

유아 특성을 고려한 AI 윤리 프레임워크와 규제 방향 도출에 관한 연구

정 광 식¹ · 양 재 훈² · 박 건 철^{3*}

¹차세대융합기술연구원 연구원

²차세대융합기술연구원 선임연구원

³차세대융합기술연구원 실장·선임연구원

Child-Centered AI: An Ethics Framework and Future Directions

Kwang Sik Jeong¹ · Jaehoon Yang² · Keon Chul Park^{3*}

¹Researcher, Advanced Institute of Convergence Technology, Suwon 16229, Korea

²Senior Researcher, Advanced Institute of Convergence Technology, Suwon 16229, Korea

³Director·Senior Researcher, Advanced Institute of Convergence Technology, Suwon 16229, Korea

[요 약]

본 연구의 목적은 유아의 특성을 고려한 AI(AI; Artificial Intelligence) 윤리 프레임워크를 제시하고, 유아의 안전하고 건전한 AI 활용을 위한 규제 방향을 제안하는 것에 있다. AI 기술이 다양한 분야로 확산됨에 따라 기술의 오남용으로 발생하는 사회적 부작용에 대한 우려가 커지고 있으며, 이에 따라 전 세계적으로 AI 윤리 기준을 마련하려는 노력이 경주되고 있다. 그러나 AI 기술 오남용에 특히 취약한 유아의 특성을 고려한 윤리 원칙에 대한 논의는 부족한 실정이다. 따라서 본 연구에서는 다기준 의사결정 방법론인 ANP(Analytic Network Process) 분석을 통해 유아 특성과 이용 환경에 적합한 AI 윤리 원칙 요소를 도출하고 각 요소의 가중치를 산정하였다. 또한, 사회적 인식 변화에 따른 요소들의 우선순위와 규제 방향의 변화를 제시하였다. 본 연구 결과는 유아의 AI 기술 오남용을 방지하는 동시에 건전한 AI 활용을 장려하여, 궁극적으로는 유아의 사회정서 발달을 지원하는 동반자 기술로의 발전을 위한 학술적, 정책적, 산업적 시사점을 제공할 것으로 기대된다.

[Abstract]

Various organizations worldwide have issued artificial intelligence (AI) principles to address ethical issues arising from AI development. However, there is a lack of research on AI principles that consider the characteristics of young children, who are particularly vulnerable to misuse of AI. This study aimed to propose an ethics framework for child-centered AI and suggest regulatory directions for the safe and sound use of AI by young children. To achieve this, we reorganized elements from existing AI principles with a focus on childhood characteristics and applied the Analytic Network Process, a multi-criteria decision-making methodology, to prioritize important ethical elements. The study further explored suitable regulatory directions for AI ethics considering children and examined changes in element priorities and regulatory directions in response to shifts in social awareness using sensitivity analysis. The results of this study are expected to provide academic, industrial, and policy-related implications for preventing misuse and promoting sound use of AI, which ultimately aids AI companion technology advancement for enhancing socio-emotional development in young children.

색인어 : 인공지능(AI), 유아, 규제, AI 윤리 프레임워크, ANP

Keyword : Artificial Intelligence(AI), Children, Regulation, AI Ethics Framework, Analytic Network Process

<http://dx.doi.org/10.9728/dcs.2024.25.8.2225>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 02 July 2024; Revised 08 August 2024

Accepted 13 August 2024

*Corresponding Author; Keon Chul Park

Tel: +82-31-888-9034

E-mail: parkkc07@snu.ac.kr

I. 서론

인공지능(AI; Artificial Intelligence)은 일상생활의 다양한 분야에 접목되어 삶의 편리함을 가져다주고 있다. 그러나 한편으로는 여러 유형의 윤리적 문제를 발생시키기도 한다[1]. 첫 번째 유형은 혐오와 차별적 발언으로 인한 문제이다. 마이크로소프트사의 AI 챗봇 ‘테이(Tay)’는 2016년에 출시되었으나, 인종차별과 혐오 발언으로 출시 16시간 만에 서비스가 전면 중단되었다. 국내에서도 2021년 12월 출시된 AI 챗봇 ‘이루다’는 차별·혐오적인 발언으로 출시 직후 논란이 되었다. 두 번째 유형은 데이터와 개인정보 침해로 인한 문제이다. 2020년 IBM사는 안면인식시스템의 데이터셋 구축 과정에서 미국 생체정보보호법(BIPA; Biometric Information Privacy Act)을 위반한 혐의로 집단 제소를 당하여 AI 관련 윤리적 문제에 직면하였다. 앞서 언급한 ‘이루다’ 또한 개발 과정에서의 개인정보 수집 및 처리에 대한 부분으로 법적 소송이 제기되면서 출시 20여 일 만에 서비스가 중단되는 사태가 발생했다. 이러한 문제들에 대응하기 위해 공공과 민간, 그리고 국제 차원의 다양한 기관에서 AI로부터 파생되는 문제에 대응하기 위해 노력하였다[2].

AI 기술을 선도하는 유럽연합(EU; European Union)과 미국에서는 AI로부터 파생되는 문제에 대응하면서 동시에 AI의 발전을 도모하기 위해 서로 다른 유형의 AI 규제를 시행하고 있다. 세계 최초 AI법(Artificial Intelligence Act)이 가결된 EU에서는 AI의 위험성을 사전에 분류하고 금지 행위를 규정하고 있다. 반면, 미국은 2023년 정부의 행정명령을 통해 AI의 잠재성을 극대화하고 산업 진흥 차원에서 자율적으로 윤리 준수를 바탕으로 한 사후 규제를 위주로 시행하고 있다.

한편, 사회의 다양한 집단은 각각 고유한 필요 사항이 있음에도 불구하고, 현재 논의되고 있는 AI 윤리 원칙은 이러한 특수성을 반영하지 못하고 있다[3]. 특히, 인구의 연령별 구분에 있어서 유아 특성을 고려한 AI 윤리에 대한 논의는 일반적인 AI 윤리에 대한 논의에 비해 부족한 실정이다[4].

유아 특성을 고려한 AI는 두 가지 측면에 대한 고려 사항을 포함한다. 첫째, 유아는 사회정서발달이 중요한 시기로 자아 형성과 자기의사결정이 불완전하다. 이로 인해 성인에 비해 혐오와 차별적 발언에 취약하며, 개인정보 유출의 위험이 높은 AI로부터 프라이버시 및 개인정보 보호가 특히 중요하다[5],[6]. 둘째, 유아는 AI의 활용을 체화하여 다양한 역량을 함양할 수 있으며, 장기적으로는 AI가 보편화된 미래 사회에 긍정적인 영향을 미칠 수 있다[7]. 유아 특성을 고려한 AI는 유아의 안전과 발달뿐만 아니라 사회적 측면에서도 중요하게 작용한다.

본 연구에서는 유아 특성을 고려한 AI 윤리의 구체적인 내용과 규제 방향을 논하고자 한다. 이를 위해서 첫째, 유아 특성을 고려한 AI 윤리 프레임워크를 도출한다. 유아 특성을 고려한 AI 윤리 프레임워크는 기존 AI 윤리 원칙에서 제시되는

윤리 요소를 유아 관점에서 재조합하고, 중요한 요소의 우선 순위를 도출한 개념이다. 이를 위해서 다기준 의사결정 방법인 ANP(Alytic Network Process)를 활용하였다. 둘째, ANP를 통한 최적의 대안 탐색으로 유아 특성을 고려한 AI의 적합한 규제 방향성을 정립한다. 셋째, AI에 대한 사회적 인식 변화를 가정했을 때의 요소 중요도 변화에 따른 규제 방향성의 변화 시나리오를 탐색한다. 이를 토대로, 최종적으로는 유아 특성을 고려한 AI 관련 정책적 시사점을 제공하고자 한다.

II. 이론적 고찰

2-1 유아 특성을 고려한 AI

AI의 사용 연령대가 확장되면서 AI와 인간의 상호작용 관련 연구의 범위도 폭넓게 확장되고 있다[2],[8],[9]. AI의 최초 사용 연령대는 유아기로 계속 낮아지고 있으며, 특히 유아용 스마트 장난감이나 스마트기기 등의 유아용 AI는 동반자로서 다양한 역할을 수행하며, 이에 따라 새로운 유형의 상호작용이 발생하고 있다[10],[11]. 유아를 대상으로 설계된 AI 기반 로봇은 유아의 인지 발달에 도움을 주며, 유아의 AI 로봇과의 놀이는 유아의 창의적·정서적·협력적 탐구 능력을 증진시킨다[12].

