



‘이루다’ 서비스로 본 AI 및 알고리즘의 윤리적·법적 쟁점과 대응방안

김 법 연*

Abstract

‘이루다’ 사건으로 AI의 윤리 문제가 수면위에 드러났고, 많은 우려와 함께 대응방안에 대한 논의가 활발하게 진행되고 있다. AI의 위험성은 이미 지속적으로 경고되어 온 것으로 이 문제를 해결하기 위해 국가와 기업, 학계와 단체 등 주체를 가리지 않고 윤리적, 규범적 논의를 다양하게 실시하고 있다. AI 기술에 대한 위험에 대응하기 위해 윤리적 규범과 원칙이 제시되고 있고 법률적·규제적 통제가 본격화되고 있기도 하고 있다. 이러한 상황속에서 발생한 ‘이루다’ 사건은 우리가 해당 논의를 함에 있어 무엇을 중요하게 여겨야 하는지에 대하여 다시 한번 고민하게 되는 계기가 되었다. 이 글에서는 ‘이루다’ 사건을 되짚어보면서 AI의 위험성을 통제하고 AI 기술과 공존하는 시대에 진입한 우리사회가 무엇을 준비해야 하는지 AI와 관련된 윤리적·법적 규범이 본질적으로 다루어야 하는 것은 무엇인지 그리고 어떻게 접근하는 것이 타당한지에 대해서 고민해보고자 한다.

I. 들어가며

AI 기술과 서비스의 위험성에 대한 논의가 ‘이루다’ 사건을 계기로 매우 활발하게 이루어지고 있다. AI에 대한 윤리 문제, AI에 대한 법률적 통제에 대한 필요성은 ‘이루다’ 사건 이전에도 지속적으로 고민해오던 것이다. 그 고민의 결과로 ‘AI의 윤리원칙’이라던가 ‘AI 규제 원칙과 기준’ 등이 마련되었다. 형식과 주체는 다양하지만 대부분 AI의 위험성을 통제하기 위하여 기술자와 설계자, 사업자 등이 지키고 준수해야 할 의무들을 제시하고 있다. 그럼에도 불구하고 ‘이루다’는 우리가 예측하고 있던 문제들을 그동안 예측한 바와 전혀 다

* 고려대학교 정보보호대학원 연구교수, kby82@korea.ac.kr

르지 않게 그대로 현실에서 나타내 주었다. 이미 앞선 AI 기술들이 보여줬던 문제였기에 예측이 틀리지 않았던 결과이기도 하지만, 그러한 문제들에 대해 고민했던 결과가 제대로 효과를 발휘하고 있지 못하다는 것을 보여주는 것이기도 하다. 그동안 제시된 AI에 대한 윤리적·법적 규범들이 정말 AI의 위험을 통제하고 우리의 두려움을 해소해 줄 수 있을 것인지에 대해 더 치열한 고민이 필요한 것으로 보인다. 이에 이 글에서는 이러한 현상의 본질적 문제가 무엇인지, 그리고 현재 우리가 고려하고 다루어야 할 핵심적 과제가 무엇인지에 대해서 생각해보고자 한다.

II. ‘이루다’ 서비스에서 나타난 AI의 윤리적·법적 문제와 쟁점

1. ‘이루다’ 서비스의 내용

‘이루다’는 ‘스캐터랩’이라는 회사에서 개발한 페이스북 메신저 채팅 기반 열린 주제 대화형 인공지능 챗봇이다.¹⁾ 페이스북 메신저를 통해 일상적인 대화를 나눌 수 있는 대화형 AI(Artificial Intelligence)로, 단순히 명령을 수행하는 AI와는 다르게 사람의 대화를 구사하는 기술을 특징으로 한다.²⁾ 딥러닝으로 대화를 학습한 알고리즘이 이전 문맥을 파악하여 가장 적절한 답변을 계산하여 답하는 것인데, 카카오톡 대화 내용을 ‘업로드’하여 애정도를 분석하는 ‘연애의 과학’ 서비스를 통해 수집한 대화를 ‘이루다’ 챗봇 서비스의 학습데이터로 활용하였다.

2. 프라이버시 및 개인정보의 침해

1) 사건과 이슈의 내용

2020년 6월 베타테스트를 시작으로 12월 정식 오픈한 ‘이루다’ 챗봇 서비스는 오픈 한 달만인 2021년 1월 갖은 논란으로 서비스를 중단하게 되었다. 특히 개인정보와 프라이버시 침해에 대한 논란이 상당하였는데, ‘이루다’ 챗봇의 학습데이터인 카카오톡 대화내용의 정보주체들이 자신의 대화가 당해 서비스에 이용되고 있다는 점을 정확하게 인지하지 못하였다는 점과, 그럼에도 불구하고 대화 내용 가운데 주소, 이름, 전화번호 등 개인을 특정할 수

1) 이루다 서비스 홈페이지, 원문: <https://luda.ai/faq>, 최종검색일: 2021. 3. 10.

2) 인공지능 신문, “스캐터랩, 세계 최고 수준의 언어능력 보유한 인공지능 ‘이루다’ 정식 출시”, 2020. 12. 23.자 보도, 원문: <http://www.aitimes.kr/news/articleView.html?idxno=18758>, 최종검색일: 2021. 3. 10.

있는 정보가 포함되기도 하였다는 점 때문이다. 뿐만 아니라 이용자가 업로드한 대화 내용을 ‘이루다’ 서비스의 운영사인 ‘스캐터랩’ 임직원이 적절한 통제장치 없이 모두 열람하기도 하였다는 의혹이 제기되기도 하였다. 조금 더 자세히 살펴보면, ‘스캐터랩’이 제공하는 서비스 중 ‘연애의 과학³⁾이라는 콘텐츠 서비스에서 카카오톡 대화 데이터를 분석해서 보고서를 만들어주는 기능이 있는데, 이는 해당 기능을 이용하는 이용자들이 데이터를 업로드하면 그 데이터를 통해 개인정보를 필터링한 뒤 AI의 학습에 활용하는 방식으로 작동한다.

데이터는 실제 연인들의 대화 내용으로 약 100억 건의 데이터가 딥러닝(컴퓨터가 인간의 뇌처럼 사물이나 데이터를 분류할 수 있도록 하는 기계학습의 일종) 방식으로 학습되었다. 문제는 ‘이루다’ 챗봇 서비스에서 주소, 계좌 등을 물어보면 실제 학습된 대화의 데이터가 필터링되지 않고 그대로 노출되기도 하고, ‘이루다’가 갑자기 특정 은행의 예금주의 실명으로 추측되는 이름을 답하거나 아파트의 동과 호수까지 포함된 주소정보를 말하는 사례가 발생한 것에서 시작되었다. 뿐만 아니라 이루다 관련 개발 기록을 오픈소스 공유 플랫폼 ‘깃허브’에 공유하는 과정에서도 개인정보가 비식별화되지 않고 그대로 활용되었는데, ‘깃허브’에 올라온 카카오톡 데이터 100건에는 이름 20여 건과 더불어 대화를 나누는 사람들의 관계가 다수 포함되어 있었다.⁴⁾ 가장 심각하게 지적된 부분은 대화에 참여한 2인 중 1인에 게만 동의를 받고도 채팅 내용을 제공할 수 있어 동의하지 않은 1인은 대화 내용이 제공되고 있는지 여부도 알 수 없었다는 점이다.

