4%), 장치기계 4명(0.8%), 단순노무 8명(1.6%), 학생 86명(17.2%), 주부 142명(28.4%)이 참여하였다. 본조사의 참여자 중 성별 분포는 남자가 대략적으로 비슷한 응답비율을 보였고, 연령대는 40대와 50대가 많았다. 직업군은 사무직, 학생, 주부 집단의 응답이 많았다.

위의 설문 참여자 특성에서 나타난 공통점은 성별에서 여성이 남성보다 많고, 직업군에서 주부 응답자가 많다는 것인데, 이것은 인공지능기술이 적용된 제품이 주로 가정에서 사용하는 가전제품이거나 여성들이 사용하는 미용제품에 인공지능기술이 적용된 사례가 많기 때문인 것으로 해석된다.

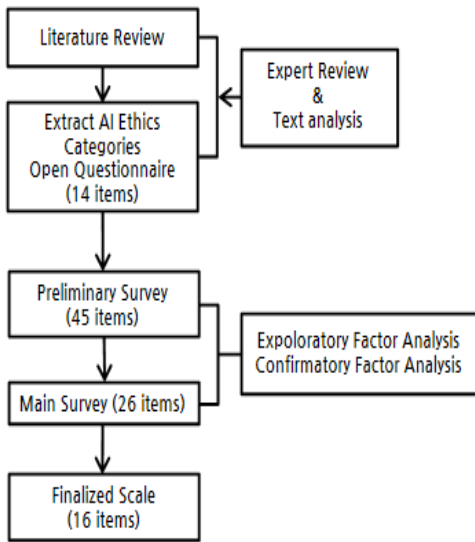


Fig. 1. Research Flow

## 4. 연구 결과

### 4.1 문헌 분석 결과

인공지능기술의 윤리성 인식 범주를 파악하기 위한 문헌 검토는 주로, 인공지능기술을 개발하고 있거나 사용하고 있는 글로벌 IT 기업 또는 정부차원의 가이드라인을 중심으로 이루어졌다. 2016년 이후 서구에서 발표된 인공지능 윤리성 관련 문헌은 80편 이상이다[16]. 2019년

현재 미국, EU, 호주와 영국에서 작성한 ‘인공지능의 윤리적 원칙과 지침’을 포함하는 문헌은 약 84편 정도이며 이 중 88%가 2016년 이후에 발간된 것이었다[6]. 84편의 문헌에서 공통적으로 언급하고 있는 인공지능 윤리성 인식의 주제는 11개의 범주로 나뉜다. 즉 투명성(transparency: 73편), 공정성(justice & fairness: 68편), 안전성(non-maleficence: 60편), 책임성(responsibility: 60편), 사생활 보호(privacy: 47편), 편익성(beneficence: 41편), 자율성(freedom & autonomy: 34편), 신뢰성(trust: 28편), 지속가능성(sustainability: 14편), 존엄성(dignity: 13편), 연대성(solidarity: 6편)이 그것이다[6]. 또한 2020년까지 발표된 인공지능 윤리성 관련 문헌자료 22편을 분석한 연구에서 윤리성 관련 내용의 언급 빈도 중 가장 자주 언급된 요인으로 공정성(18회), 책임성(17회), 투명성(16회), 안전성(16회)이 선정되었다[13].

이상의 문헌 분석을 통해 위의 범주 중 60편 이상(전체의 71.4%)에서 언급된 상위 네 개의 범주와 최다빈도로 언급된 윤리성 범주가 일치함을 확인하였으므로 투명성, 공정성, 안전성, 책임성이 본 연구의 인공지능기술 윤리성 범주로 선정되었다.

Table 1. Selected AI Ethical Awareness Categories

Selected Ethical Awareness Categories	Attributes
Transparency (Jobin, et al., 2019; Rothenberger, et al., 2019)	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing
Fairness (Jobin, et al., 2019)	Justice, fairness, consistency, inclusion, equality, equity, (non-) bias, (non-)discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Safety (Jobin, et al., 2019; Rothenberger, et al., 2019)	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion
Responsibility (Jobin, et al., 2019; Cointe, et al., 2016; Rothenberger, et al., 2019)	Responsibility, accountability, liability, acting with integrity

### 4.2 1단계 분석 결과

개방형 질문에 대한 응답을 내용분석한 결과, ‘(1)인공지능은 ( )와 ( )을 연결시켜준다’의 응답에는 ‘인간과 로봇, 기계, 컴퓨터, 사물 등’이 90% 이상으로 나타났다. ‘(2)인공지능이라고 하면, ( )이 먼저 떠오른다’의 응답에는 ‘AI, 로봇, 알파고, 인공지능피커, 자율주행, 컴퓨터,

Table 2. Open Questionnaire Items

No.	Items
1	Artificial intelligence connects ( ) and ( ).
2	When you think of artificial intelligence, ( ) comes to mind first.
3	Artificial intelligence I know makes ( ) possible.
4	What artificial intelligence can make our lives easier.
5	Write freely about what you know about artificial intelligence.
6	What do you know about artificial intelligence technology.
7	What products use artificial intelligence.
8	What are the intangible services that use artificial intelligence?
9	Should the AI development process be open to users.
10	If it should be made public, how should it be made public.
11	Should the application process of artificial intelligence technology be fair.
12	How to be fair if it should be fair?
13	What makes artificial intelligence technology uncomfortable or a threat to us.
14	Who is responsible for the development and use of artificial intelligence technology.
15	What parts should users be protected when applying artificial intelligence technology.
16	Gender
17	Age
18	Occupation