한편, 3-5세의 유아는 생물과 무생물의 차이를 인지하나, 로봇과 같은 무생물은 생물학적 특성을 가지지 않더라도 심리적인 특성을 가질 수 있다고 믿는다[13]. 유아의 주변 환경은 유아의 고등 정신 기능 발달에 상당한 영향을 미치기 때문에 유아의 환경에 녹아들 수 있는 AI에 대한 특별한 주의가 필요하다[14]. 하지만 AI에 대한 논의는 시스템의 고도화와 기술 활용을 통한 기회 창출에 초점이 맞춰져 있다[4]. AI가 유아에게 미칠 수 있는 위험이 간과될 수 있음에도 불구하고, 유아의 권리와 안전에 미치는 영향은 학계와 산업계의 주목을 받지 못하고 있다[5]. 유아와 AI 간 상호작용이 유아의 사회정서발달에 미치는 영향에 대해 아직 충분히 알려지지 않은 만큼, 유아의 환경에 노출된 AI에 대해서는 유아 고유의 특성을 고려한 세심한 접근이 필요하다[12],[15].

AI 윤리 관련 연구는 최근 국내에서 다방면으로 활발히 진행되고 있다[16]-[20]. 반면, 유아 특성을 고려한 AI 윤리 연구에서는 AI 교육에 대한 인식과 요구를 알아보거나, AI 기기와 유아의 상호작용에 대한 사례 연구, 동향 연구 등 기초적인 연구가 주를 이루고 있다[7],[21]-[24].

Kim and Kim[7]은 유아 AI 교육에 대한 학부모 인식과 요구사항을 고찰하였으나, 결과적으로 유아 AI 교육의 필요성과 AI 활용을 촉진하기 위한 환경에 대해서만 논의가 이루어졌다. Kim[22]은 유아 교육 분야와 관련된 AI에 대해 국내 연구 동향을 종합적으로 살펴보았으나 향후 관련 연구가 이루어져야 할 방향에 대한 개괄적인 제언을 하는 것에 그쳤다. Song[24]

은 유아를 위한 AI 융합 교육에서 윤리적 중요성을 강조하며 정서와 도덕성 등 논의가 필요한 쟁점에 대해 언급하고 있으나, 유아를 위한 구체적인 고려 사항이 무엇인지에 대한 논의는 이루어지지 않았다. 국제적으로는 유아 특성을 고려한 AI 윤리 원칙을 제시하고 있는 곳은 국제기구인 세계경제포럼(WEF; World Economic Forum)과 유니세프(UNICEF; United Nations Children's Fund) 두 곳뿐이다[5], [6].

종합적으로 일반적인 AI 윤리 연구는 다수 수행되고 있으나 유아 특성을 고려한 AI 윤리에 대한 구체적인 논의는 부족한 실정이다.

2-2 AI 윤리의 규제 방향

과학기술의 발전 속도는 법·제도 차원에서의 규제 대응 속도보다 빠르며, AI 기술 또한 예외가 아니다[25]. AI의 발전으로 일상생활에서의 편의가 증대되었지만, 법·제도의 테두리를 벗어나 발생하는 예상치 못한 윤리적 문제들 또한 많아졌다. AI의 발전 방향을 탐색하기 위해 본 절에서는 특정 행위가 이루어지기 전에 미리 규제하거나 금지하는 방식인 사전 규제와 행위가 이미 이루어진 후에 이를 규제하거나 필요한 조치를 취하는 사후 규제의 두 가지 관점을 살펴보고자 한다.

1) 사전 규제의 관점

EU는 AI법을 2021년에 발의했으며, 해당 법안은 2024년 3월부터 가결되어 시행되고 있다. 이는 기존에 연성법(soft law)으로만 존재하던 AI 윤리가 경성법(hard law)으로 구체화한 결과라고 볼 수 있다. 법안의 세부 내용을 살펴보면 고위험 등급에 해당하는 의료, 교육 분야에서는 위험관리시스템을 구축하고 사람이 최종 관리·감독하게 되어 있다. 또한 AI를 활용한 원격 생체인식, 데이터 기반 개인 평가 서비스인 소셜 스코어링 등 금지 항목 위주로 법안이 구성되어 있다. 종합적으로 AI법은 데이터 및 개인정보 보호에 중점을 두면서 AI의 잠재적 위험 요소를 사전에 규제하여 방지하는 성격의 법안이다. 기술영역 이론에서 정부 규제는 반드시 경제 성장을 동반하지는 않지만, 시장실패를 교정하고 방지하는 것에 있어서 중요한 역할을 한다[26]. 이는 기존에 자율적으로 준수되던 윤리 원칙이 공공의 성격이 강한 법·제도 차원으로 넘어오면서 행정적 구속력을 갖게 된 것에 의의가 있다[27].

2) 사후 규제의 관점

반드시 경제 성장을 담보로 하지는 않으면서도 다양한 측면에서 사회적 효용성을 추구하는 정부규제론과는 달리 혁신 확산이론(Innovation Diffusion Theory)에서는 개인이나 조직 차원에서 정보기술을 수용하고 혁신을 선택하는 과정이 경제적 이득, 사회적 위신, 편리성, 만족도 등의 상대적 이점으로 이어지는 부분을 설명한다[28]. EU와는 달리 미국은 이러한 혁신과 진흥에 방점을 두는 국가로, 과잉 규제에 의한 혁신 저해 가능성을 근거로 하는 사후 규제가 일반적인 규제 방법

론으로 자리 잡았다[29]. 일례로 미국의 국가과학기술위원회(NSTC; The National Science and Technology Council)는 2016년에 발표한 보고서에서 기존 규제의 리스크 관리 가능성을 먼저 판단하고, 규제 신설이나 개선이 규제 준수 비용을 증가하거나 기술개발 속도를 저해할 경우, 이를 보완할 수 있는 방안을 마련하도록 명시하였다. 사후 규제를 통한 민간 혁신 위주의 발전은 공공의 성격이 강한 사전 규제와는 다른 방향의 발전을 추구하며, 기술 자체의 진보를 위해서는 중요하게 고려해야 할 관점이다.

2-3 소결

유아 특성을 고려한 AI에 대한 접근은 일반 AI에 대한 접근과는 다르게 되어야 함에도 이에 대한 논의가 저조하다. 유아용 AI가 유아의 신체적·정서적 취약성에 해가 되지 않으면서 유아의 발달을 돕기 위해서는 기존의 AI 윤리 원칙과는 다르게 유아 특성을 고려한 AI 윤리 프레임워크가 필요하며, 윤리 프레임워크에서 중요한 요소들이 고려된 적합한 규제 방향성에 대한 논의도 필요하다. 더 나아가, AI에 대한 사회적 인식이 변화했을 때의 고려 요소의 변화나 규제 방향성의 변화에 대해서도 살펴볼 필요가 있다. 이를 종합적으로 고려한 본 연구의 연구문제는 아래와 같다.

- 연구문제 1. 유아 특성을 고려한 AI 윤리 프레임워크에서 중요하게 고려되어야 할 요소는 무엇인가?
- 연구문제 2. 유아 특성을 고려한 AI 윤리 프레임워크의 적절한 규제 방향성은 무엇인가?
- 연구문제 3. 사회적 인식 변화에 따라 추구해야 할 규제 방향성은 어떻게 바뀌며 여기에 미치는 요소는 무엇인가?

III. 분석 개요

본 연구를 수행하는 방법은 다음과 같다. 첫째, 국내·외 AI 윤리 원칙과 유아 특성에 초점을 맞춘 문헌 조사를 통해 유아 특성을 고려한 AI 윤리의 대원칙과 핵심 요소를 도출한다. 둘째, ANP를 활용하여 윤리 요소의 우선순위와 최적의 대안을 도출한다. 셋째, 민감도 분석을 통해 대안의 우선순위 변화를 시뮬레이션한다.

3-1 유아 특성을 고려한 AI 윤리 프레임워크 요소 도출

AI 윤리 원칙은 2007년 유럽로봇연구연합(EURON; European Robotics Research Network)에서 로봇윤리 13원칙이 발표된 이후 본격적으로 논의되기 시작했으며, 신뢰할 수 있는 AI의 구현을 위해 다양한 정책을 추진하고 있다. 2016년 이후로는 공공, 민간, 비정부기구(NGO; non-governmental organization) 등 전 세계의 다양한 주체에 의해 80개 이상의

AI 윤리 원칙이 발표되었다[30]. 이 중 중요하게 주목되는 내용을 공통으로 담고 있으면서 유아 특성에 초점을 맞춘 문헌을 검토하여 유아용 AI에서 고려되어야 할 대원칙과 핵심 요소를 도출하였다. 도출된 대원칙과 핵심 요소는 표 1과 같다.

1) 공정성(Fairness)

AI 윤리에서 공정성은 인권 보장과 사회 번영의 관점에서 중요하게 작용하는 요소로 많이 언급되는 개념 중 하나이다 [8]. 사람을 닮아가는 AI는 사회적 편견을 내재하고 그것을 대규모로 배포할 수 있다는 점에서 공정성 추구는 공공과 민간 모두에게 큰 과제로 다가온다. 특히, 미래 세대를 형성하는 중요한 주체인 유아는 사회에서 가장 취약한 구성원이고 차별이나 편견 등의 위협에 노출되어 있기 때문에 유아용 AI의 공정성에 대한 의무는 더욱 커지고 있다[5]. 이는 우리 사회에 각각 존재하는 집단들의 연대를 통해 유아를 위한 공정한 AI를 만드는 과정에 참여가 필요함을 시사한다[31]. 이를 종합하였을 때, 공정성은 AI 시스템의 개발 및 활용 전 주기에 걸쳐 특정 유아에 대한 차별이나 편견이 없도록 보장하며, 이를 통해 기술에 의한 번영이 인류에게 공유될 수 있도록 하는 원칙으로 정의될 수 있다.