2) 법률적 문제와 쟁점

‘이루다’ 서비스는 학습한 데이터 내에 개인정보가 포함되어 챗봇의 답변을 통해 개인정보가 노출이 되었다는 점, 개인정보가 정보주체가 제공받지 않는 제3의 서비스의 학습데이터로 활용된다는 것을 정보주체에게 충분히 숙지시키지 못하고 동의를 획득한 점, 연인의 대화를 학습 데이터로 활용하면서 대화의 상대방 모두에게 대화 내용에 대한 활용 동의를 획득하지 못하였다는 점 등이 이슈가 되고 있다. 즉, 인공지능의 데이터 학습 과정에서의 개인정보 및 프라이버시 침해 문제와 「개인정보보호법」 위반 여부가 쟁점이 되고 있다. ‘스캐터랩’은 ‘이루다’의 데이터 학습과정과 서비스 제공과정에서 비식별화 조치를 하였기 때문에

3) ‘연애의 과학’ 서비스에는 ‘카톡으로 보는 속마음’ 심리테스트라는 서비스가 있는데 연인이나 호감 가는 사람과 나누는 카카오톡 대화 내용을 입력하고 3,000원~5,000원 정도의 비용을 지불하면 애정도 수치를 분석해 준다(뉴스1, “집주소·계좌정보 ‘술술’... ‘AI 이루다’ 개인정보 유출 논란”, 2021. 1. 13.자 보도,

원문: <https://www.donga.com/news/It/article/all/20210113/104899076/1>, 최종검색일: 2021. 3. 10.

4) 뉴데일리 경제, “AI ‘이루다’에 쓰인 카톡 데이터, 4개월간 온라인서 공유”, 2021. 1. 13.자 보도,

원문: <http://biz.newdaily.co.kr/site/data/html/2021/01/13/2021011300160.html>, 최종검색일: 2021. 3. 10.

개인정보가 노출되거나 유출되지는 않았다고 한다. 또한 사전에 동의가 이루어진 개인정보 취급방침의 범위 내에서 개인정보를 활용하였다고 입장을 밝히고 있다.⁵⁾ ‘이루다’의 「개인정보보호법」 위반 여부가 조사 중에 있고 시민단체 등은 철저한 조사와 처벌을 주장한다.⁶⁾

‘이루다’의 「개인정보보호법」 위반 여부에서의 쟁점은 대화 내용 자체와 대화의 내용 중에 노출된 ‘개인정보’가 법률이 보호하고 있는 ‘개인정보’에 해당하는지, ‘개인정보’에 해당한다면 개인정보에 대한 활용 여부를 ‘동의’ 받는 과정에서 충분히 설명하지 못한 채 획득한 동의가 유효한 동의가 되는지, 대화의 상대방에게 ‘동의’를 받지 않은 것이 법률 위반이 되는지 등일 것이다. 이에 대해 자세히 살펴보면 ‘스캐터랩’은 ‘이루다’의 비식별화 조치에 대하여 ① 주소, 계좌번호, 전화번호 등의 개인정보가 포함될 수 있는 메시지 제거, ② 특정 단어가 실명인지 판단하는 알고리즘을 이용하여 실명이 들어갔다고 판단되는 문장을 모두 삭제, ③ 내부 테스트 및 베타테스트를 통해 발견되는 실명을 리스트로 관리하여, 해당 실명이 들어간 문장들을 모두 삭제하는 등의 조치를 하였다고 밝혔다.⁷⁾ 비식별화 조치를 시행하였지만 기계적 필터링과 예상 못한 변수들로 인해 완전한 비식별화가 되지 못한 것으로 판단된다. 따라서 이러한 문제는 사업자가 비식별화 조치를 취하였음에도 불구하고 발생하는 개인정보 노출에 대하여 어떤 법적 책임을 묻게 할 것인가에 대해 다툼이 일어나게 될 것으로 보인다. 검색된 은행 계좌번호나 주소와 같은 정보의 경우 노출되었다는 점은 심각한 문제이지만 그 외의 정보 즉, 대화의 문맥을 통해 개인을 식별해낼 확률은 매우 낮을 것으로 예상되는데, 1억 개 이상의 문장이 개별적이고 독립적인 형태로 저장되어 있고, 발화 형태에 따라 일정 확률 내에서 랜덤하게 출력되므로 ‘이루다’의 답변 내용을 조합하여 개인을 특정하는 것은 상당히 어렵기 때문이다. 개인정보의 수집 및 제공 동의에 관해서도 동의 과정에서 이용자들에게 수집 및 활용 등에 대한 충분한 설명이 없었기에 이용자들이 이를 인지하지 못하고 개인정보를 제공하였다는 점이 문제이지만, 동의 획득에 대한 법적 의무의 이행과 책임 준수 문제만을 두고 보았을 때 이것이 명백한 법률상 위법행위에 해당하는지가 법적 판단의 핵심 쟁점이 될 것으로 보인다.⁸⁾

5) 스캐터랩, “스캐터랩의 공식적인 입장을 알려드립니다.”, 2021. 1. 11.,

원문: <https://blog.pingpong.us/luda-official-statement/>, 최종검색일: 2021.3.12.

6) 조선일보, “시민단체 ‘이루다’ 개발사 개인정보법 위반... 철저히 조사·처벌해야”, 2021. 1. 13.자 보도,

원문: https://biz.chosun.com/site/data/html_dir/2021/01/13/2021011302834.html, 최종검색일: 2021. 3. 12.

7) 뉴스1, “[전문]이루다 “비식별화 조치에도 문맥 따라 인물 이름 남은 점 사과”, 2021. 1. 12.자 보도,

원문: <https://www.news1.kr/articles/?4178621>, 최종검색일: 2021.3.10.

8) 다만 우리 대법원은 2017년 대형마트가 고객의 개인정보를 수집하여 판매할 목적으로 경품행사를 진행하면서 개인정보 처리에 관한 사항을 1mm 크기의 글씨로 기재하고 동의를 받은 사안에 대해 개인정보보호법 위반이라 판단한 바 있다

3. 편향성과 기타 윤리적 문제

1) 사건과 이슈의 내용

‘개인정보’ 침해 문제뿐만 아니라 ‘이루다’는 특정 집단에 대한 편향적 견해를 나타내거나 혐오적 표현을 하여 문제가 되었다. ‘이루다’에게 성소수자를 표현하는 단어를 입력하고 의견을 물었을 때 싫어한다거나 혐오한다는 등의 부정적이고 차별적인 답변을 하기도 하였고,⁹⁾ 여성이나 장애인에 대해서도 마찬가지로 혐오발언을 하여 데이터의 편향성에 대한 논란이 제기된 것이다.¹⁰⁾

그림 1 이루다가 지하철 임신부석에 관하여 표현한 대화내용 사진 캡처본 **그림 2** 이루다가 성소수자에 관하여 표현한 대화내용 사진 캡처본



자료: SNS 등, 머니투데이 재인용(2021)¹¹⁾

자료: SNS 등, 뉴스1 재인용(2021)¹²⁾

(대법원 2017. 4. 7. 선고 2016도13263).

9) 중앙일보, “20살女 AI에 ‘레즈비언’ 꺼내자 한 말 “질 떨어져 소름끼친다”, 2021. 1. 10.자 보도, 원문: <https://news.joins.com/article/23966208>, 최종검색일: 2021.3.10.

10) 연합뉴스, “AI가 동성애·장애인 혐오?”... 이루다가 불붙인 ‘AI 윤리’ 논쟁”, 2021.1.11.자 보도, 원문: <https://www.yna.co.kr/view/AKR20210109055052017>, 최종검색일: 2021. 3. 10.