편리함 등이 90% 이상으로 나타났다. ‘(3)내가 아는 인공지능은 ( )을 가능하게 해준다’에 대해서는 ‘불가능, 편리한 생활, 학습, 정보처리, 대화 등이 대부분을 차지했다. ‘(4)인공지능이 우리 생활을 편리하게 해주는 것에는 무엇이 있나요’의 응답에는 ‘음성인식, 자율주행, 재택근무, 무인점포, 전자제품, 청소, 시간절약, 자동화 등이 대부분으로 나타났다. ‘(5)인공지능에 대해서 어느대로 자유롭게 적어보세요’의 응답에는 ‘가상현실, 로봇, 자율적인 선택지 생성, 인간의 사고방식과 유사, 미래기술, 빅데이터, 사람의 능력을 넘어선 능력, 인간을 대신하여 생각함 등이 있었다. ‘(6)인공지능 기술 중 알고 있는 것은 무엇인가요, (7)인공지능이 사용된 제품에는 무엇이 있나요, (8)인공지능이 사용된 무형의 서비스에는 무엇이 있나요’의 질문에는 ‘음성인식, 알파고, 자율주행, 컴퓨터, 학습능력, 생활가전, 휴대폰, 자동차, 챗봇, 서빙로봇, 로봇청소기, 드론, 인터넷쇼핑, 일기예보, SNS, 원격조정, 검색기능, 금융, 의료, 구인구직, 추천서비스 등의 응답이 있었다.

‘(9)인공지능 개발 과정이 사용자에게 공개되어야 할까요’의 투명성과 관련된 질문에는 ‘개발과정을 사용자에게 공개할 필요는 없다’는 응답과 ‘사용자의 민감정보 부분은 공개할 필요가 있다’는 응답이 나뉘었다. ‘(10)공개

되어야 한다면 어떻게 공개되면 좋을까요’에 대해서는 ‘인공지능 개발과 도입과정의 공개, 개념 위주로 적정 수준에서 공개, 제품구입 시 설명서 첨부, 대응매체나 교육기관을 통한 홍보, 홈페이지 게시, 학술지 게재, 포털사이트 통한 홍보 등’의 응답이 있었다. ‘(11)인공지능기술의 적용과정이 공정해야 할까요’의 질문에 대하여 ‘공정해야 신뢰할 수 있다, 악의적 이용 차단, 윤리적 측면 충족 필요함, 개발자의 허락이나 특허권 침해야 없어야 함 등’의 응답이 있었다. ‘(12)공정해야 한다면 어떻게 공정해야 할까요’의 질문에는 ‘법적 규제책 마련, 악의적 의도 차단, 지적재산권 침해 없어야 함, 차별없이 일관된 적용 등’의 응답이 있었다. ‘(13)인공지능기술이 우리를 불편하게 하거나 위협이 되는 점에는 무엇이 있을까요’에 대한 질문에는 ‘사생활 침해, 개인정보유출, 윤리적 탈선, 오작동에 의한 사고, 일자리 감소, 군사적 이용으로 평화 위협, 범죄에 이용 위험성, 인간의 나태함으로 연결, 인간생존권 위협, 고령층 이용의 어려움 등’의 응답이 많았다. ‘(14)인공지능기술 개발과 사용에 대한 책임은 누구에게 있을까요?’에 대해서는 ‘정부의 제도적 뒷받침 필요, 기업의 윤리의식 제고, 사용자의 윤리적 사용, 법적 책임소재 규정 필요 등’의 응답이 있었다. ‘(15)인공지능기술 적용 시 사용자가 보호받아야 할 부분은 무엇일까요?’에 대한 질문에는 ‘개인정보에 따른 차별금지, 사용자가 요구하는 기술 적용, 사회적 약자에 대한 배려, 개인정보유출 방지 등’의 응답이 많았다.

위와 같이 개방형 질문에 대한 응답에서 추출한 키워드를 중심으로 총 45개의 예비조사 문항을 생성하였다.

Table 3. Experts Review of the Open Questionnaire

Major	Qualification	Affiliation
Computer Science/ Machine Learning	Ph. D.	***Institute of Science & Technology Senior Researcher
Computer Science/ AI Development	Ph. D.	***University Lecturer

### 4.3 2단계 분석 결과

예비조사 결과에 따라 측정 항목을 선별하여 변인을 정교화하기 위해 기술통계분석(Descriptive Statistics)과 탐색적 요인분석(Exploratory Factor Analysis)[23]을 실시하였다. 요인분석모형은 공분산행렬을 이용하여 다음과 같은 공식에 의해 산출되었다.

$$\Sigma = \Lambda\Phi\Lambda' + \Psi \quad [\text{formula 1}]$$

$\Sigma$ : 변수( $x$ )의 모집단 공분산 행렬  
 $\Lambda$ : 요인부하 행렬  
 $\Lambda'$ : 요인부하 행렬의 전치 행렬  
 $\Phi$ : 요인의 공분산 행렬  
 $\Psi$ : 측정오차의 공분산 행렬

공통요인을 찾아내기 위한 요인분석모형은 측정변수와 잠재변수 또는 요인 간의 상관을 추정하도록 공분산 행렬을 이용하여 구한다. 2단계 분석에 의한 척도 정제 결과는 다음과 같다.

첫째, 문항평균은 2.93에서 4.53까지의 분포를 보였고, 표준편차는 모두 0.55 이상으로 지나치게 낮은 표준

편차값을 보인 문항은 확인되지 않았다. 정규분포 여부를 확인하기 위한 왜도(<math>\pm 3</math>)와 첨도값(<math>\pm 8</math>)[9,10]은 모두 범위 내에 분포하고 있어 정규성 기준을 충족하는 것을 확인하였다. 타당도를 확인하기 위한 KMO값은 .854로 상당히 좋은 편이었으며 Bartlett의 구형성 검정 값 역시 적절하여 요인분석을 실시하기에 적합한 데이터임이 확인되었다.