2) Inclusiveness(포용성)

국제적으로 인권에 대한 인식이 높아지면서 포용성 또한 AI의 중요한 윤리적 고려 요소로 대두되고 있다. 포용성은 인간의 감정적, 신체적, 심리적 안정감을 형성하는데 필수 요소지만, 포용성이 부족할 때 유아가 느낄 수 있는 배제감은 유아의 자신감과 인지적, 감정적 발달에 해가 될 수 있다[5]. 이를 위해서 시각, 청각, 촉각 등 다양한 매개체를 통해 장애가 있는 유아 또한 AI로부터 혜택을 받을 수 있도록 접근성을 확보하는 것이 중요하다[5],[6],[32]. 다양한 배경의 유아 데이터를 포함하여 포괄적인 AI 시스템이 될 수 있도록 다양성을 추구하는 것과 AI를 사용할 유아를 위한 유아 중심 디자인을 추구하는 것 또한 포용성의 중요한 요소이다[5],[6],[32]. 이러한 핵심 요소를 바탕으로 포용성은 AI 시스템의 개발 및 활용 과정에서 유아의 다양한 데이터가 포함되어 가능한 많은 유아의 참여 및 사용이 가능하도록 하는 원칙으로 정의할 수 있다.

3) 책무성(Accountability)

AI 책무성은 어떤 행위에 대한 알고리즘 단위의 인과적 사건을 밝히는 것이 아닌, 행위가 발생한 본질적인 이유를 밝히는 것이다. 스스로 인지하고 판단하는 능력이 부족한 유아는 AI 기술이 접목된 매체와의 접촉에서 발생할 수 있는 상황에 대해 스스로 책임을 질 수 없다. 이에 기술 사용에 대한 책무 규칙의 확보가 필요하다[31]. 책무성을 확보하기 위해서는 AI 개발 및 설계, 활용의 전 주기에 걸쳐 책임 소재를 밝히는 것이 필요하다. 이러한 책임식별성은 AI 시스템의 개발 및 활용 과정에서 책임 주체를 설정하여 피해를 최소화하는 것으로 정의되고 있다. 책임 주체 설정과는 별개로 책무성 있는

AI는 정상적 작동에 대한 피드백과 매커니즘이 필요하며, 이를 위해서는 AI 시스템의 알고리즘, 데이터, 거버넌스 등 구성요소의 기준에 대한 검증을 가능하게 해야 한다. 또한, 유아 관점에서 책무성 있는 AI는 유아의 AI에 대한 이해 수준에 관계 없이 유아 사용자를 보호하는 방향으로 개발되어야 한다[6]. 종합적으로 책무성은 AI 시스템을 사용하는 유아에 대한 책임을 자체적으로 내재화하며, AI 시스템의 개발 및 활용 과정에서 발생가능한 문제들에 대한 책임 주체와 개발 과정에서 기준 설정과 검증을 통해 AI의 책무성을 충족하는 원칙으로 정의할 수 있다.

4) 안전성(Safety)

유아는 물리적, 심리적으로 자신을 보호할 수 있는 방어 기제의 발달이 미숙하고 AI로부터 받을 수 있는 악영향에 대한 안전 확보가 중요하기 때문에 특별히 데이터와 개인정보의 보호가 필요하다[5],[6]. 이와 관련된 요소는 프라이버시로, 이는 사용자의 데이터 및 개인정보를 익명화하여 사이버 안전을 보장하는 요소이다. AI 시스템의 기술적 보안성 또한 안전한 AI의 중요한 요소로서, 데이터 유출, 알고리즘 조작 등 외부의 위험 요인으로부터 시스템을 보호한다. 이 외에도 WEF과 UNICEF는 유아의 신체적, 사회적, 정서적 발달에 해가 되는 요인을 사전에 방지하는 것이 유아의 안전성 확보 차원에서 중요하다고 논하고 있다[5],[6]. 안전을 확보하려는 노력에도 불구하고 AI 시스템은 오작동할 수 있는 위험이 충분히 존재하며, 인간의 개입을 통해 제어가 가능하도록 만드는 것도 중요하다. 이러한 요소들을 복합적으로 아우를 수 있는 안전성은 AI 시스템을 이루는 데이터와 알고리즘의 유출, 오작동, 부작용 등의 문제를 방지하며, 특히 유아를 대상으로 신체적, 정서적, 사회적 해를 끼치지 않고 문제 발생 시 AI 시스템의 통제가 가능하도록 하는 원칙으로 정의가 가능하다.

5) 투명성(Transparency)

투명성 또한 AI 윤리에서 많이 언급되는 개념 중 하나이다 [7]. 일반적으로 AI의 투명성은 AI 시스템의 수명 주기인 설계, 개발 및 배포 과정 전체에 걸친 개방성을 의미하며 설명 가능성과 추적가능성을 전제로 한다. 유아 관점에서 AI의 설명 가능성은 유아뿐만 아니라 구성원 모두 AI 시스템과 작동 방식에 대한 이해를 증진시키는 요소이다. 추적가능성은 AI 시스템에서 문제 발생 시, 문제 발생 원인을 신속하게 추적할 수 있는 요소이다. AI 시스템의 예상되는 위험과 관련된 정보를 공개하는 사전 고지 또한 투명성에 포함될 수 있다. 이러한 개념들을 종합했을 때, 투명성은 AI 시스템의 작동 방식, 데이터 출처, 발생가능한 위험 요소 등을 투명하게 공개하여 AI에 대한 유아 및 관계자의 이해도와 대응력을 증진하는 원칙으로 정의할 수 있다.

6) 지속가능성(Sustainability)

특정한 과정이나 상태를 유지할 수 있는 능력을 의미하는

표 1. 유아 특성을 고려한 AI 윤리 프레임워크의 대원칙과 핵심 요소

Table 1. Basic principles and key factors of AI ethics framework focused on children

Base principles	Definition	Key factors	Definition	Source
Fairness	Ensuring that there is no discrimination or prejudice against certain children throughout the life cycle of AI systems, and that economic prosperity through technology is shared with humanity	Bias Prevention	Factor that ensures AI systems are not biased towards one side or the other	WEF, EU, Dubai, Rome, Harvard University
		Publicness	Factor that must be used to ensure that AI systems are utilized to promote social good and the common benefit of humanity	Ministry of Science & ICT, NIA, Harvard University
		Solidarity	Factor that maintains relationship solidarity between various groups and ensure opportunities for participation by various entities throughout the entire AI cycle	Ministry of Science & ICT, Université de Montréal
Inclusiveness	Including diverse data from children in the development and use of AI systems to enable participation and utilization by as many children as possible	Accessibility	Factor that allows children with disabilities to benefit from the AI system through various media such as sight, hearing, and touch	Ministry of Science & ICT, UNICEF, Beijing, Harvard University
		Diversity	Factor that enables a comprehensive AI system, including data from children from diverse backgrounds	Ministry of Science & ICT, UNICEF, Beijing, Rome, Université de Montréal
		Design Inclusiveness	Inclusive design factor for toddlers	WEF, UNICEF
Accountability	Ensuring accountability of AI systems and internalizing the responsibility for their use in infants, identifying the responsible entities for potential issues arising during the development and utilization of AI systems, and establishing standards and verification processes during the development process	Child Centricity	Factor that is appropriate for the infant's age and cognitive development stage and allows the infant to be the center of the development process	WEF, UNICEF
		Responsibility Identification	Factor that minimizes damage by establishing responsible parties in the process of developing and utilizing AI systems	IEEE, Ministry of Science & ICT, NIA, Beijing, Rome, Dubai, Université de Montréal
		Verifiability	Factor that enables verification of the standards of AI system components such as algorithms, data, and governance	NIA, Dubai, Harvard University
Safety	Preventing problems such as leaks, malfunctions, and side effects of the data and algorithms that make up the AI system, and enabling control of the artificial intelligence system when problems occur without causing physical, emotional, or social harm, especially to infants	Privacy	Factor that ensures infant privacy by anonymizing infant data and personal information	WEF, UNICEF, EU, IEEE, Ministry of Science & ICT, Rome, Beijing, Harvard University, Université de Montréal
		Security	Factor that protects AI systems from external risk factors such as data leaks and algorithm manipulation	WEF, UNICEF, EU, IEEE, Ministry of Science & ICT, NIA, Rome, Beijing, Université de Montréal
		Hazardousness	Factor that protects infants from factors that are harmful to their physical, social, and emotional development	WEF, UNICEF
		Control	Factor that enables human intervention and control for AI system malfunctions, etc	EU, IEEE, Beijing, NIA, Harvard University, Université de Montréal
Transparency	Improving the understanding and responsiveness of children and stakeholders about the AI system by transparently disclosing operation mechanism, data sources, and possible risk factors	Explainability	Factor that promotes understanding of AI systems and how they work, not only for children but also for members	WEF, UNICEF, EU, IEEE, Ministry of Science & ICT, NIA, Rome, Dubai, Harvard University, Merck KGaA
		Traceability	Factor that can quickly trace the cause of a problem when a problem occurs in an AI system	EU, IEEE, NIA, Dubai, Harvard University, Merck KGaA
		Informed Consent	Factor that disclosew information related to the expected risks of AI systems	Ministry of Science & ICT, NIA, Beijing
Sustainability	Ensuring that current and future AI developments for young children continue to occur	Well-being	Factor that promotes long-term social and emotional development in young children	UNICEF, EU, IEEE, Université de Montréal
		Competency	Factor that can cultivate AI utilization abilities of children and related parties	UNICEF, IEEE