11) 머니투데이, “지하철 임신부석에 “혐오스러워”... 끝없는 AI 이루다 논란”, 2021. 1. 10.자 보도,

일부 가치판단적 질의에 대해 편향적(biased) 답변이 제시된 것에 대하여 설명을 제시하지 못할 뿐만 아니라 ‘이루다’에게 인위적 인격을 부여(의인화)하는 과정에서 발생할 수 있는 논란에 대하여 사전적 고려가 미흡하다는 지적도 있었다. ‘이루다’를 20대 여성으로 캐릭터화하였는데, 이를 상대로 선정적인 말과 음담패설을 하는 등의 성희롱적 대화를 하는 이용자가 발생하였지만 이러한 대화를 차단하지 못하였다.¹³⁾ 일부에서는 AI 서비스가 여성을 캐릭터화하였다는 것에 대하여 성역할을 고착화한다는 점에 대해 불편함을 드러내기도 한다.¹⁴⁾

2) 법률적 문제와 쟁점

‘이루다’의 특정 집단에 대한 차별적이고 혐오적 답변은 ‘AI의 편향성 문제’라고 할 수 있다. AI는 철저히 학습한 데이터를 통해 작동되고 서비스를 제공하므로 어떤 데이터를 어떻게 처리하고 분석할 것이냐에 대해 결과가 달라질 수밖에 없다. 가치판단을 필요로 하는 분석결과를 도출해야 하는 경우 이러한 편향성의 위험은 더욱 커지게 된다. 이러한 AI의 편향성의 문제에 대해서는 우리 사회가 지속적으로 위험성을 경고하고 이를 통제할 수 있는 질서체계가 만들어져야 한다고 논의하고 있지만 아직까지는 어떤 법적 책임을 부과할 수 있을 것인가에 대해 답하기 어려운 상황이다. 우리의 법체계에서는 인공지능과 그 알고리즘이 분석을 통해 결과를 도출하거나, 특정 서비스를 이용자에게 제공할 때의 위험이나 그에 대한 법적 책임에 대한 문제를 수용하지 못하고 있기 때문이다. AI와 같은 기술이 주는 변화 이전에 만들어진 체계이기 때문이기도 하지만 법률적으로 판단하기에 AI 기술 자체가 주는 복잡함 때문이기도 하다. 또한 위험의 원인과 결과를 명확하게 특정 개인이나 집단의 책임이라 하기에 어려운 구조이다. 따라서 ‘이루다’의 편향적 발언 등 문제가 된 사안을 이유로 ‘스캐터랩’에게 법적 책임 또는 법률에 근거한 처벌이나 행정적 규제 등을 하기는 쉽지 않을 것으로 보인다. 특히 인공지능이 학습한 데이터가 개발자나 사업자의 조작이나 고의에 의해 편향될 경우가 아닌, ‘이루다’ 서비스와 같이 이용자가 생산한 데이터에 근거하여 AI가 학습한 경우에 개발자나 사업자에 대해서 AI의 결과에 대한 책임을 묻는 것

원문: <https://news.mt.co.kr/mtview.php?no=2021011013422680331>, 최종검색일: 2021. 3. 10.

12) 뉴스1, “성희롱 이어 성소수자·장애인·인종혐오까지… AI챗봇 ‘이루다’ 중단 요구”, 2021. 1. 10.자 보도, 원문: <https://www.news1.kr/articles/?4176267>, 최종검색일: 2021. 3. 10.

13) 조선일보, “‘뒤편?’ “만져”... 진짜 여성 같다는 AI와 성희롱 채팅”, 2021. 1. 8.자 보도,

원문: https://www.chosun.com/economy/tech_it/2021/01/08/CKU46VV3QNB7ROWFQ7BNKLY3SA/, 최종검색일: 2021. 3. 10.

14) 뉴스민, “[민교협 시사 칼럼] 그 AI는 왜 여성이었나?: ‘이루다’ 성희롱 논란 다시 보기”, 2021. 1. 20.자 보도, 원문: <http://www.newsmmin.co.kr/news/55044/>, 최종검색일: 2021. 3. 10.

이 타당할 것인가에 대하여 결정하는 것은 더욱 복잡한 문제이다. 물론 법리적으로는 위험을 야기한 자가 책임을 지게 된다는 ‘위험책임’이나 소비자와의 균형을 유지하기 위한 관점에서 제조자에 대한 ‘제조물 책임’과 같은 법리를 통해 책임을 묻는 것이 이론적으로 가능할 수는 있지만 그렇다고 해서 어떤 책임을 얼마나 지우게 할 것인지, 그것은 과연 타당한 접근인 것인지, 이 문제를 개발자나 사업자의 책임만으로 치부할 수 있는지 등에 대한 문제는 매우 어려운 질문이다.

Ⅲ. AI에 대한 규범적 논의의 동향

1. AI 및 알고리즘에 대한 윤리적 접근

‘이루다’ 사건에서 확인한 AI의 편향성과 차별 발생, 개인정보 또는 프라이버시에 대한 침해 등에 대한 문제에 대해서는 이미 오랜 기간 지속적으로 논의가 이루어져 왔다. AI 기술이 발전하고 정교해질수록 기존의 질서 체계로는 포섭할 수 없는 현상들이 다양하게 발생하거나 그럴 것이 예측되었기 때문이다. 이와 같은 인공지능의 윤리 논의는 국가나 정부의 관심은 물론 민간의 단체나 기업 등을 통해서도 활발하게 전개되고 있는데, AI 윤리 원칙이나 가이드라인, 윤리헌장 등의 형식으로 발표되거나 공표되고 있다. 우리나라는 소프트웨어정책연구소의 ‘인공지능 연구윤리 가이드라인(2017)’, 과학기술정보통신부의 ‘지능정보사회 윤리헌장(2018)’, 방송통신위원회의 ‘이용자 중심의 지능정보사회를 위한 원칙(2019)’, 과학기술정보통신부의 ‘국가 인공지능 윤리기준(2020)’, ‘카카오 알고리즘 윤리헌장(2018)’ 등이, 미국의 경우 ‘아실로마 인공지능 원칙(Asilomar AI Principles, 2017)’, ‘ACM의 알고리즘 투명성과 책무성에 대한 성명(Statement on Algorithmic Transparency and Accountability, 2017)’ 등이 대표적이다. 유럽연합도 ‘신뢰할 수 있는 인공지능을 위한 윤리 가이드라인(Ethics Guideline for Trustworthy AI, 2019)’를 공표한 바 있다. 국내외에서 발표된 인공지능 윤리원칙의 대강의 내용은 다음과 같다.