둘째, 탐색적 요인분석에서는 문헌분석을 통해 추출한 4 요인(투명, 공정, 안전, 책임)을 반영하여 고정된 요인수를 4개로 설정하고 베리맥스 회전에 의해 주성분 분석

Table 4. Preliminary Questions

No.	Items
1	Users should be able to know the AI development process.
2	The types of personal information collected by artificial intelligence should be disclosed.
3	The areas where AI is used need to be made public.
4	Users need to know what AI features are used in the product.
5	The AI development process and results are intellectual property and do not need to be disclosed.
6	Guidelines for transparency in AI development and use should be promoted.
7	There is a need for a clear concept of transparency in the development and use of AI.
8	Transparency in AI development and use should be searchable by users.
9	There is a need for an open discussion on the transparency of AI development and use.
10	In order to secure transparency in the development and use of artificial intelligence, it is necessary to promote the concept through news, portal sites, and SNS.
11	A manual on the transparency of AI development and use should be created.
12	Gender discrimination should be considered when using AI technology.
13	Age discrimination should be considered when using AI technology.
14	Racial discrimination should be considered when using AI technology.
15	Judgment should be considered in terms of human rights when using artificial intelligence technology.
16	Cultural differences should be considered when using AI technology.
17	When using artificial intelligence technology, the weak should be considered.
18	Consideration should be given to illegal activities when using artificial intelligence technology.
19	There should be no users at a disadvantage when using artificial intelligence technology.
20	When using artificial intelligence technology, there must be no malicious development intention.
21	There must be no infringement of intellectual property rights when using artificial intelligence technology.
22	When using artificial intelligence technology, institutional support for securing fairness should be considered.
23	I felt that artificial intelligence technology is being used fairly.
24	When using artificial intelligence technology, users may be exposed to harmful content on the Internet and SNS.
25	Personal information leakage or invasion of privacy may occur when using artificial intelligence technology.
26	Artificial intelligence technology feels safe.
27	Jobs may be lost due to artificial intelligence technology, and labor rights may be infringed.
28	Users may feel threatened about the change of life caused by artificial intelligence technology.
29	Repression of power using artificial intelligence technology may appear.
30	The ownership of AI technology should not be biased towards a specific class or group.
31	There are insufficient laws and systems to support changes caused by artificial intelligence technology.
32	Artificial intelligence technology could be used to develop weapons.
33	Artificial intelligence technology could be used in crime.
34	Artificial intelligence technology can be used in cyberbullying.
35	There is a possibility of malfunction or hacking of artificial intelligence technology.
36	Personal information should be protected when developing and using artificial intelligence technology.
37	When developing and utilizing artificial intelligence technology, it is necessary to respond to the needs of users.
38	Gender, age, and racial discrimination should be prevented when developing and using artificial intelligence technology.
39	Intellectual property rights should be protected when developing and utilizing artificial intelligence technology.
40	The legal responsibility for the development and use of artificial intelligence technology should be clearly defined.
41	It is the government's responsibility to establish laws and systems for the development and use of artificial intelligence technology.
42	The responsibility for accidents caused by artificial intelligence technology lies with the developer (company).
43	Responsibility for accidents caused by artificial intelligence technology rests with the government (competent agencies and departments).
44	Users are responsible for accidents caused by artificial intelligence technology.
45	Problems related to the use of AI technologies are the responsibility of developers, users, and governments all at the same time.



Table 5. Selected Interim Questions

Selected Ethical Awareness Categories	No.	Items
Transparency	3	The areas where AI is used need to be made public.
	4	Users need to know what AI features are used in the product.
	6	Guidelines for transparency in AI development and use should be promoted.
	7	There is a need for a clear concept of transparency in the development and use of AI.
	8	Transparency in AI development and use should be searchable by users.
	9	There is a need for an open discussion on the transparency of AI development and use.
	10	In order to secure transparency in the development and use of artificial intelligence, it is necessary to promote the concept through news, portal sites, and SNS.
Fairness	11	A manual on the transparency of AI development and use should be created.
	12	Gender discrimination should be considered when using AI technology.
	13	Age discrimination should be considered when using AI technology.
	14	Racial discrimination should be considered when using AI technology.
	15	Judgment should be considered in terms of human rights when using artificial intelligence technology.
Safety	16	Cultural differences should be considered when using AI technology.
	24	When using artificial intelligence technology, users may be exposed to harmful content on the Internet and SNS.
	25	Personal information leakage or invasion of privacy may occur when using artificial intelligence technology.
	27	Jobs may be lost due to artificial intelligence technology, and labor rights may be infringed.
	28	Users may feel threatened about the change of life caused by artificial intelligence technology.
Responsibility	29	Repression of power using artificial intelligence technology may appear.
	36	Personal information should be protected when developing and using artificial intelligence technology.
	37	When developing and utilizing artificial intelligence technology, it is necessary to respond to the needs of users.
	22	When using artificial intelligence technology, institutional support for securing fairness should be considered.
	39	Intellectual property rights should be protected when developing and utilizing artificial intelligence technology.
	40	The legal responsibility for the development and use of artificial intelligence technology should be clearly defined.
	18	Consideration should be given to illegal activities when using artificial intelligence technology.
19	There should be no users at a disadvantage when using artificial intelligence technology.	
20	When using artificial intelligence technology, there must be no malicious development intention.	

을 실시하였다. 탐색적 요인분석 결과, 공통성과 요인적 재량이 .40 이상인 문항을 유의한 것으로 판단하여 .40 이하로 나타나거나 두 개 이상의 요인에서 .40 이상의 요인적재량을 나타낸 문항 16개를 삭제하였다.

셋째, 위의 요인분석 후에 문항 간의 상관분석 결과를 검토하여 상관계수가 .20 이하로 나타나거나 문항 간 유사성이 매우 높은 문항 3개를 삭제하였다.