표 2. 대안으로서 AI 윤리 원칙의 유형

Table 2. Types of AI ethics guidelines as alternatives

Type		Definition	Examples
Current ethics principles	International organizations and NGOs AI ethics principles	Principles and guidelines centered on human values distributed by various public and private entities at the international level in order to reach international social consensus at the level of international organizations and NGOs where discussions on human rights have continued and been active.	<ul style="list-style-type: none"> • OECD 「OECD Principles on AI」 (2019) • UNICEF 「Policy guidance on AI for Children」 (2021)
Alternatives	Public sector AI ethics principles	Prior-regulation-based public principles and guidelines that prioritize the impact and ethical issues of AI new technologies and are presented and specified in the public at the national level, which is the basis for laws, systems and regulations	<ul style="list-style-type: none"> • EU 「Ethics Guidelines for Trustworthy AI」 (2019) • Ministry of Science and ICT 「The National Guidelines for AI Ethics」 (2020)
	Private sector AI ethics principles	Post-liability-based private principles and guidelines set autonomously by the private sector to fulfill social responsibilities while embodying the company's goals in the capital market	<ul style="list-style-type: none"> • Kakao Algorithm Ethics • Naver AI Code of Ethics • Microsoft Responsible AI Standard

지속가능성은 유아용 AI에 대한 지속적인 발전 가능성을 의미한다[6]. 유니세프(UNICEF)는 특히 유아의 사회적, 정서적 발달을 장기적으로 증진하는 것이 유아 웰빙에 중요하며, 유아 및 관계자의 AI 활용 능력을 함양하는 것이 고도화되고 확산되는 AI에 대한 대응책임을 주장하고 있다[6]. 디지털 사회에 도래한 만큼, AI 역량 함양은 중요한 부분을 차지한다[33]. 이를 종합했을 때 지속가능성은 현재와 미래의 AI 발전이 유아를 위해 지속적으로 이루어지도록 하는 원칙으로 정의할 수 있다.

3-2 AI 윤리 원칙의 유형

AI 윤리 원칙은 다양한 기관에서 발표되고 있으며, 다양한 기관 중 정부기관, 민간기업·기관, NGO가 가장 많은 AI 윤리 원칙 문헌을 발표하는 것으로 나타났다[34]. AI 윤리 원칙을 발표하는 기관에 따라 그 성격이 다르게 나타나고 있는데, 공공의 경우 대중에 대한 책임을, 민간의 경우 고객에 대한 책임을, NGO의 경우 공공과 민간이 다루는 부분보다 폭넓게 윤리적 측면을 다루고 있으며, 민간보다는 공공과 NGO가 보다 법률과 규제에 더 많이 관여하고 보다 참여적인 프로세스를 통해 발전된다[30].

유아 특성을 고려한 AI 윤리 원칙은 인권에 대한 논의가 활발하게 이어져 온 국제기구 및 NGO 위주로 발표되었다[5],[32]. 국제적 합의 도출을 위해 국제 차원의 공공과 민간의 다양한 주체가 참여하여 배포하는 인간 가치 중심의 기존 원칙과 지침을 ‘국제기구 및 NGO 윤리 원칙’으로 정의하였다. 해당 유형의 원칙은 유아 특성을 고려한 AI의 윤리적 쟁점을 언급하고 있지만, 이러한 AI를 개발하고 사용하는 개인 수준의 특수성을 반영하기가 어려운 문제와 실천에 옮기도록 하는 구속력이 없다는 문제가 존재한다.

이에 대한 대안으로 사전 규제 관점에서 AI 기술이 가지는 파급력과 윤리적인 문제를 먼저 고려하고, 국가 차원의 공공에서 발표 및 명시하여 법·제도 및 규제 사항의 기초 자료가

되면서 준(準) 행정적 구속력을 갖는 AI 윤리 원칙과 지침을 ‘공공 위주 AI 윤리 원칙’으로 정의하였다. 이는 2-2절에서 논의한 AI에 대한 규제 관점과 혁신 관점 중 규제를 통한 AI 기술의 통제 및 안정화를 목표로 한다.

한편, 앞서 논의한 사후 규제의 관점에 따라 또 다른 대안을 제시하였다. 자본주의 시장에서 기업이 목표로 하는 상(像)을 담으면서도 사회적 책임을 다하기 위해 민간에서 자율적으로 설정하는 AI 윤리 원칙과 지침을 ‘민간 위주 AI 윤리 원칙’으로 정의하였다. 이는 사후 규제를 통한 자율적 윤리 준수를 바탕으로 AI 기술 발전에 초점을 맞춘 혁신을 목표로 한다.

3-3 ANP(Alytic Network Process)

유아 특성을 고려한 AI 윤리 요소의 중요도를 산정하기 위해 ANP를 활용하였다. ANP는 Saaty(1996)에 의해 개발된 기법으로, AHP(analytic hierarchy process)의 한계를 극복하기 위해 등장하였다[35]. ANP는 의사결정 체계에서 요소 간의 양방향적인 영향을 고려하여 상호작용, 상호의존, 피드백을 평가할 수 있다. AHP가 한 방향으로만 종속성을 표현하는 데 반해, ANP는 쌍방향의 관계를 고려하여 계층별 차이와 요소 간의 관계를 정량적으로 우선순위 도출이 가능하며, 궁극적으로 의사결정 문제를 처리할 때 복잡한 관계들을 쉽게 다룰 수 있는 장점을 제공한다[36].

3-4 연구 모형

3-1절에서 도출한 대원칙과 핵심 요소를 바탕으로 3-2절의 ANP를 수행하기 위한 비교 항목을 도출하였다. 비교 항목의 상관성 분석을 통해 계층화된 네트워크 도출을 위해서 1차 전문가 설문을 실시하였다. 설문은 상관성 도출을 위한 쌍대 비교 방식으로 진행되었으며, 유아용 AI의 개발과 활용의 윤리적 측면을 다양한 관점에서 고려할 수 있도록 6인의 전문가 풀을 구성하였다. 세부적으로는 AI 분야 교수 1인, AI 분야

연구자 1인, AI 개발자 1인, 데이터 사이언스 연구자 1인, 유아 발달심리학 연구자 1인, 교육공학 교수 1인으로 구성하였다. AI 분야는 최근 기관을 막론하고 많은 연구와 개발이 진행되고 있기에 전문가를 직업별로 1인씩 구성하였으며, 데이터 분야는 AI 분야와 밀접한 관련이 있기에 1인을 구성하였다. 발달심리학과 교육공학은 유아 특성을 고려하기 위한 핵심 분야이기 때문에 전문가를 분야별로 1인씩 구성하였다.

설문을 진행한 뒤, 설문지의 각 항목에 대해 6인 중 과반을 초과하는 4인이 상관관계가 있다고 응답한 항목에 대하여 최종적으로 상관성 네트워크 구조를 도출했으며, 이를 토대로 분석 모형을 그림 1과 같이 도식화하였다.

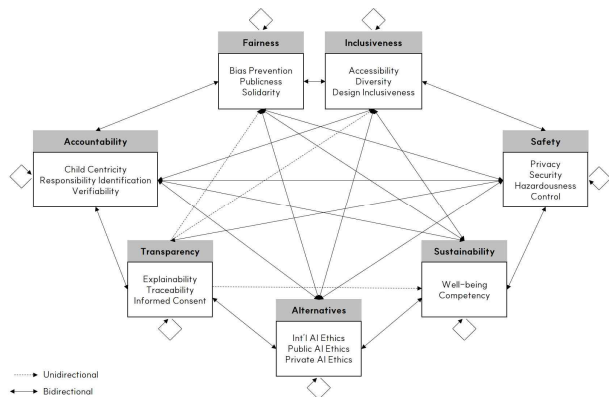


그림 1. 유아 특성을 고려한 AI 윤리 프레임워크의 ANP 모델
 Fig. 1. ANP model for deriving AI ethics framework focused on children

1차 설문에서 도출된 상관관계를 바탕으로 9점의 리커트 척도를 적용한 쌍대 비교를 위해 유아 교육, 데이터, AI 전문가로 구성된 집단 5인을 대상으로 2차 전문가 설문을 실시하였다. 이후 설문지 응답에 대한 일관성을 확보하기 위해 쌍대 비교 항목에 대해 일관성 비율(CR; consistency ratio)을 확인하였다. CR이 0.1 이상인 항목은 일관성이 없기 때문에 분석에서 제외하나, 본 분석에서는 모든 항목의 CR이 0.1 이하로 도출되어 모든 항목을 분석 대상에 포함하였다.