표 1 주요 인공지능 윤리원칙 및 규범의 내용

구분	주요내용	
미국 아실로마 인공지능 원칙 ¹⁵⁾	안전 (Safety)	- AI 시스템은 작동 수명 전반에 걸쳐 안전하고 안정적이어야 하며, 적용과 실현이 가능하다면 검증할 수 있어야 한다.
	오류 투명성 (Failure Transparency)	- AI 시스템이 해를 입히는 경우 그 이유를 확인할 수 있어야 한다.
	사법의 투명성 (Judicial Transparency)	- 사법 결정에 있어 자동화된 시스템이 개입할 경우, 권한있는 인간 기관이 감사할 수 있는 충분한 설명을 제공해야 한다.
	책임성 (Responsibility)	- 고급 AI 시스템의 설계자와 제조자는 그것의 사용, 오용 및 행위의 도덕적 함의에 있어서, 그것을 형성할 책임과 기회가 있는 이해관계자이다.
	가치의 준수 (Value Alignment)	- 고도로 자율적인 AI 시스템은 그것이 작동하는 동안 목표와 행동이 인간의 가치와 반드시 일치하도록 설계되어야 한다.
	인간의 가치 (Human Values)	- AI 시스템은 인간의 존엄성, 권리, 자유 및 문화 다양성의 이상과 양립할 수 있도록 설계되고 운영되어야 한다.
	개인정보보호 (Personal Privacy)	- AI 시스템이 개인정보 데이터를 분석하고 활용할 수 있는 경우, 사람들은 자신이 생성한 데이터에 접근해 관리 및 제어할 권리를 가져야 한다.
	자유와 개인정보 (Liberty and Privacy)	- 개인정보에 대한 AI 의 적용이 사람들의 실제 또는 인지된 자유를 부당하게 침해해서는 안된다.
	이익공유 (Shared Benefit)	- AI 기술은 가능한 많은 사람들에게 혜택을 주고 역량을 강화해야 한다.
	공동번영 (Shared Prosperity)	- AI에 의해 만들어진 경제적 번영은 모든 인류에게 이익이 되도록 널리 공유되어야 한다.
	인간의 통제 (Human Control)	- 인간은 인간이 선택한 목적을 달성하기 위해, 의사 결정을 AI 시스템에 위임할 것인지 여부와 방법에 대해 선택할 수 있어야 한다.
	비전복 (Non-subversion)	- 고도로 발전된 AI 시스템의 통제를 통해 부여되는 권력은 건강한 사회가 의존하는 사회적 시민적 과정을 전복하기보다, 존중하고 개선해야 한다.
AI 무기 경쟁 (AI Arms Race)	- 치명적인 자동화 무기의 군비 경쟁은 피해야 한다.	
EU 신뢰할 수 있는 인공지능 윤리원칙 ¹⁶⁾	인간의 집행과 감시 (Human Agency and Oversight)	- 인공지능은 인간의 자율성과 의사결정을 지원해야 한다. - 인공지능은 인간의 기본적 권리를 함양하고 인간의 감시를 허용해야 한다.
	기술적 견고성 및 안전 (Technical robustness and Safety)	- 인공지능은 의도하지 않은 또는 예기치 않은 위해를 최소화하고 방지하는 한편, 인간이 의도한 대로 안정적으로 동작할 수 있도록 예방적으로 접근하여 개발하여야 한다.

15) Future of Life, Asilomar AI Principle, 2017.

16) European Commission, The Ethics Guidelines for Trustworthy Artificial Intelligence, 2019.

구분	주요내용	
EU 신뢰할 수 있는 인공지능 윤리원칙	프라이버시 보호 및 데이터 거버넌스 (Privacy and Data governance)	<ul style="list-style-type: none"> - 인공지능 시스템의 전체 수명주기 동안 프라이버시와 정보 보호를 보장해야 한다. - 인공지능으로 인한 프라이버시 침해 방지를 위해 정보의 품질과 무결성 등을 보호하는 적절한 데이터 거버넌스가 필요함. 개인정보가 불법적으로 수집되거나 부당한 차별에 사용되어서는 안된다.
	투명성 (Transparency)	<ul style="list-style-type: none"> - 인공지능 시스템에 의해 내려진 결정은 인간이 이해하고 추적할 수 있어야 함. 특히 인공지능 시스템의 결정에 의해 국민 생활에 큰 영향을 미치는 경우는 그에 대한 적절한 설명을 요구할 수 있어야 한다.
	다양성, 비차별성 및 공정성 (Diversity, Non-discrimination and Fairness)	<ul style="list-style-type: none"> - 인공지능이 사용하는 데이터는 편향, 불완전성 등이 포함될 수 있으므로 가능한 한 수집 단계에서부터 차별적 편견을 식별·제거해야 한다.
	환경 및 사회 복지 (Environmental and Societal well-being)	<ul style="list-style-type: none"> - 보다 넓은 사회 및 환경적 삶의 질에 관련된 문제들도 인공지능의 이해 관계자로 간주되어야 하며, 미래 세대를 포함한 모든 인류에게 혜택을 주기 위하여 인공지능 시스템을 이용하는 것이 이상적이다.
	책임성 (accountability)	<ul style="list-style-type: none"> - 인공지능의 개발, 배치, 사용에 있어 그 결과에 대한 책임을 지기 위한 장치가 마련되어야 한다. - 인공지능의 알고리즘, 설계 프로세스, 데이터 등에 대한 평가가 수반되어야 한다.
영국 FAST Track Principles ¹⁷⁾	공정성 (Fairness)	<ul style="list-style-type: none"> - 데이터 공정성(data fairness): 공정한 데이터 세트만 활용 - 설계 공정성(design fairness): 합리적 기능과 절차 및 분석 구조를 포함하는 모델 설계 - 결과 공정성(outcome fairness): 시스템의 결과가 차별을 낳지 않도록 방지 - 이행 공정성(implementation fairness): 시스템은 편향되지 않도록 구현
	책임성 (Accountability)	<ul style="list-style-type: none"> - 프로젝트 설계 및 구현 체계 전반의 모든 역할에 관한 책임 구조를 확립 - 프로젝트 설계 및 구현 체계 전반에 관한 감독 및 검토 등 활동 모니터링 체계 구현
	지속성 (Sustainability)	<ul style="list-style-type: none"> - 인공지능 시스템이 개인과 사회에 미칠 수 있는 변화와 영향 이해 - 인공지능 시스템이 환경에 미칠 수 있는 변화와 영향 이해
	투명성 (Transparency)	<ul style="list-style-type: none"> - 인공지능 시스템의 영향을 받는 모든 이해관계자에게 특정 상황에서 어떤 목적으로, 어떤 방식으로 결과가 나타났는지 설명 - 설계와 사용의 윤리적 허용 가능성, 비해악성, 비차별성, 전반 과정과 결과에 관한 공공 신뢰 확보
일본 AI 활용 가이드라인 ¹⁸⁾	적정한 이용	<ul style="list-style-type: none"> - 이용자는 인간과 인공지능 및 이용자 간 적절한 역할 분담 하에 적절한 범위와 방법으로 인공지능 시스템 또는 서비스를 이용하여야 한다.
	적정한 학습	<ul style="list-style-type: none"> - 이용자 및 데이터 제공자는 인공지능 시스템의 학습에 이용하는 데이터 품질을 고려하여야 한다.

17) Government Digital Service·Office for Artificial Intelligence, Understanding artificial intelligence ethics and safety, 2019.

18) AI 네트워크 사회推進會議, AI 利活用ガイドライン-AI 利活用のためのプラクティカルリファレンス, 2019.