이상에서 기술통계분석, 탐색적 요인분석 및 상관분석 결과를 검토한 후 문항을 정제하여 총 4개 요인에 포함된 26문항이 본조사 설문 문항으로 도출되었으며 문항의 신뢰도는 모든 요인에서 .70 이상으로 적합하였다(Table 5, Table 6).

#### 4.4 3단계 분석 결과

##### 4.4.1 척도개발 결과

본조사 결과에 따라 투명성, 공정성, 안전성, 책임성 범주의 측정 항목에 대한 확인적 요인분석(Confirmatory Factor Analysis)[23]을 실시하여 최종적으로 척도를 정제하고 요인모형을 도출하여 적합도를 검증한 결과는 다음과 같다.

첫째, 확인적 요인분석 step 1에서 문항의 요인적재량

Table 6. The Results of Exploratory Factor Analysis

Factors (Reliability: Cronbach's $\alpha$ )	Item No.	Factor Loading	Communality	Total Variance	Cumulative Variance
Transparency (.866)	10	.750	.585	18.161	18.161
	8	.742	.577		
	9	.732	.575		
	6	.720	.567		
	7	.684	.541		
	11	.673	.527		
	3	.620	.415		
	4	.609	.426		
Fairness (.798)	12	.842	.733	16.536	34.697
	14	.824	.705		
	13	.755	.596		
	16	.606	.525		
	15	.545	.552		
Safety (.722)	28	.821	.674	10.725	45.423
	29	.700	.530		
	27	.648	.426		
	25	.567	.448		
	24	.556	.405		
Responsibility (.877)	20	.791	.655	9.383	54.806
	40	.786	.690		
	39	.760	.609		
	36	.709	.509		
	22	.695	.606		
	19	.658	.475		
	18	.643	.471		
	37	.598	.427		
KMO & Bartlett Test .851, $\chi^2 = 6231.060$ , $df = 325$ , $p = .000$					

(factor loading)이 .60 이하인 10개 문항을 삭제하고 (Table 7, Fig. 2), 확인적 요인분석 step 2를 실시하여 총 4요인, 16문항을 최종 도출하였다(Table 8, Table 12, Fig. 3).

둘째, 인공지능기술 윤리성 인식 척도의 최종 요인모형을 설정하여 적합도를 검증하였다. 모형의 적합도는 다양한 지수에 의해 판단할 수 있으며 본 연구에서는  $\chi^2$ , RMSEA(root mean square error of approximation), CFI(comparative fit index), IFI(incremental fit index)를 통해 모형 적합도를 확인하였다.

$\chi^2$  분포의 평균( $\chi_M^2$ )은 그 분포의 자유도(df)이므로  $\chi_M^2$ 을  $df_M$ 으로 나누어 모형의 적합도를 확인할 수 있다(formula 2).

$$\chi_M^2 = \frac{\chi_M^2 - df_M}{\sqrt{2df_M}} \quad [\text{formula 2}]$$

다음으로, 증분 적합도 지수 중 대표적 지수인 CFI 값을 다음의 공식으로 산출하였다(formula 3).

$$CFI = 1 - \frac{Max(\chi_M^2 - df_M, 0)}{Max(\chi_M^2 - df_M, \chi_N^2 - df_N)} \quad [\text{formula 3}]$$

$df_N$ : 기저모형의 자유도  
 $df_M$ : 연구자 설정 자유도

확인적 요인분석 모형의 적합도는 ( $\chi^2=351.079$ ,  $df=98$ ,  $CMIN/df=3.582$ ,  $RMSEA=0.072$ ,  $IFI=.928$ ,  $CFI=.928$ ) 양호한 것으로 확인되었다 (Table 8). 요인모형 분석에 따른 세부적인 경로계수와 유의도 결과는 Table 9와 같다.

#### 4.4.2 구성타당도 결과

Table 7. AI Ethical Awareness Fit Index for Step 1

Model	$\chi^2$	df	Absolute Fit Index		Incremental Fit Index	
			CMIN/df	RMSEA	IFI	CFI
AI Ethical Awareness	1716.244	293	5.857	.094	.776	.773
acceptance level (significance probability)	$p < .001$		$\leq 3$	$\leq .08$	$\geq .9$	$\geq .9$

Table 8. AI Ethical Awareness Fit Index for Step 2

Model	$\chi^2$	df	Absolute Fit Index		Incremental Fit Index	
			CMIN/df	RMSEA	IFI	CFI
AI Ethical Awareness	351.079	98	3.582	.072	.928	.927
acceptance level (significance probability)	$p < .001$		$\leq 3$	$\leq .08$	$\geq .9$	$\geq .9$

최종 선정된 인공지능기술 윤리성 인식 척도의 수렴타당도와 판별타당도를 검증한 결과는 다음과 같다.

첫째, 인공지능기술 윤리성 인식 척도의 수렴타당도를 확인하기 위해 개념신뢰도 (Construct Reliability)와 평균분산추출지수를 각 잠재변인(투명성, 공정성, 안전성, 책임성)의 표준화된 회귀계수와 측정 변인의 표준오차를 사용하여 산출하였다(Table 10). 개념신뢰도는 투명성 .859, 공정성 .841, 안전성 .795, 책임성 .857로 모두 .70이상의 값을 나타냈다. 동시에 각 요인의 평균분산추출지수(AVE)가 모두 .50 이상[11,12]으로 나타나 인공지능기술 윤리성 인식 척도의 수렴타당도가 적합한 것으로 확인되었다.