IV. 분석 결과

4-1 ANP 분석 결과

ANP를 통해 도출된 분석 결과는 다음과 같다. 먼저, 대안 선택에 영향을 미치는 대원칙 간 가중치를 구하였다. 대원칙 간 가중치는 대안 선택에서 고려되는 중요도와 영향 관계를 파악할 수 있으며, 이는 핵심 요소의 우선순위를 분석하기 위한 가중 대행렬과 극한 대행렬을 연산하기 위한 기초 값이다. 대원칙 간 가중치를 표현하는 클러스터 매트릭스(cluster matrix)는 표 3과 같다.

표 3에서 기술된 대원칙 간 가중치는 각 열의 합이 1이 되는 특징이 있으며, 좌측 요소가 상단 요소에 영향을 미친다. AI 윤리 프레임워크의 대안 선택에 안전성(39.86%)이 가장 영향을 많이 미치는 것으로 나타났으며, 대안의 자체적 영향력을 제외하고 책무성(17.47%), 지속가능성(10.93%), 투명성(7.77%), 공정성(5.81%), 포용성(5.38%) 순으로 영향력이 있는 것으로 나타났다. AI 윤리 프레임워크에 안전성이 가장 큰 영향을 미치는 것은 데이터와 알고리즘 유출 등 AI의 기술적 측면뿐만 아니라 유아에게 잠재적 신체적·정서적·사회적 해를 방지하는 것이 가장 중요하다는 것을 의미한다. 반대로, AI의 개발 및 활용 과정에서 유아의 다양한 데이터가 포함되도록 하는 포용성은 상대적으로 덜 고려되고 있음을 파악할 수 있다.

세부 요소 간 쌍대비교를 통해 도출된 초기 대행렬과 대원칙 간 가중치를 바탕으로 가중 대행렬과 극한 대행렬을 도출하여 AI 윤리 프레임워크의 대안 선택을 위한 핵심 요소의 우선순위를 도출하였다. 종합 결과는 표 4와 같다.

ANP 분석 결과, AI 윤리 프레임워크의 적합한 대안을 선택하기 위해 가장 중요하게 고려해야 할 요소 5가지는 책임식별성(21.325%), 검증가능성(17.015%), 보안성(11.756%), 통제성(9.947%), 그리고 유해성(6.255%)인 것으로 나타났다. 이는 유아용 AI를 개발하고 활용할 때 책임 주체를 설정하여 발생가능한 피해를 최소화하는 것(책임식별성), AI 시스템을 구성하는 알고리즘, 데이터 등의 기준에 대해 검증이 가

표 3. ANP 모형의 클러스터 매트릭스(cluster matrix)
 Table 3. Cluster matrix for the ANP model

	Fairness	Inclusiveness	Accountability	Safety	Transparency	Sustainability	Alternatives
Fairness	0.485662	0.190888	0.053456	0.046028	0	0.062886	0.058114
Inclusiveness	0.113309	0.510508	0.023132	0.029439	0	0.047246	0.053845
Accountability	0.072637	0.057316	0.545304	0.196486	0.238679	0.062865	0.174736
Safety	0.131346	0	0.179934	0.528244	0.157817	0.157142	0.398589
Transparency	0.070202	0.040153	0.071643	0.074392	0.515614	0	0.077744
Sustainability	0.062354	0.063939	0.045321	0.0645	0	0.587835	0.109314
Alternatives	0.064489	0.137195	0.081209	0.060911	0.08789	0.082025	0.127658

표 4. 클러스터(대원칙)와 노드(핵심 요소)의 우선순위
Table 4. Cluster(basic principles) and node(key factors) priorities

Cluster	Node	Value	Priority
Fairness (0.0695)	Bias Prevention	0.02554	10
	Publicness	0.039574	7
	Solidarity	0.004383	18
Inclusiveness (0.1587)	Accessibility	0.022929	11
	Diversity	0.015327	15
	Design Inclusiveness	0.014628	16
Accountability (0.3606)	Child Centricity	0.013569	17
	Responsibility Identification	0.193038	1
Safety (0.3044)	Verifiability	0.154029	2
	Privacy	0.05138	6
	Security	0.106417	3
	Hazardousness	0.056623	5
Transparency (0.0769)	Control	0.090044	4
	Explainability	0.026172	9
	Traceability	0.031212	8
Sustainability (0.0409)	Informed Consent	0.019532	13
	Well-being	0.021956	12
	Competency	0.018889	14

능하도록 하는 것(검증가능성), 그리고 데이터 유출과 알고리즘 조작 등 외부의 위협 요인으로부터 AI 시스템을 보호하는 것이(보안성) AI 윤리 원칙 유형 선택에 중요하다는 것을 의미한다. 위의 요소만큼은 아니지만 AI 시스템의 오작동 등에 대해 인간의 개입이나 제어가 가능하게 하는 요소와(통제성) AI가 유아에 미치는 영향 중 유아의 신체적·사회적·정서적 발달에 해가 되는 요소로부터 유아를 보호하는 것 또한 중요하다(유해성).

이러한 결과는 오늘날 유아 특성을 고려한 AI가 추구해야 할 방향이 공정성이나 포용성 등과 같이 장기적으로 도달해야 할 목표보다는 AI 시스템의 발생가능한 오류에 대한 실질적인 책임 식별과 이로 인한 피해를 최소화하기 위한 유아의 안전 보장이 중요하다는 것을 시사한다.

유아 특성을 고려한 AI 윤리 프레임워크의 가장 적합한 대안은 공공 위주의 AI 윤리 원칙(77.26%)인 것으로 나타났다. 이는 AI의 파급력을 우선적으로 고려하고, 법·제도 및 규제 사항의 기초 자료가 되는 국가 차원의 공공에서 발표 및 명시하여 준(準) 행정적 구속력이 우선시 되어야 함을 의미한다. 법은 반드시 지켜야 하는 최소한의 도덕이고 강제규범이라면, 통상 지칭되는 도덕은 내면적 동기에 관한 자율적 규범이라고 할 수 있다. 도덕적인 AI의 실현보다는 실질적으로 안전한 AI의 실현을 위해 공공 위주의 AI 윤리 원칙이 적합하다고 할 수 있다.

표 5. 대안(AI 윤리 원칙 유형)의 우선순위
Table 5. Alternative(Type of AI ethics principles) priorities

Alternatives	Weight	Priority
International Organizations and NGOs AI Ethics Principles	0.2260	2
Public Sector AI Ethics Principles	0.7226	1
Private Sector AI Ethics Principles	0.0711	3

국제기구 및 NGO 위주의 AI 윤리 원칙(22.60%)은 대안으로서 적합성이 다소 떨어지는 것으로 나타났는데, 이는 해당 유형의 윤리 원칙은 주로 포괄적인 내용을 바탕으로 하고 있으며, 실제 정책 입안자인 공공 차원에서나 제품 개발자인 민간 차원에서 적용하기에는 사회적 요구사항을 반영하지 못하여 실질성이 떨어지는 것에서 기인한다고 판단된다.

민간 위주의 AI 윤리 원칙(7.11%)은 나머지 두 대안에 비해 적합도가 매우 저조한 것으로 도출되어 AI 윤리 프레임워크의 대안으로 적용하기에는 다소 부적절한 것으로 나타났다.

가중 대행렬 분석을 통해 대안에서 가장 중요하게 고려해야 할 요소를 추가로 도출하였다. 공공 위주의 AI 윤리 원칙에서는 유해성(20.13%), 책임식별성(19.62%), 보안성(13.21%)을 가장 고려해야 하는 것으로 도출되었다. 이는 인간의 생애 주기 중 가장 취약하다고 볼 수 있는 유아에게 해가 될 수 있는 요소를 제거하고 데이터나 알고리즘의 유출을 방지하여 안전한 AI를 구축하고, AI의 개발 및 활용 과정에서 책임 주체의 명확한 설정을 통해 예측 불가한 피해를 최소화하는 데 있어서 공공의 역할이 중요하다는 점을 시사한다.