구분	주요내용	
일본 시 활용 가이드라인	연계 가능성 고려	- 이용자 및 데이터, 서비스 제공자는 인공지능 시스템 또는 서비스 상호 간의 연계 가능성을 고려하고 위험이 확산될 수 있음을 유의하여야 한다.
	안전	- 이용자는 인공지능 시스템 또는 서비스를 활용하여 이용자 및 제3자의 생명, 신체, 재산에 위해를 끼치지 않도록 하여야 한다.
	보안	- 이용자 및 데이터 제공자는 인공지능 시스템 또는 서비스의 보안에 유의하여야 한다.
	프라이버시 보호	- 이용자 및 데이터 제공자는 인공지능 시스템 또는 서비스 활용 과정에서 자신 또는 타인의 프라이버시를 침해하지 않아야 한다.
	존엄과 자율의 존중	- 이용자는 인공지능 시스템 또는 서비스를 활용하면서 인간의 존엄과 개인의 자율을 존중하여야 한다.
	공평성	- 이용자 및 데이터, 서비스 제공자는 인공지능 시스템 또는 서비스의 판단이 편향될 수 있음을 유의하고 개인이 부당하게 차별받지 않도록 하여야 한다.
	투명성	- 이용자 및 서비스 제공자는 인공지능 시스템 또는 서비스의 검증 가능성 및 결과의 설명 가능성을 확보하여야 한다.
	책임성	- 이용자는 이해관계자에게 책임을 질 의무가 있다.
과학기술 정보통신부, 국가 인공지능 윤리기준 10대 핵심요건 ¹⁹⁾	인권보장	- 인공지능의 개발과 활용은 모든 인간에게 동등하게 부여된 권리를 존중하고, 다양한 민주적 가치와 국제 인권법 등에 명시된 권리를 보장하여야 한다. - 인공지능의 개발과 활용은 인간의 권리와 자유를 침해해서는 안 된다.
	프라이버시 보호	- 인공지능을 개발하고 활용하는 전 과정에서 개인의 프라이버시를 보호해야 한다. - 인공지능 전 생애주기에 걸쳐 개인 정보의 오용을 최소화하도록 노력해야 한다.
	다양성 존중	- 인공지능 개발 및 활용 전 단계에서 사용자의 다양성과 대표성을 반영해야 하며, 성별·연령·장애·지역·인종·종교·국가 등 개인 특성에 따른 편향과 차별을 최소화하고, 상용화된 인공지능은 모든 사람에게 공정하게 적용되어야 한다. - 사회적 약자 및 취약 계층의 인공지능 기술 및 서비스에 대한 접근성을 보장하고, 인공지능이 주는 혜택은 특정 집단이 아닌 모든 사람에게 골고루 분배되도록 노력해야 한다.
	침해금지	- 인공지능을 인간에게 직간접적인 해를 입히는 목적으로 활용해서는 안 된다. - 인공지능이 야기할 수 있는 위험과 부정적 결과에 대응 방안을 마련하도록 노력해야 한다.
	공공성	- 인공지능은 개인적 행복 추구 뿐만 아니라 사회적 공공성 증진과 인류의 공동 이익을 위해 활용해야 한다. - 인공지능은 긍정적 사회변화를 이끄는 방향으로 활용되어야 한다. - 인공지능의 순기능을 극대화하고 역기능을 최소화하기 위한 교육을 다방면으로 시행하여야 한다.
	연대성	- 다양한 집단 간의 관계 연대성을 유지하고, 미래세대를 충분히 배려하여 인공지능을 활용해야 한다. - 인공지능 전 주기에 걸쳐 다양한 주체들의 공정한 참여 기회를 보장하여야 한다. - 윤리적 인공지능의 개발 및 활용에 국제사회가 협력하도록 노력해야 한다.

구분	주요내용	
과학기술 정보통신부, 국가 인공지능 윤리기준 10대 핵심요건	데이터 관리	<ul style="list-style-type: none"> - 개인정보 등 각각의 데이터를 그 목적에 부합하도록 활용하고, 목적 외 용도로 활용하지 않아야 한다. - 데이터 수집과 활용의 전 과정에서 데이터 편향성이 최소화되도록 데이터 품질과 위험을 관리해야 한다.
	책임성	<ul style="list-style-type: none"> - 인공지능 개발 및 활용과정에서 책임주체를 설정함으로써 발생할 수 있는 피해를 최소화하도록 노력해야 한다. - 인공지능 설계 및 개발자, 서비스 제공자, 사용자 간의 책임소재를 명확히 해야 한다.
	안전성	<ul style="list-style-type: none"> - 인공지능 개발 및 활용 전 과정에 걸쳐 잠재적 위험을 방지하고 안전을 보장할 수 있도록 노력해야 한다. - 인공지능 활용 과정에서 명백한 오류 또는 침해가 발생할 때 사용자가 그 작동을 제어할 수 있는 기능을 갖추도록 노력해야 한다.
OECD 이사회 권고안 5대 원칙 ²⁰⁾	<ul style="list-style-type: none"> - 인공지능은 포용 성장, 지속가능한 개발 및 복지를 추진하여 인류와 지구의 이익을 증진하여야 한다. - 인공지능은 법치, 인권, 민주적 가치 및 다양성을 존중하는 방식으로 설계되어야 하며 공정하고 정의로운 사회를 보장하기 위해 필요한 경우 인간의 개입을 가능하게 하는 등 적절한 안전장치를 포함해야 한다. - 인공지능은 사람들이 그 결과를 이해하고 이에 도전할 수 있도록 투명성과 책임성을 보장하여야 한다. - 인공지능은 해당 시스템의 생명주기에 걸쳐 견고하고 안전하며 보안을 보장해야 하고 잠재적 위험을 지속적으로 평가하고 관리하여야 한다. - 인공지능을 개발, 운영, 배치하는 개인 또는 조직은 위의 원칙에 따라 적절한 책임을 져야 한다. 	

인공지능의 리스크와 역기능에 대한 대처방안으로 인공지능의 윤리에 대한 논의는 최근에는 더욱 구체화되고 있는데, 규제적 접근이 필요하다는 견해와 함께 법적·규제적 프레임워크가 제시되기도 한다. 윤리적 접근으로도 대응이 어려운 부분에 대해서는 규제적 방식을 통해 해결하고자 하는 것이다. 기본적으로 인공지능 사용에 따른 리스크의 문제는 구체화되기 전에는 윤리적 규범을 통해 대응을 하되, 인공지능의 사용으로 이용자의 기본권이 침해되거나 법률을 위반하게 될 경우 법적 규율을 통해 대응이 필요하다고 보는 것으로 해석된다.

2. 인공지능 및 알고리즘 규제에 대한 논의 동향과 주요내용

1) 인공지능에 대한 규제의 실시

EU는 2020년 2월 인공지능에 대한 백서(WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust)를 발간하고 AI에 대한 신뢰성을 위한 환

19) 과학기술정보통신부, 국가 인공지능 윤리기준, 2020.

20) OECD, Council Recommendation on Artificial Intelligence, 2018.

경 조성 방안으로 AI 알고리즘에 대한 법적 규제의 필요성을 강조하였다. 백서에서 강조한 규제의 필요성에 따라 유럽연합 의회는 2020년 10월 AI를 규제할 수 있는 방안을 담은 보고서를 발간하여 입법체계의 개선사항을 제시하였다. 해당 보고서에서는 기존 법률의 체계를 적용하는 방안과 새롭게 입법이 이루어져야 할 영역들에 대해서 동시에 다루고 있는데 인공지능과 알고리즘을 다루는 법률은 인공지능 시스템의 기능 변화, 공급망에서 서로 다른 경제 사업자 간의 책임 배분에 관한 불확실성, 안전성 개념의 변화에 대해 다루어야 한다고 강조한다. 기술에 대해서는 고위험 인공지능과 로봇틱스 관련 기술과 이외의 기술로 구분하면서 고위험 기술 등에 대해서는 보다 강력한 의무를 부과하는데, 이러한 의무는 규정의 대상인 인공지능, 로봇틱스 및 관련 기술들에 의해 사용되거나 생산되는 소프트웨어, 알고리즘 및 데이터를 포함한 모든 인공지능, 로봇틱스 및 관련 기술의 내용과 관련된 의무와 해당 기술의 안전성 및 보호와 관련된 의무로 구분된다. 기술의 내용과 관련된 의무로는 비편향 및 비차별 의무, 사회적 책임 및 성 평등 의무, 환경 지속가능성 의무가 있으며, 기술의 안전성 및 보호와 관련된 의무로는 인간 중심의 감독 및 통제 의무, 안전, 투명성 및 그에 의한 책임 의무, 개인정보보호 의무, 위험 평가 의무, 컴플라이언스 평가 의무가 있다.