둘째, 인공지능기술 윤리성 인식 척도의 판별타당성 검증을 위해 평균분산추출지수값을 확인한 결과, 모두 .50을 상회하면서 윤리성 인식 요인 간 상관계수 제곱값(결정계수,  $r^2$ ) 보다 큰 것으로 나타나 각 요인 간 판별타당도가 확보된 것을 확인하였다(Table 11).

#### 4.4.3 신뢰도 결과

인공지능기술 윤리성 인식 척도의 신뢰도 분석결과 문항의 평균은 2.906-4.637의 범위였으며 표준편차는 .59 이상으로 특별히 높거나 낮은 문항은 존재하지 않았다. 정규분포 충족여부를 확인하기 위해 왜도와 첨도를 확인한 결과, 왜도( $\leq \pm 3$ )와 첨도값( $\leq \pm 8$ )[9,10]은 모두 범위 내에 분포하고 있어 정규성 기준을 충족하였다. 또한 문항 간 상관은 모두 유의한 것으로 확인되었다.

신뢰도 분석 결과, 전체 문항의 신뢰도 Cronbach's  $\alpha = .817$ , 요인별 신뢰도는 투명성 .855, 공정성 .833, 안전성 .791, 책임성 .852로 신뢰도 기준( $\alpha > .60$ )보다 높게 나타남으로써 모든 문항들이 각 요인의 개념을 잘 설명하고 있는 것으로 확인되었다(Table 13).

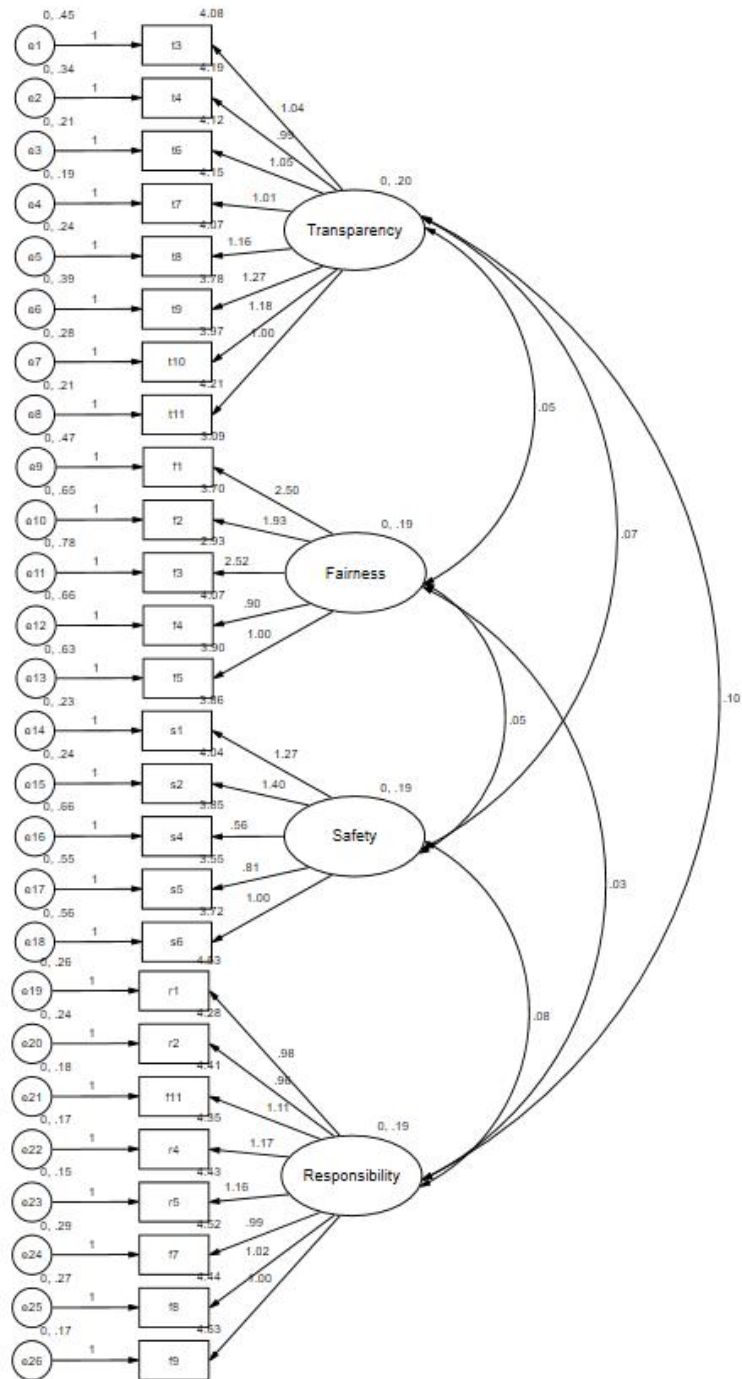


Fig. 2. Confirmatory Factor Analysis for Step 1

Table 9. Final Results of Confirmatory Factor Analysis

Model	Item No.	Standardized Regression Weights	Unstandardized Regression Weights	C.R.
Transparency	6	.730	1.08	14.475***
	7	.755	1.056	14.895***
	8	.703	1.109	14.001***
	9	.672	1.280	13.449***
	10	.701	1.175	13.964***
	11	.696	1.000	1.000***
Fairness	12	.913	1.052	17.406***
	13	.677	.713	15.371***
	14	.794	1.000	1.000***
Safety	24	.764	1.000	1.000***
	25	.859	1.214	9.403***
Responsibility	37	.670	.923	14.176***
	39	.800	1.102	16.882***
	40	.820	1.104	17.269***
	20	.657	.833	13.889***
	22	.738	1.000	1.000***