4-2 사회적 인식 변화에 따른 규제 방향성 변화

AI에 대한 사회적 인식이 바뀌고 이에 따른 요소 중요도가 변화할 때, 규제 방향성의 우선순위가 어떻게 변화하는지 탐색하고자 민감도 분석을 진행하였다. 민감도 분석 그래프에서 세로축은 대안의 가중치를, 가로축은 요소의 가중치를 의미한다. 가중치는 중요도 백분율로도 해석된다[35]. 민감도 분석 결과, 요소 가중치가 변화할 때 대안의 우선순위가 동시에 변화하는 요소는 프라이버시, 웰빙, 역량, 다양성, 그리고 유아 중심성의 5개 요소로 나타났다.

프라이버시의 경우, 그림 2와 같이 요소 가중치 전 구간에 걸쳐 공공 위주의 AI 윤리 원칙이 제일 적합한 대안으로 나타나고 있으나, 가중치가 0.78을 초과하는 지점에서 민간 위주의 AI 윤리 원칙이 국제기구 AI 윤리 원칙보다 대안으로서 더 적합해지는 것을 확인할 수 있다. 비록 민간 위주의 AI 윤리 원칙이 가장 적합한 대안은 아니지만, 프라이버시에 대한 중요도 인식이 높아질수록 자율성 높은 민간 차원에서의 AI 윤리 원칙을 차순위 대안으로 고려해 볼 수 있다.

웰빙, 역량, 그리고 다양성에 대한 민감도 분석 결과는 매우 유사한 형태로 나타나고 있다(그림 3-5). 가중치가 0.7~0.8 구간에 이르기까지는 공공 위주의 AI 윤리 원칙이 가장

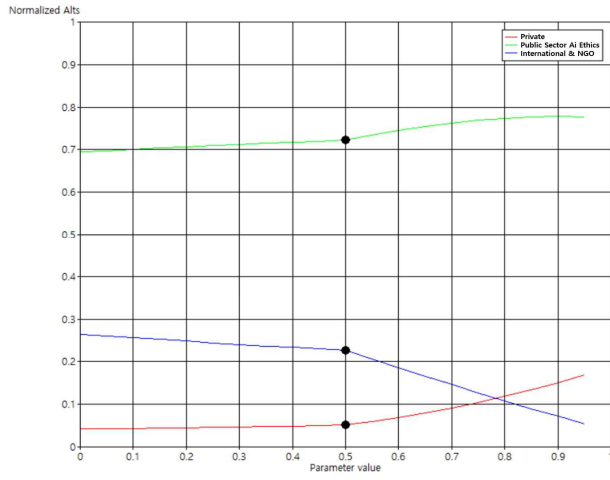


그림 2. 프라이버시에 대한 민감도 분석
 Fig. 2. Sensitivity analysis result for privacy

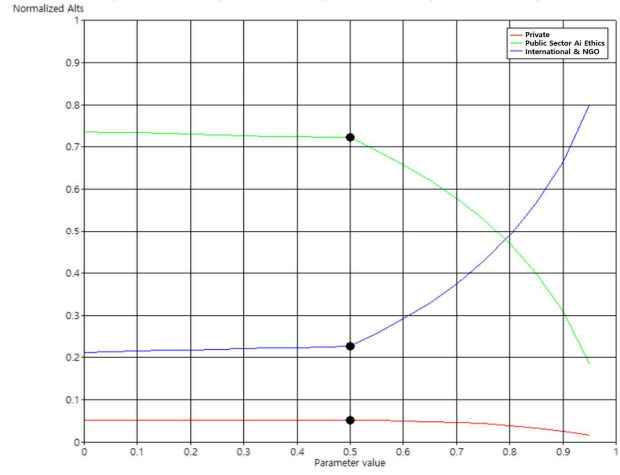


그림 3. 웰빙에 대한 민감도 분석
 Fig. 3. Sensitivity analysis result for well-being

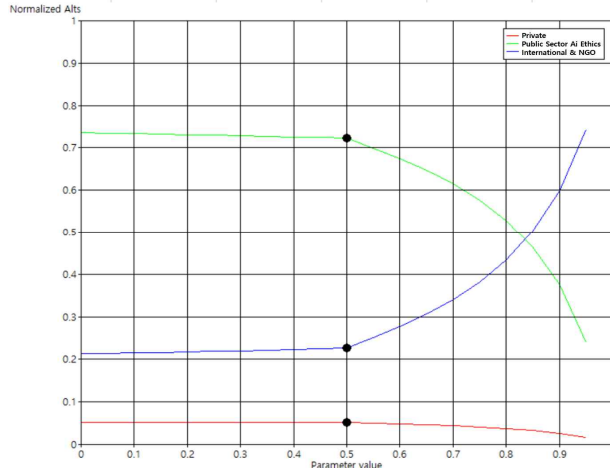


그림 4. 역량에 대한 민감도 분석
 Fig. 4. Sensitivity analysis result for competency

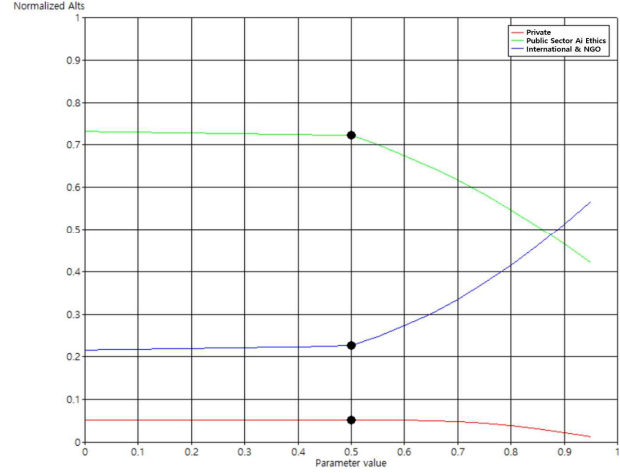


그림 5. 다양성에 대한 민감도 분석
 Fig. 5. Sensitivity analysis result for diversity

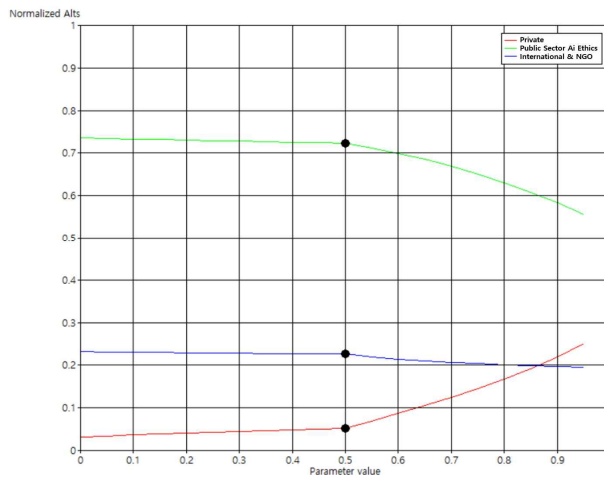


그림 6. 유아중심성에 대한 민감도 분석
 Fig. 6. Sensitivity analysis result for child centricity

적합한 대안으로 나타나다가, 이후 각 요소가 더욱 중요해지는 시점이 오면 국제기구 AI 윤리 원칙이 적합한 대안으로 나타난다. 웰빙과 역량, 그리고 다양성은 4-2절에서 도출된 핵심 요소 순위에서는 상대적으로 중요하지 않은 요소들로, 이는 유아용 AI의 기술적 성숙도 관점에서 단기적으로 해결해야 할 과제보다는 장기적으로 지향해야 할 고차원적인 목표로 볼 수 있다. 이는 유아용 AI가 기술적으로 충분히 성숙해지고 웰빙, 역량, 다양성과 같은 요소들이 더욱 중요해질 때 행정적 구속력으로 이어질 수 있는 공공 위주의 AI 윤리 원칙보다는 행정적 구속력은 없지만 장기적으로 고차원적 목표를 제시할 수 있는 국제기구 및 NGO AI 윤리 원칙의 역할이 더 커질 수 있다고 해석이 가능하다.

유아중심성 또한 그림 6에서 볼 수 있듯이 가중치 변화의 전 구간에 걸쳐서 공공 위주의 AI 윤리 원칙이 가장 적합한 대안으로 도출되었다. 다만, 유아중심성이 중요해질수록 공공 AI 윤리 원칙은 대안으로서 적합도가 다소 감소하고, 특정 시점에서 민간 위주의 AI 윤리 원칙이 국제기구 AI 윤리 원칙보다는 더 적합한 차선책 대안으로 나타나고 있다.

V. 결 론

AI의 발전은 여러 윤리적 문제를 동반하고 있으며, 사회적으로 이에 대한 대응이 필수적인 과제로 다가오고 있다. AI를 활용하는 연령대는 확장되고 있으며, 특히 유아가 활용하는 유아용 AI는 새로운 상호작용의 매체로 자리 잡고 있다. 기술을 통해 유아의 건강과 웰빙을 지원하고자 하는 시도의 증가는 매우 고무적이지만, 유아에게 발생할 수 있는 문제와 부작용을 방지하기 위한 윤리적인 논의는 여전히 부족한 실정이다. 한편, AI의 규제 방향성에 대해서도 국내에서는 아직 사회적 합의가 이루어지지 않았다. 2023년에 사후 규제 방식에 방점을 두는 「인공지능 산업 육성 및 신뢰 기반 조성에 관한 법률안」이 발의되었으나, 국가인권위원회의 사후 규제 원칙에 대한 반대를 비롯한 사회적 비판 여론으로 인해 해당 법안은 1년이 넘게 국회에서 계류되고 있다. AI 기술의 발전은 세계적 수준을 따라가고 있으나, 윤리 규제 방향성은 아직 정립되지 않은 실정이다.