인공지능에 대한 규제적 접근은 미국에서도 유사하게 나타나고 있다. 미국은 2016년 오바마 행정부에서 ‘인공지능의 미래를 위한 준비에 관한 보고서(Preparing for the Future of Artificial Intelligence)’를 통해 인공지능 윤리, 보안, 개인정보보호, 안전 등에 관한 주제를 포함하여야 한다고 강조한 바 있다. 2020년 트럼프 행정부는 ‘미국의 인공지능 선도에 관한 행정명령 제13859호(Executive Order on Maintaining American Leadership in Artificial Intelligence)’를 발표하고 후속조치로 백악관 과학기술정책국(Office of Science and Technology Policy, OSTP)은 ‘인공지능 적용 규제 지침(Guidance for Regulation of Artificial Intelligence Applications)’ 초안을 공개하였다.²¹⁾ 이 초안에서는 인공지능 응용 프로그램을 설계, 개발, 배치, 운용할 때 고려하여야 하는 규제 원칙을 제시하고 있으며, 원칙은 ① AI에 대한 공공 신뢰(Public Trust in AI), ② 대중 참여(Public Participation), ③ 과학적 무결성 및 정보 품질(Scientific Integrity and Information Quality), ④ 위험 평가 및 관리(Risk Assessment and Management), ⑤ 이익과 비용(Benefits and Costs), ⑥ 유연성(Flexibility), ⑦ 공정성과 비차별성(Fairness and Non-Discrimination), ⑧ 공

21) OSTP, Guidance for Regulation of Artificial Intelligence Applications, 2019. 2. 11. (Available at: https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf?utm_source=morning_brew)

개와 투명성(Disclosure and Transparency), ⑨ 안전성과 보안성(Safety and Security), ⑩ 기관 간 협력(Interagency Coordination)을 내용으로 삼고 있다.

2) 분야별 구체화

인공지능 윤리 규범에서 법적·규제적 관점에서의 논의로 흘러가던 인공지능의 규범과 관련한 논의는 최근에는 특정한 영역의 인공지능 서비스가 구체화되고 적용되면서 나타나는 현상들을 중심으로 규제가 형성되는 모습을 보이고 있기도 하다. EU는 2020년 12월 EU 디지털서비스법안(Digital Service Act)을 발표하였는데, 동 법안은 온라인 플랫폼에서 추천시스템의 투명성을 강화하는 데 그치지 않고, 이용자에게 주요 매개변수를 수정하는 등의 옵션을 제공할 것을 의무화하는 내용을 담고 있다.²²⁾ 미국의 상원에서는 인공지능과 머신러닝을 적용한 디지털콘텐츠 위조기술의 현황을 보고하도록 하는 내용을 다루는 딥페이크 보고법(Deepfake Report Act of 2019)을 제안한 바 있으며,²³⁾ 하원은 2020년 인공지능 관련 직업의 요건을 기술하고 분류하기 위한 체계 및 어휘를 개발하거나 자격증명의 기준을 만드는 등의 내용을 포함하는 인공지능 직무법(AI Carrers Act of 2020)을 제안하기도 하였다.²⁴⁾ 우리나라에서도 「온라인 플랫폼 이용자 보호에 관한 법률안」이 2020년 12월 국회에 발의되었었는데, EU와 유사하게 온라인 플랫폼 사업자는 플랫폼에서의 콘텐츠 노출 방식 및 노출 순서에 대한 결정 기준을 공개하여야 하고(안 제11조 제1항), 개인화된 기준의 적용 여부와 내용과 범위에 대한 이용자의 선택권을 보장할 것을 의무화하고 있다(안 제11조 제2항).²⁵⁾

IV. AI 규제의 근본적 한계와 접근 방법

1. AI 기술 및 서비스에 대한 법적 통제의 한계

22) European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services(Digital Service Act) and amending Directive 2000/31/EC, 2020.12.15.(Available at: <https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital>)

23) S.2065, Deepfake Report Act of 2019, 2019.7.24.(Available at: <https://www.congress.gov/bill/116th-congress/senate-bill/2065?q=%7B%22search%22%3A%5B%22Deepfake+Report+Act+of+2019%22%5D%7D&s=1&r=3>)

24) H.R.7339, AI Careers Act of 2020, 2020.6.25. (Available at: <https://www.congress.gov/bill/116th-congress/house-bill/7339/text?r=16&s=1>)

25) 의안번호: 2106369, 제안일: 2020.12.11., 제안자: 전혜숙의원 등 12인

AI의 기술이나 서비스 등의 위험을 통제하여야 한다는 필요성에 대해 공감대가 형성되었지만, 지금부터의 속제는 어떤 방식으로 대응하여야 하는가에 대한 것이다. AI의 윤리를 강조하고 기술개발과 서비스 구현에 적용하기 위한 접근과 시도를 하고 있고, 최근에는 법률적으로 성문화하는 방식으로 대응하려고 하고 있기도 하다. 법규범을 신속하게 마련하고 규제를 통해 역기능과 문제에 대응해야 한다는 관점이다. 그러나 AI를 법률적으로 규제의 방식을 통해 통제하려는 것은 여러 한계가 존재한다. AI 기술의 사용으로 어떤 작용과 결과가 도출되는지 명확하지 않은 상황에서 예측을 통해 위험을 통제하려는 것은 ‘법치’를 근간으로 하는 우리의 법원리에 반하는 결과를 초래한다. 뿐만 아니라 AI 기술과 산업 발전을 저해할 수도 있으며, 규제가 정확하게 AI 기술의 변화를 예측하고 실효적으로 적용될 수 있을지도 명확하지 않기 때문이다.²⁶⁾ 나아가 AI 기술이 가진 복잡성과 예측 불가능성은 개발자, 사업자, 사용자 등 개별 행위의 주체들이 윤리 규범의 판단과 도덕률을 준수하는 것에 기댈 수밖에 없게 한다. 이러한 점이 그간의 논의가 AI의 사회적 수용의 노력을 윤리적으로 접근하였던 이유라 할 수 있다. 반면 AI에 대한 문제를 윤리 규범에 근거하는 것만으로 해결할 수 있는 것은 아니다. 윤리는 법규범에 비해 강제적이지 못하고 단순한 윤리적 명제만으로 기술에 대한 구체적 기준을 제시해주지 못하기 때문이다.²⁷⁾ 이러한 점에서 윤리적 접근과 제도적 통제가 유기적으로 이루어져야 하며,²⁸⁾ 법규범과 윤리를 유형을 구분하고 배타적 관계로 이해하는 것을 지양해야 한다는 주장²⁹⁾은 매우 설득력이 높다고 할 수 있다.