Table 10. Convergent Validity of Confirmatory Factor Analysis

Model	Item No.	Standardized Regression Weights	Standard Error	Construct Reliability	AVE
Transparency	6	.730	.075	.859	.504
	7	.755	.071		
	8	.703	.079		
	9	.672	.095		
	10	.701	.084		
	11	.696			
Fairness	12	.913	.060	.841	.641
	13	.677	.046		
	14	.794			
Safety	24	.764		.795	.661
	25	.859	.129		
Responsibility	37	.670	.065	.857	.548
	39	.800	.065		
	40	.820	.064		
	20	.657	.060		
	22	.738			

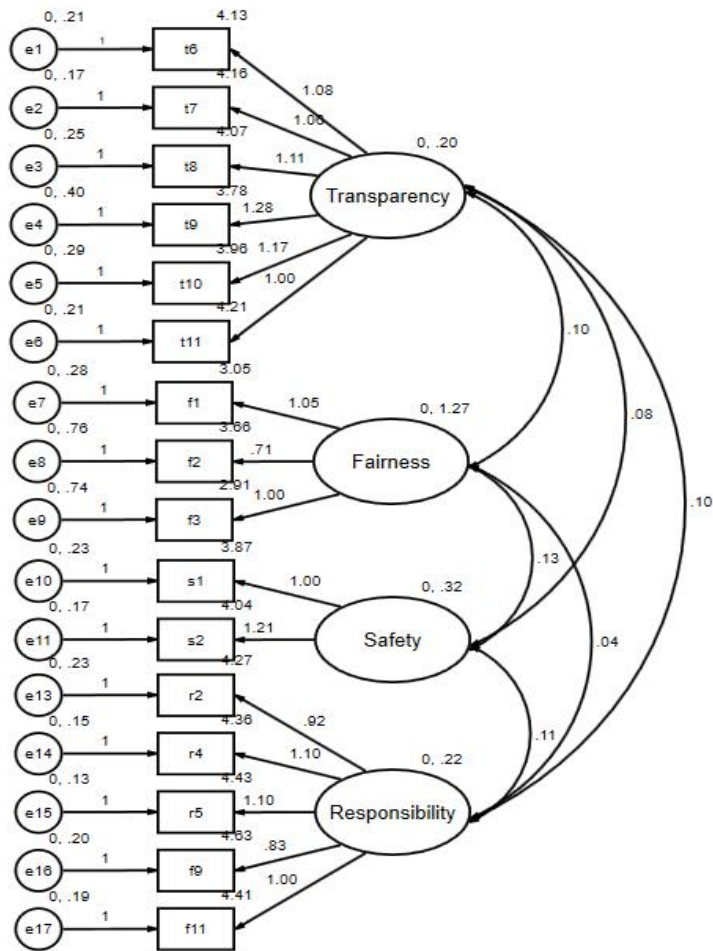


Fig. 3. Finalized Confirmatory Factor Analysis for Step 2

Table 11. AI Ethical Awareness Scale: Coefficient of Determination and AVE

Factors	Transparency	Fairness	Safety	Responsibility	AVE
Transparency	1				.504
Fairness	.032	1			.641
Safety	.075	.034	1		.661
Responsibility	.189	.006	.109	1	.548

Table 12. Finalized Scale of AI Ethical Awareness

Selected Ethical Awareness Categories	No.	Items
Transparency	6	Guidelines for transparency in AI development and use should be promoted.
	7	There is a need for a clear concept of transparency in the development and use of AI.
	8	Transparency in AI development and use should be searchable by users.
	9	There is a need for an open discussion on the transparency of AI development and use.
	10	In order to secure transparency in the development and use of artificial intelligence, it is necessary to promote the concept through news, portal sites, and SNS.
	11	A manual on the transparency of AI development and use should be created.
Fairness	12	Gender discrimination should be considered when using AI technology.
	13	Age discrimination should be considered when using AI technology.
	14	Racial discrimination should be considered when using AI technology.
Safety	24	When using artificial intelligence technology, users may be exposed to harmful content on the Internet and SNS.
	25	Personal information leakage or invasion of privacy may occur when using artificial intelligence technology.
Responsibility	37	When developing and utilizing artificial intelligence technology, it is necessary to respond to the needs of users.
	39	Intellectual property rights should be protected when developing and utilizing artificial intelligence technology.
	40	The legal responsibility for the development and use of artificial intelligence technology should be clearly defined.
	20	When using artificial intelligence technology, there must be no malicious development intention.
	22	When using artificial intelligence technology, institutional support for securing fairness should be considered.

Table 13. AI Ethical Awareness Reliability

Categories	No.		Communality	Mean	Standard Deviation	Cronbach's $\alpha$
Transparency	1	Guidelines for transparency in AI development and use should be promoted.	.600	4.128	.6637	.855
	2	There is a need for a clear concept of transparency in the development and use of AI.	.617	4.158	.6276	
	3	Transparency in AI development and use should be searchable by users.	.569	4.072	.7077	
	4	There is a need for an open discussion on the transparency of AI development and use.	.597	3.780	.8538	
	5	In order to secure transparency in the development and use of artificial intelligence, it is necessary to promote the concept through news, portal sites, and SNS.	.606	3.960	.7514	
	6	A manual on the transparency of AI development and use should be created.	.548	4.209	.6442	
Fairness	7	Gender discrimination should be considered when using AI technology.	.828	3.052	1.2979	.833
	8	Age discrimination should be considered when using AI technology.	.653	3.660	1.1863	
	9	Racial discrimination should be considered when using AI technology.	.770	2.906	1.4189	
Safety	10	When using artificial intelligence technology, users may be exposed to harmful content on the Internet and SNS.	.829	3.870	.7418	.791
	11	Personal information leakage or invasion of privacy may occur when using artificial intelligence technology.	.819	4.036	.8000	
Responsibility	12	When developing and utilizing artificial intelligence technology, it is necessary to respond to the needs of users.	.540	4.271	.6476	.852
	13	Intellectual property rights should be protected when developing and utilizing artificial intelligence technology.	.716	4.359	.6480	
	14	The legal responsibility for the development and use of artificial intelligence technology should be clearly defined.	.748	4.426	.6335	
	15	When using artificial intelligence technology, there must be no malicious development intention.	.611	4.637	.5938	
	16	When using artificial intelligence technology, institutional support for securing fairness should be considered.	.626	4.410	.6377	