본 연구는 이러한 공백을 채우고자 유아 특성을 고려한 AI 윤리 프레임워크와 규제 방향성을 제시하였다. 첫째, 기존 AI 윤리 원칙 문헌에서 제시되고 있는 윤리 요소를 재구성하고, ANP를 통해 이들 요소의 중요도를 분석하여 유아 특성을 고려한 AI 윤리 프레임워크를 도출하였다. 둘째, 도출된 윤리 프레임워크에 대한 규제 방향성을 담을 수 있는 최적의 AI 윤리 원칙의 유형을 대안으로써 제시하였다. 셋째, 사회적 인식 변화에 따른 규제 방향성의 변화 시나리오를 탐색하였다.

본 연구 결과에 대한 요약은 다음과 같다. 첫째, 대원칙 간 가중치 분석에서 안전성이 가장 큰 영향을 미치는 것으로 나

타났으며, 책무성과 지속가능성도 준수한 영향력을 가지는 것으로 나타났다. 이는 안전성이 유아에게 잠재적인 신체적, 정서적, 사회적 위협을 방지하는 데 가장 중요한 역할을 하며, 포용성은 상대적으로 덜 고려되고 있음을 의미한다. 이는 UNICEF와 WEF에서 유아 특성을 고려한 윤리에서 포용성과 다양성이 중요하게 고려되어야 한다는 주장과는 다른 양상으로 나타나고 있다[5],[6]. 둘째, 유아 특성을 고려한 AI 윤리 프레임워크의 핵심 요소 우선순위와 이에 대한 적합한 규제 방향성을 담는 윤리 원칙 유형을 도출하였다(연구문제 1). 핵심 요소 중 가장 중요한 요소는 책임식별성, 검증가능성, 보안성, 통제성, 그리고 유해성으로 나타났다. 이러한 요소들은 안전성과 같이 AI 시스템의 안전 보장과 데이터, 알고리즘의 검증가능성을 강조하고 유아에게 잠재적인 위협을 최소화하는 데 중요한 역할을 한다. 유아 특성을 고려한 AI 윤리 프레임워크에서 가장 적합한 윤리 원칙의 유형은 공공 위주의 AI 윤리 원칙인 것으로 나타났다(연구문제 2). 이는 국내에서 아직 AI 규제 철학에 대한 합의가 충분히 이루어지지 않은 가운데, 유아에 대한 예측 불가한 피해를 최소화하고 안전한 AI 시스템을 구축하기 위해서는 공공의 사전 규제 방향이 적절하다는 것을 시사한다. 셋째, 대안의 우선순위 변화를 시뮬레이션하기 위해 민감도 분석을 실시하였다(연구문제 3). 눈여겨볼 만한 결과는 웰빙, 역량, 다양성과 같은 정치적 중립성이 있어야 하는 요소들이 더욱 중요해질 때, 행정적 구속력으로 이어질 수 있는 공공 위주 AI 윤리 원칙보다는 국제기구 AI 윤리 원칙의 역할이 커질 수 있다는 점이다. 해당 결과는 사회적 인식 변화에 따라 적용해야 하는 적절한 윤리 원칙의 유형이 변화할 수 있음을 의미한다.

AI의 활용 패러다임은 작업 중심에서 상호작용 중심으로 넘어가고 있으며, 이를 통해 AI 기술은 유아의 사회기술적 역량을 함양할 수 있는 동반자로서 역할이 가능하다. 본 연구 결과는 유아용 AI 동반자 기술의 발전을 위해 학술적, 정책적, 산업적 측면에서 시사점을 제공할 수 있다. 첫째, 유아 교육 현장에서 AI 활용 지침을 마련하고, 관련 정책을 수립하는 유용하게 사용될 것이다. 구체적으로, 유아 교사들은 AI 기술의 올바른 사용 방법과 잠재적인 위험을 인식하고 관리할 수 있는 구체적인 지침을 얻을 수 있으며, 교육 당국은 이를 바탕으로 안전하고 효과적인 AI 교육 프로그램을 개발할 수 있을 것이다. 둘째, 정책 입안자들은 유아의 발달 단계에 적합한 AI 기술 사용 규제와 보호 장치를 마련하여 유아들이 AI와 상호작용하는 과정에서 안전을 보장할 수 있을 것이다. 셋째, 산업계에서는 유아를 대상으로 한 AI 제품을 개발할 때, 본 연구에서 제시된 윤리적 기준을 준수함으로써 제품의 신뢰성을 크게 높일 수 있을 것이다. 이는 소비자들의 신뢰를 얻고, 장기적으로는 시장에서의 경쟁력을 강화하는 데 중요한 역할을 할 것이다. 궁극적으로, 본 연구를 통해 AI를 사용하는 유아의 안전과 사회정서적 발달을 지원하는 AI 동반자 기술의 발전을 도모하여 유아와 AI 기술이 상생하는 사회적 모델을 제시할 수 있을 것이다.

한편, 본 연구는 유아 특성을 고려한 AI 윤리 연구의 초기 단계로 조사를 위한 설문을 소수의 AI, 유아, 교육 분야 전문가를 대상으로만 한정했다는 것에 한계가 있다. 이는 ANP 방법론 특성상 비교 문항이 많을 경우 전문가 의견이 과다하게 많아질수록 오히려 전체 분석에 대한 일관성 척도가 낮게 나타날 수 있음을 감안한 것으로, 분야별 관점을 수렴하면서도 설문 타당성을 저해하지 않는 수준에서 전문가 표본 수를 설정한 것이다[35]. 그럼에도 불구하고 향후에는 다학제간 다양성 차원에서 AI 개발자뿐만 아니라 UI/UX 전문가, AI 엔지니어, 로보틱스 엔지니어 등 AI의 개발 단계별로 참여하는 전문가 집단과 AI의 실사용자인 유아의 관리·감독 역할을 수행하는 부모 집단으로 대상을 확대하여 연구를 진행할 필요가 있다. 또한, 후속 연구에서는 유아 특성을 고려한 AI에서 사용자 경험 관점의 다양한 요소를 파악하여 AI 개발 과정에서 단계별로 적용할 수 있는 실질적인 가이드라인을 구축하고, AI 사용자 입장에서의 요구사항을 파악하는 실증적 연구를 통해 구체적인 사용 가이드라인을 구축할 수 있을 것이다. 이를 통해 유아 특성을 고려한 AI의 개발과 사용에 대한 실질적인 학술적, 정책적, 산업적 시사점을 제공하고, 궁극적으로는 유아의 발달을 지원하는 안전하고 건전한 AI의 사용을 장려할 수 있을 것이다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00251378).