2. AI에 대한 윤리적 접근의 필요성과 방향

우리나라를 포함한 세계 각국이 AI의 문제를 윤리 규범을 통해서 대응하려고 하는 것은 법률을 통한 접근과 규제가 AI 산업의 발전을 저해하고 글로벌 산업 시장에서 자국의 경쟁력이 상실될지도 모른다는 위기감 때문이기도 하다. 그러나 핵심적인 이유는 법과 규제라는 강제적 수단을 활용한다고 해서 AI 기술에 대한 통제가 의도한대로 이루어지지 않는다는 것을 이미 인식하고 있기 때문이다. 그렇다면 여기서 우리는 조금 더 본질적인 문제를 고민해볼 필요가 있다. 인공지능을 비롯한 사물인터넷, 빅데이터 등 소위 지능정보기술이라

26) 선지원(2020), “인공지능의 사회적 수용을 위한 국가법의 과제”, 「국가법연구」 제16집 제3호, 한국국가법학회, 14면.

27) 오테원(2020), “인공지능 윤리에서의 차별 방지”, 「The Digital Ethics」 제4권 제2호, 한국인터넷윤리학회, 23면.

28) 오테원(2017), “초연결사회의 미래규범”, 「The Digital Ethics」 제1권 제1호, 한국인터넷윤리학회, 19면.

29) 고학수·박도현·이나래(2020), “인공지능 윤리규범과 규제 거버넌스의 현황과 과제”, 「경제규제와 법」 제13권 제1호, 서울대학교 공익산업법센터, 30면.

는 새로운 기술들을 사회적으로 수용하기 위해 통제장치를 규제와 법률을 통해 마련하더라도 근본적인 문제를 해소하고 위험을 제거하는 것이 가능할 것인가? 우리가 그동안 AI의 윤리 문제, 규제의 가능성과 타당성을 검토해오고 고민해왔던 가장 큰 이유는 AI 기술의 복잡한 알고리즘으로 인해 왜 그런 결정을 도출했는지 예측할 수 없고 검증할 수 없다는 점 때문이다. 기술을 통제할 수 없을지도 모른다는 우려에서 기인하는 것이었다. 그러나 지금 우리 앞에 나타난 ‘이루다’의 현상은 이런 통제 불가능의 문제라고 볼 수는 없다. 인공지능의 핵심 요소는 알고리즘과 학습데이터이고, 이를 구성하고 움직이게 하는 것은 설계자이고 개발자인, 그리고 그 데이터를 제공한 이용자인 사람에 해당한다. 따라서 이를 설계하는 전문가와 그를 이용하는 사람에 대한 주목을 하지 않을 수 없다. 이러한 점은 ‘이루다’ 사건에서 더욱 극명해졌다. ‘이루다’가 보여준 편향성과 혐오의 문제 그리고 프라이버시 침해의 문제 모두 AI 자체의 기술적 위험성에 기인하는 것은 아니었다. ‘이루다’의 답변의 기초가 되었던 데이터는 이용자의 대화를 학습한 결과이며, 이는 현재 우리 사회의 모습을 아주 잘 보여주고 있다고 생각된다. 기술을 사용하는 사람의 수준과 역량이 ‘이루다’를 통해 고스란히 투영된 것이 아닐까 한다. 두려워할 것은 인공지능이 아니라 인공지능을 사용하는 인간이며, 그들의 도덕적 결정과 가치판단을 법률과 규제만으로 통제할 수 없다는 것을 잘 보여주고 있다고 여겨진다.³⁰⁾

국가별로 AI에 대한 문제를 윤리적으로 해결할 것인지 법률이나 규제 등의 강제적 수단을 통해서 대응할 것인지 형태도, 주체도 다양하지만 그 내용은 대동소이하다. 모두 AI 기술과 시스템에 대하여 이용자를 보호하고, 그들에게 서비스에 대하여 충분히 설명하고, 어떻게 결과를 도출하는지 투명하게 공개할 것을 핵심적 내용으로 삼고 있다. 특히 개인정보와 프라이버시 침해에 대해서는 강하게 경계하고 있으며, 보호의 필요성에 대하여 특별히 강조하고 있다. ‘이루다’의 경우 ‘스캐터랩’이 이용자에게 대화 활용에 대한 충분한 고지를 하였고, 개인정보가 전부 비식별화 조치가 잘 되어 「개인정보보호법」을 위반하지 않았거나 특정한 누군가에게 위해가 발생하지 않았다고 가정해보자. 그렇다고 해서 이용자의 프라이버시 침해 문제로부터 자유로워질 수 있는 것인가? ‘이루다’의 경우 이용자들에게 충분히 설명하지도 못하였지만, 설명을 잘 하였다 하더라도 아주 은밀하고 사적인 영역의 대화를 들여다보도록 하는 것이 그렇게 간단하게 설명이 되고, ‘누가’ 한 대화의 내용인지 모르게만 한다면 다른 누군가가 그 대화를 알 수 있게 하는 것이 과연 옳은 것인가 하

30) 윤상필(2018), “인공지능 담론의 역설과 인간윤리의 새로운 지평”, 『The Digital Ethics』 제2권 제2호, 한국인터넷 윤리학회, 63면.

는 의문이 든다. 기술을 통해 어떤 가치를 실현시킬 것인가 혹은 어떤 의미를 찾을 것인가, 무엇을 위해 기술을 활용할 것인가에 대한 진지한 고민 없이 우리가 가진 기술의 혁신이 주는 신비로움에 감탄하기만 하고 그저 편익성과 효율성만을 좇는 것은 아닌가 하는 생각이 든다.

한편, 우리나라를 포함한 세계 각국이 AI 윤리 규범과 법률들을 만들어가는 과정에 대해서도 고민해볼 필요가 있다. 기술을 받아들일 내구성이 없는 상황에서, 그리고 그러한 규범에 대한 숙의와 합의가 충분하지 않은 상황에서 규범이 질서의 축과 잣대로서 제 기능을 할 수 있을 것인가? 그저 공허한 부르짖음에 불과한 것은 아닐까. 지능정보사회라 일컬어지는 시대에 있어 이용자는 더 이상 기업과 시장에 의하여 공급되던 기술의 수요자가 아닌 기술과 함께 공진하고 참여하는 능동적 주체가 된다. 이러한 세상에서는 규범을 국가나 정부가 만들어서 하달하는 방식으로는 질서를 유지하는 데는 한계가 있다. 그동안의 윤리 규범을 만드는 과정을 보면(그것이 윤리 규범이건 원칙이건 법률이건 간에) 국가가, 정부가 지켜야 할 것들을 마련하고 이를 지켜야 하는 대상들에게 공포하는 방식으로 실시되어 왔다. 공동체의 가치 기준을 정하는 일에는 적어도 충분한 당사자의 참여를 보장하고 그 규범을 도입해야 하는 이유에 대한 판단과 도입방식, 검증 절차 등에 대해 숙의와 논의를 거쳐야만 그 규범이 성공할 수 있다.³¹⁾ 그것이 강제적이건 강제적이지 않건 중요하지 않다. 이해관계자와 이를 수용할 국가사회 구성원이 함께 고민한다면 AI에 대한 기술적 통제와 사회적 수용방식을 강제하지 않아도 지켜질 수밖에 없다. 이러한 합의가 없다면 법률과 규제 등으로 통제한다 하더라도 의도한 효과를 거두기 어렵다. 제도가 현상을 규율하지 못하는 한계에 대해서 우리는 이미 여러 경험을 통해 익히 알고 있다.