## 5. 결론 및 제언

본 연구의 목적은 4차산업혁명 시대를 이끌어 갈 인공지능기술의 윤리적 사용에 대한 인식을 판단할 수 있는 근거로서 양적 측정도구인 인공지능기술 윤리성 인식 척도를 개발하는 것이다. 특히, 인공지능기술이 적용된 제품과 서비스를 이용하는 사용자의 인공지능 윤리에 대한 개념과 인식을 가늠해볼 수 있는 기준으로서 인공지능 윤리성 인식 척도를 제공하여 인공지능 윤리에 대한 개념과 문제의식을 높이는 데 목적을 두었다.

이에 본 연구에서는 인공지능기술의 윤리적 사용에 대한 인식을 판단할 수 있는 근거로서 양적 측정도구인 인공지능기술 윤리성 인식 척도를 개발하여 인공지능기술이 윤리적 측면에서 올바른 방향으로 사용될 수 있는 토대를 마련하고자 하였다. 연구절차와 분석결과를 통해 인공지능기술 윤리성 인식 척도의 타당도와 신뢰도를 모두 검증하여, 투명성(6문항), 공정성(3문항), 안전성(2문항), 책임성(5문항)의 총 16문항을 개발하였다. 각 문항은 Likert 5점 척도를 사용하여 각 문항에 대해 응답자가 자기보고식으로 반응하도록 구성하였다. 문항의 응답에 대한 채점방식은 각 요인별 응답 점수를 합산하고 문항 수로 나눈 평균값이 높을수록 인공지능기술 윤리성 인식의 정도가 높은 것으로 해석한다.

개발된 척도는 현재 문장의 형태가 일반 진술문으로 구성되었다. 이러한 기본 척도의 문장형식은 향후 척도를 사용하는 용도에 따라 질문의 형태나 진행형 등의 문장으로 수정하여 사용할 수 있다. 개발된 척도의 의도와 의미를 유지하면서 문장의 형식을 수정함으로써 실제로 측정하고자 하는 인공지능기술 윤리성 인식의 상태 및 응답자의 특성에 맞게 활용할 수 있다. 본 연구의 인공지능기술 윤리성 인식 척도는 인공지능기술의 윤리 관련 분야에 기개발되거나 사용된 척도 또는 문항이 매우 희박한 상황에서 개발되었다. 따라서 본 연구에서 개발한 척도의 각 요인을 구성한 문항에 대한 선행연구를 찾아보기 어려운 측면이 있다. 이것은 거의 처음으로 시도된 인공지능기술 윤리성 척도 개발의 결과물인 각 문항들을 향후 다양한 상황과 연구에 적용함으로써 문항이 더욱 정제되어야 할 것이다.

본 연구에서 선정된 투명성 변인의 문항에는 투명성을 위한 가이드라인의 홍보, 개념정의를 사용자가 이해할 수 있도록 해야 한다는 내용이 포함되었다. 또한 투명성 확보를 위한 대중적 홍보나 매뉴얼 등의 필요성을 담고 있다. 이러한 내용은 인공지능기술 사용자가 자신이 사용하

는 인공지능기술 적용 제품이나 서비스에 대해 알 수 있도록 적극적인 홍보와 교육이 필요함을 의미한다. 다음으로 공정성 변인의 문항은 성별, 연령, 인종에 대한 차별 문제가 포함되었다. 이것은 인공지능기술 적용에 있어서 가장 핵심적인 논의에 해당하는 것으로 인식되고 있다. 안전성 변인의 문항은 인공지능기술에 의해 유해콘텐츠에 노출될 위험성과 개인정보 유출의 위험성을 포함했다. 이 내용 또한 인공지능기술 적용 시에 가장 중요하게 고려하는 부분이다. 마지막으로 책임성 변인의 문항에는 이용자의 요구에 부응하고 지적 재산권, 법적 책임 소재, 제도적 뒷받침, 악의적 개발의도의 차단이 포함되었다. 이는 인공지능기술의 개발과 이용에 있어서 지켜져야 할 공식적인 규칙 또는 원칙들을 사용자가 인식하고 있음을 말해준다.

본 연구의 결과에 따라, 인공지능 기술 적용 시 윤리성의 고려에 대한 정책의 전환이 필요함을 제언하고자 한다. 첫째, 기업, 공공기관, 학교 등 다양한 분야의 제품과 서비스에 대하여 인공지능 기술을 적용할 때 인공지능 기술의 편리성과 함께 잠재적으로 유발될 수 있는 위험 또는 불편함의 가능성을 함께 고지하도록 의무화할 필요성이 있다. 둘째, 인공지능 기술이 적용된 제품과 서비스의 사용자가 지켜야 할 행동강령을 작성하여 사용자에게 제공할 필요가 있다. 셋째, 인공지능기술에 의해 각종 차별과 유해 콘텐츠 노출이 있을 경우, 이에 대해 사용자와 제3자가 시정을 요구할 수 있도록 하고 요구사항에 대해 인공지능기술보유 주체가 즉각적인 시정노력을 했는지에 대한 감찰 프로세스를 마련할 필요성이 제기된다. 넷째, 새로운 기술의 등장과 발전은 그 속도가 점점 빨라지고 있으므로 위의 내용에 해당하는 모든 사항과 절차가 법적, 제도적으로 효력을 가질 수 있도록 입법과 정책개발이 조속히 이루어져야 할 것이다. 다섯째, 인공지능기술 적용 제품과 서비스 사용 후에 본 연구에서 개발된 척도를 응용하여 사용자에게 설문을 실시함으로써 인공지능기술의 윤리성에 대한 고려의 적절성을 지속적으로 모니터링 할 필요성이 있다.