참고문헌

- [1] S. So and S. Ahn, "A Study on the Classification Model and Components of Artificial Intelligence Ethical Principles," *The Journal of Korean Association of Computer Education*, Vol. 24, No. 6, pp. 119-132, November 2021. <https://doi.org/10.32431/kace.2021.24.6.010>
- [2] A. Jobin, M. Ienca, and E. Vayena, "The Global Landscape of AI Ethics Guidelines," *Nature Machine Intelligence*, Vol. 1, No. 9, pp. 389-399, September 2019. <https://doi.org/10.1038/s42256-019-0088-2>
- [3] J. P. Singh, "AI Ethics and Societal Perspectives: A Comparative Study of Ethical Principle Prioritization among Diverse Demographic Clusters," *Journal of Advanced Analytics in Healthcare Management*, Vol. 5, No. 1, pp. 1-18, January 2021.
- [4] E. Fosch-Villaronga, S. van der Hof, C. Lutz, and A. Tamò-Larrioux, "Toy Story or Children Story? Putting Children and Their Rights at the Forefront of the Artificial Intelligence Revolution," *AI & Society*, Vol. 38, No. 1, pp. 133-152, February 2023. <https://doi.org/10.1007/s00146-021-01295-w>
- [5] World Economic Forum, *Artificial Intelligence for Children: Toolkit*, Author, Geneva, Switzerland, March 2022.
- [6] UNICEF (United Nations Children's Fund), *Policy Guidance on AI for Children (Version 2.0)*, Author, New York: NY, November 2021.
- [7] K. Kim and M. Kim, "Perception and Demand of Kindergarten Parents for Artificial Intelligence (AI) Education for Infants," *Korean Journal of Early Childhood Education Research*, Vol. 25, No. 2, pp. 114-143, June 2023. <https://doi.org/10.15409/riece.2023.25.2.5>
- [8] G. Yoo and S. Kim, "The Effects and Aspects of Artificial Intelligence Speakers Activity on Children's Interaction and Creative Problem Solving Skills," *The Journal of Korea Open Association for Early Childhood Education*, Vol. 26, No. 5, pp. 209-244, October 2021. <https://doi.org/10.20437/KOAECE26-5-09>
- [9] G. Son, J. So, J. Ko, J.-W. Lee, J. R. Lee, and W.-S. Shin, "Enhanced AI Model to Improve Child Speech Recognition," *Journal of Digital Contents Society*, Vol. 25, No. 2, pp. 547-555, February 2024. <https://doi.org/10.9728/dcs.2024.25.2.547>
- [10] R. Kim, H. Hyun, and M. Song, "The Development of STEAM Program for Young Children Using Picture Books and Smart Devices," *The Journal of Korean Association of Computer Education*, Vol. 25, No. 6, pp. 85-102, November 2022. <https://doi.org/10.32431/kace.2022.25.6.007>
- [11] E.-S. Oh and Y.-H. Kim, "The Core Competencies of Young Children and the Direction of Early Childhood Education in the Fourth Industrial Revolution," *Journal of Digital Contents Society*, Vol. 20, No. 5, pp. 1011-1021, May 2019. <https://doi.org/10.9728/dcs.2019.20.5.1011>
- [12] X. Wang, N. Yin, and Z. Zhang, "Smart Design of Intelligent Companion Toys for Preschool Children," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, Vol. 35, No. 2, pp. 151-164, May 2021. <https://doi.org/10.1017/S0890060420000499>
- [13] S. L. Calvert, "Intelligent Digital Beings as Children's Imaginary Social Companions," *Journal of Children and Media*, Vol. 15, No. 2, pp. 291-296, 2021. <https://doi.org/10.1080/17482798.2021.1896200>
- [14] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*, Cambridge, MA: Harvard University Press, 1978.
- [15] S. J. Chang, S. Han, and K. C. Park, "Social Features of Artificial Intelligence Companions for Children," *Journal*

- of Digital Contents Society*, Vol. 24, No. 12, pp. 3199-3211, December 2023. <https://doi.org/10.9728/dcs.2023.24.12.3199>
- [16] J. N. Kim, E. C. Bang, H. Y. Jang, and S. Y. Song, "A Study on Developing the Ethical Framework of AI Learning Data," *Journal of Ethics*, No. 140, pp. 283-309, March 2023. <http://doi.org/10.15801/je.1.140.202303.283>
- [17] J. Bang, B.-T. Lee, and P. Park, "A Study on Guidelines for AI Ethics in the Process of Aligning Conversation Topics by Conversational Bot," *Journal of Broadcast Engineering*, Vol. 28, No. 5, pp. 564-577, September 2023. <https://doi.org/10.5909/JBE.2023.28.5.564>
- [18] S. Yoo, "Ethical Issues Posed by 'Generative-AI' (G-AI) - Response Strategies for 'Good AI Society'," *Journal of the Korean Bioethics Association*, Vol. 24, No. 1, pp. 1-29, June 2023. <https://doi.org/10.37305/JKBA.2023.06.24.1.1>
- [19] H.-Y. Park, "What is Post-Humanistic AI Ethics?: Focusing on the Term of 'Ethical AI'," *The Journal of Korean Association of Computer Education*, Vol. 25, No. 6, pp. 71-83, November 2022. <https://doi.org/10.32431/kace.2022.25.6.006>
- [20] S.-Y. Choi, "A Study on the AI Literacy Framework," *The Journal of Korean Association of Computer Education*, Vol. 25, No. 5, pp. 73-84, September 2022. <https://doi.org/10.32431/kace.2022.25.5.007>
- [21] B. Kim and H. Choi, "A Case Study on the Interaction of 3-Year-Old Infants Using Kongsooni AI TalkTalk, an AI Speaker," in *Proceedings of 2023 Fall Conference on Korean Association for Children's Media & Education*, Cheongju, pp. 127-133, November 2023.
- [22] D. H. Kim, "An Analysis of Research Trends in Artificial Intelligence for Young Children," *The Journal of Korea Open Association for Early Childhood Education*, Vol. 28, No. 1, pp. 205-222, February 2023. <https://doi.org/10.20437/KOAECE28-1-08>
- [23] B. Son and H.-J. So, "A Study on Parents' Perceptions of Acceptance Factors of the AI-Based Learning Content for Children," *The Journal of Educational Information and Media*, Vol. 28, No. 3, pp. 763-789, September 2022. <https://doi.org/10.15833/KAFEIAM.28.3.763>
- [24] S. Song, "The Study of Convergent AI Education and Its Ethical Task for Children in the Digital Age," *The Journal of Korea Open Association for Early Childhood Education*, Vol. 29, No. 1, pp. 81-100, February 2024. <https://doi.org/10.20437/KOAECE29-1-04>
- [25] K.-H. Lee, "Technological Regulations: Cases and Policy Implications," *Journal of Law & Economic Regulation*, Vol. 9, No. 2, pp. 143-160, November 2016.
- [26] W. K. Viscusi, J. M. Vernon, and J. E. Harrington Jr., *Economics of Regulation and Antitrust*, 2nd ed. Cambridge, MA: The MIT Press, 1995.
- [27] S. H. Park, B. Y. Kim, and H. Y. Kwon, "A Study on the Regulation of Artificial Intelligence in the European Union's Bill," *Public Law*, Vol. 49, No. 3, pp. 349-374, February 2021. <https://doi.org/10.38176/PublicLaw.2021.02.49.3.349>
- [28] E. M. Rogers, *Diffusion of Innovations*, New York, NY: Free Press, 1962.
- [29] J. Bang, "The Transition of Regulatory Governance on AI Algorithms -Focused on US Algorithmic Regulations and AI Ethical Principles-," *Public Law*, Vol. 49, No. 3, pp. 375-406, February 2021. <https://doi.org/10.38176/PublicLaw.2021.02.49.3.375>
- [30] D. Schiff, J. Biddle, J. Borenstein, and K. Laas, "What's Next for AI Ethics, Policy, and Governance? A Global Overview," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*, New York: NY, pp. 153-158, February 2019. <https://doi.org/10.1145/3375627.3375804>
- [31] C. Adams, P. Pente, G. Lernermeier, and G. Rockwell, "Ethical Principles for Artificial Intelligence in K-12 Education," *Computers and Education: Artificial Intelligence*, Vol. 4, 100131, 2023. <https://doi.org/10.1016/j.caeai.2023.100131>
- [32] M. Wald, "AI Data-Driven Personalisation and Disability Inclusion," *Frontiers in Artificial Intelligence*, Vol. 3, 571955, January 2021. <https://doi.org/10.3389/frai.2020.571955>
- [33] S. Y. Choi, "A Study on the Digital Competency for the Fourth Industrial Revolution," *The Journal of Korean Association of Computer Education*, Vol. 21, No. 5, pp. 25-35, September 2018. <https://doi.org/10.32431/kace.2018.21.5.003>
- [34] N. C. Corrêa, C. Galvão, J. W. Santos, C. Del Pino, E. P. Pinto, C. Barbosa, ... and N. de Oliveira, "Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance," *Patterns*, Vol. 4, No. 10, 100857, October 2023. <https://doi.org/10.1016/j.patter.2023.100857>
- [35] T. L. Saaty, *Decision Making with Dependence and Feedback: The Analytic Network Process*, Pittsburgh, PA: RWS Publications, 1996.
- [36] K. C. Park, J. W. Shin, and B. G. Lee, "Analysis of Authentication Methods for Smartphone Banking Service Using ANP," *KSII Transactions on Internet and Information Systems*, Vol. 8, No. 6, pp. 2087-2103, June 2014. <https://doi.org/10.3837/tiis.2014.06.016>

정광식(Kwang Sik Jeong)



2019년 : 한동대학교 도시환경공학, 건설공학 (공학사)
2021년 : 서울대학교 건설환경공학 (공학석사)
2023년 : 서울대학교 스마트도시공학 (공학박사 수료)

2023년~현 재: 차세대융합기술연구원 연구원
※ 관심분야 : 스마트시티, 인공지능 등

양재훈(Jaehoon Yang)



2009년 : 연세대학교 전기전자공학 (공학사)
2015년 : 서울대학교 기술경영경제정책협동과정 (공학박사 수료)

2021년~현 재: 차세대융합기술연구원 선임연구원
※ 관심분야 : 스마트시티, 바이오제약산업 등

박건철(Keon Chul Park)



2011년 : 연세대학교 정보대학원 (정보시스템 석사)
2015년 : 연세대학교 정보대학원 (정보시스템 박사)

2015년~2016년: 성균관대학교 박사후연구원
2016년~2021년: 서울디지털재단 데이터혁신팀장
2021년~현 재: 차세대융합기술연구원 도시데이터혁신연구실 실장
※ 관심분야 : 도시정보학(Urban Computing), 데이터과학 등