V. 맺음말

‘이루다’ 사건을 통해 확인한 것은 AI의 문제는 결국 우리 사회의 윤리의식의 결핍에서 기인한다는 점이었다. 우리나라를 포함한 세계 각국 정부는 인공지능이 불러올 변화에 대한 두려움으로 인해 기술에 대한 윤리적 규범을 만들고 법적·제도적 통제 수단을 마련하는 등 많은 노력을 하고 있다. 그러나 보다 본질적인 문제에 대하여 고민은 없었기 때문에 그렇게 수많은 노력을 하였음에도 예측한 문제가 여지없이 드러난 것이라 생각된다. 문제의 출발이 어디인지에 대한 명확한 판단은 뒤로 한 채 윤리적 접근으로도 해결이 가능한지

31) 권현영(2017), “디지털 시민주권(citizenship)과 시민참여형 인터넷윤리”, 『The Digital Ethics』 제1권 제1호, 한국인터넷윤리학회, 5면.

규제를 하는 것이 필요한지 등 외형적인 문제에 집착하고 있는 것은 아닐까 하는 의문이 든다. 오히려 현재의 상황은 국가별로 경쟁적으로 규범을 만들고 이를 선점하여 시장 경쟁의 토대로 활용하거나 질서의 축과 헤게모니를 자국으로 갖고 오기 위한 탐욕이 저변에 존재하는 것 같기도 하다. 우리는 '사람으로서 기본적으로 지켜야 할 도리'를 의미한다. 사회 구성원으로서 집단내 인간관계를 유지하기 위한 기초질서라 할 수 있다. 보다 본질적인 문제가 무엇인지를 생각하고, 이를 해결하기 위한 공동체적 논의와 숙의가 함께 이루어져야 한다. 한 사회와 공동체가 지금까지 경험해 보지 못한 변화를 수용하는 과정에서는 반드시 사회를 구성하는 구성원 모두 함께 그 과정에 참여할 수 있어야 하고 그 질서의 내용과 방법에 대해 합의하여야 한다. 그렇게 해야만 기초질서가 정착되고 보다 성숙한 사회로 나아갈 수 있을 것이다.

참고문헌

- 고학수·박도현·이나래(2020), “인공지능 윤리규범과 규제 거버넌스의 현황과 과제”, 『경제규제와 법』 제13권 제1호, 서울대학교 공익산업법센터, 7-36면.
- 권현영(2017), “디지털 시민주권(citizenship)과 시민참여형 인터넷윤리”, 『The Digital Ethics』 제1권 제1호, 한국인터넷윤리학회, 1-6면.
- 선지원(2020), “인공지능의 사회적 수용을 위한 국가법의 과제”, 『국가법연구』 제16집 제3호, 한국국가법학회, 39-65면.
- 오태원(2017), “초연결사회의 미래규범”, 『The Digital Ethics』 제1권 제1호, 한국인터넷윤리학회, 7-21면.
- _____(2020), “인공지능 윤리에서의 차별 방지”, 『The Digital Ethics』 제4권 제2호, 한국인터넷윤리학회, 18면-26면.
- 윤상필(2018), “인공지능 담론의 역설과 인간윤리의 새로운 지평”, 『The Digital Ethics』 제2권 제2호, 한국인터넷윤리학회, 59-66면.
- 방송통신위원회, 이용자 중심의 지능정보사회를 위한 원칙, 2019.
- European Commission, The Ethics Guidelines for Trustworthy Artificial Intelligence, 2019.
- European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services(Digital Service Act) and amending Directive 2000/31/EC, 2020. 12. 15.(Available at: <https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital>)

Future of Life, Asilomar AI Principle, 2017.

Government Digital Service·Office for Artificial Intelligence, Understanding artificial intelligence ethics and safety, 2019.

OECD, Council Recommendation on Artificial Intelligence, 2018.

OSTP, Guidance for Regulation of Artificial Intelligence Applications, 2019.2.11. (Available at: https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf?utm_source=morning_brew)

A I 네트워크 사회推進會議, A I 利活用ガイドライン-A I 利活用のためのプラクティカルリファレンス-, 2019.

뉴데일리경제, “AI ‘이루다’에 쓰인 카톡 데이터, 4개월간 온라인서 공유”, 2021. 1. 13.자 보도, 원문: <http://biz.newdaily.co.kr/site/data/html/2021/01/13/2021011300160.html>, 최종검색일: 2021. 3. 10.

뉴스1, ““성희롱 이어 성소수자·장애인·인종혐오까지”... AI챗봇 ‘이루다’ 중단 요구”, 2021. 1. 10.자 보도, 원문: <https://www.news1.kr/articles/?4176267>, 최종검색일: 2021. 3. 10.

뉴스1, “[전문]이루다 ‘비식별화 조치에도 문맥 따라 인물 이름 남은 점 사과’”, 2021. 1. 12.자 보도, 원문: <https://www.news1.kr/articles/?4178621>, 최종검색일: 2021. 3. 10.

뉴스1, “집주소·계좌정보 ‘술술’... ‘AI 이루다’ 개인정보 유출 논란”, 2021. 1. 13.자 보도, 원문: <https://www.donga.com/news/It/article/all/20210113/104899076/1>, 최종검색일: 2021. 3. 10.

뉴스민, “[민교협 시사 칼럼] 그 AI는 왜 여성이었나?: ‘이루다’ 성희롱 논란 다시 보기”, 2021. 1. 20.자 보도, 원문: <http://www.newsmin.co.kr/news/55044/>, 최종검색일: 2021. 3. 10.

대법원 2017. 4. 7. 선고 2016도13263).

머니투데이, “지하철 임신부석에 ‘혐오스러워’... 끝없는 AI 이루다 논란”, 2021. 1. 10.자 보도, 원문: <https://news.mt.co.kr/mtview.php?no=2021011013422680331>, 최종검색일: 2021. 3. 10.

연합뉴스, ““AI가 동성애·장애인 혐오?”... 이루다가 불붙인 ‘AI 윤리’ 논쟁”, 2021. 1. 11.자 보도, 원문: <https://www.yna.co.kr/view/AKR20210109055052017>, 최종검색일: 2021. 3. 10.

인공지능신문, “스캐터랩, 세계 최고 수준의 언어능력 보유한 인공지능 ‘이루다’ 정식 출시”, 2020. 12. 23.자 보도, 원문: <http://www.aitimes.kr/news/articleView.html?idxno=18758>, 최종검색일: 2021.3.10.

조선일보, ““뭐해?” “만져”... 진짜 여성 같다는 AI와 성희롱 채팅”, 2021. 1. 8.자 보도, 원문: https://www.chosun.com/economy/tech_it/2021/01/08/CKU46VV3QNB7ROWFQ7BNKLY3SA/, 최종검색일: 2021. 3. 10.

조선일보, “시민단체 “이루다 개발사 개인정보법 위반... 철저히 조사·처벌해야””, 2021. 1. 13.자 보

도, 원문: https://biz.chosun.com/site/data/html_dir/2021/01/13/2021011302834.html, 최종검색일: 2021. 3. 12.

중앙일보, “20살女 AI에 ‘레즈비언’ 꺼내자 한 말 “질 떨어져 소름끼친다””, 2021. 1. 10.자 보도, 원문: <https://news.joins.com/article/23966208>, 최종검색일: 2021. 3. 10.

이루다서비스홈페이지, 원문: <https://luda.ai/faq>, 최종검색일: 2021. 3. 10.

스캐터랩, “스캐터랩의 공식적인 입장을 알려드립니다.”, 2021. 1. 11., 원문: <https://blog.pingpong.us/luda-official-statement/>, 최종검색일: 2021. 3. 12.