인공지능기술이 적용된 제품과 서비스의 사용자에게 해 문제의식이 제기되지 않는다면 인공지능 윤리성 문제에 대한 근본적 접근이 어려울 것이다. 따라서 사용자의 인공지능 윤리성 부분에 대한 인식 정도와 인공지능 기술이 적용된 제품과 서비스의 이용 여부 결정에 윤리성에 대한 고려가 포함되는지를 측정하여 그 결과를 인공지능 연구와 기술개발에 반영하는 프로세스가 정착된다면 인공지능 기술의 안전하고 긍정적인 발전에 기여할

것으로 기대한다.

REFERENCES

[1] L. Rothenberger, B. Fabian, E. Arunov. (2019). Relevance of Ethical Guidelines for Artificial Intelligence – A Survey and Evaluation, In *Proceedings of the 27<sup>th</sup> European Conference on Information Systems(ECIS)*,1–11.

[2] Junggi Economy. (2019). How to ensure transparency and accountability in AI decisions, 2019. 7. 25. Retrieved from <http://www.junggi.co.kr/article/articleView.html?no=23729&totalSearchField=&totalSearchText=%C0%CC%3%A2%C8%A3&totalSearchDate1=2019-01-01&totalSearchDate2=2019-12-31&totalSearchDGubun=4&prevPageName=searchMain.html&page=12>

[3] The Arm. (2020). *AI Today, AI Tomorrow: The Arm 2020 Global AI Survey*.

[4] Y.H. Ko, C.S. Leam. (2021). The Influence of AI Technology Acceptance and Ethical Awareness towards Intention to Use. *Journal of Digital Convergence*, 19(3), 217–225.

[5] National Information Society Agency. (2019). *Guidelines for Artificial Intelligence Ethics—Cases of Japan and EU*, NIA.

[6] N. Cointe, G. Bonnet, O. Boissier. (2016). Ethical Judgment of Agents’ Behaviors in Multi-agent Systems, *AAMAS*, 1106–1114.

[7] K. Siau, W. Wang. (2020). Artificial Intelligence(AI) Ethics: Ethics of AI and Ethical AI, *Journal of Database Management*, 31(2), 74–87.

[8] A. Jobin, M. Ienca, E. Vayena. (2019). The Global Landscape of AI Ethics Guidelines, *Nature Machine Intelligence*, 1(September), 389–399.

[9] R.B. Kline. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

[10] R.B. Kline. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

[11] C. Fornell, D. F. Larcker. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50.

[12] J. Anderson, D. W. Gerbing. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin*, 103(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>

[13] T. Hagendorff. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120.

[14] J. Whittlestone, R. Nyrupe, A. Alexandrova, S. Cave (2019). The role and limits of principles in AI ethics: towards a focus on tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200.

[15] J. Morley, L. Floridi, L. Kinsey, A. Elhalal. (2021). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. In *Ethics, Governance, and Policies in Artificial Intelligence*, pp. 153–183. Springer, Cham.

[16] D. Schiff, J. Biddle, J. Borenstein, K. Laas. (2020, February). What’s next for ai ethics, policy, and governance? a global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153–158.

[17] M. Coeckelbergh. (2020). *AI Ethics*. The MIT Press.

[18] H. Kim. (2019). *Artificial Intelligence and Ethics*, Communication Books.

[19] H.S. Ko, H.B. Jung, D. H. Park. (2019). Artificial Intelligence and Discrimination. *Justice*, (171), 199–277.

[20] S.P. Cho. (2017). The Study on threats of information security and their solutions in the fourth industrial revolution. *Korean Security Journal*, (51), 11–35.

[21] E.K. Lee. (2016). Exploring Educational Possibilities for Posthumans. *Theology and World*, (86), 83–111.

[22] R. Hanna, E. Kazim. (2021). Philosophical foundations for digital ethics and AI Ethics: a dignitarian approach. *AI and Ethics*, 1(4), 405–423. <https://doi.org/10.1007/s43681-021-00040-9>

[23] C. H., Choi, Y.W. Yoo. (2017). The Study on the Comparative Analysis of EFA and CFA. *Journal of Digital Convergence*, 15(10),103–111.

김도연(Doeyon Kim)

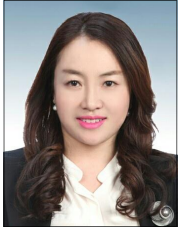
[상화임]



- 2001년 2월 : 서울대학교 사범대학 협동과정 환경교육 전공 박사수료
- 2002년 : 청주교육대학교, 한국방송통신대학교 겸임교수
- 2002년 ~ 2003년 : 호주 Griffith University 방문연구원
- 2008년 ~ 2009년 : 한양대학교(에리카) 교수학습개발센터 전문연구원
- 현재 : (주)로코모션뷰 대표
- 관심분야 : 조사, 분석, 통계, 인공지능, 플랫폼, 환경적 비용 편익분석
- E-Mail : gould55@snu.ac.kr

고 영 화(Younghwa Ko)

[장학원]



- 2018년 2월 : 연세대학교 경영대학 석사
- 2022년 2월 : 연세대학교 융합기술경영공학 박사졸업
- 현재 : 한양대학교 겸임교수 (경영학)
- 현재 : (주)국제탐어학원 & 유학원 대표
- 관심분야 : AI인공지능, 융합기술경영, 경영매니지먼트, 교육컨설팅

· E-Mail : mizstock2000@naver.com