인공지능 시스템 이용자 보호 관련 해외 규율 동향과 시사점*

선 지 원**

Abstract

인공지능에 대한 규율을 어떤 방식으로 해야 할 것인지에 대해서는 의견이 엇갈리지만, 현실적으로 제22대 국회가 개원한 지 얼마 되지 않은 시점임에도, 각종의 인공지능 관련 입법이 발의된 상태이다. 이 글에서는 국외의 인공지능 관련 규율들의 사례를 통해 우리나라에서의 인공지능 관련 입법의 방향성을 검토해 보고자 한다. 유럽연합은 과거 윤리적인 접근 위주로 인공지능을 제어하던 태도에서, 이른바 AI Act라고 불리는 입법을 통해 인공지능 시스템을 직접적이고 포괄적으로 규율하려 시도한다. OECD, IEEE 및 UNESCO 등의 국제기구들은 각 기구가 지향하는 가치관과 목표에 따라 인공지능을 규율하고 있으나, 그 규율 정도는 연성 규범에 그친다. 미국은 최근 대통령이 발하는 행정명령을 통해 인공지능에 대해 규율하고 있을 뿐만 아니라, 몇몇 법안 역시 발의되어 있는 상황이다. 점차 많은 법제들이 인공지능에 대한 규율을 제정하고 있는 상황 속에서, 우리나라 법 문화와 인공지능 관련 시장의 특성을 반영하여 슬기로운 입법 정책의 고려가 필요할 것으로 보인다.

I. 머리말

어떠한 제품이나 서비스를 막론하고 주어진 환경 아래에서 이용자가 해당 제품 및 서비스를 신뢰하고 안전하게 사용하거나 영위할 수 있도록 보장하는 일은 중요하다. 이는 제품이나 서비스를 제공하는 기업의 입장에서는 사업을 유지하고 확장성을 가지기 위한 조건이 되지만, 법제도적인 측면에서는 일반 시민에 해당하는 이용자를 법질서가 보호한다는 점에

* 이 글은 2024년 한국공법학자대회에서 발제한 것의 일부를 대폭 수정 및 보완한 것이며, 관련한 내용이 한국법제연구원의 연구보고서 〈인공지능시대 미래사회 변화에 대응하는 공법의 과제〉(2024. 12. 31. 발간 예정)에 수록될 예정입니다.
** 한양대학교 법학전문대학원 조교수, 법학박사, sunji1@hanyang.ac.kr

서 의미를 가진다. 따라서 이용자 보호는 법적인 측면이나 정책 영역뿐만 아니라, 기업의 활동 영역에서도 중요한 과제로 인식되어 왔다. 이에 따라 일반적인 소비자보호법제뿐만 아니라 많은 영역에서 특유의 이용자 보호 제도들이 발달하고 있다. 방송통신의 영역에서는 전기통신사업법을 비롯한 관련 법제들이 이용자 보호에 대해 규율하고 있다.[1] 주로 통신서비스를 매개로 이루어지는 인공지능 시스템에 대한 이용자 보호 역시 인공지능에 대한 일반법이 존재하고 있지 않은 우리나라의 법적 상황 속에서는 전기통신사업법을 근거로 한 이용자 보호의 거버넌스에 포함될 수 있을 것이다. 특히나 방송통신정책이 더 이상 네트워크의 관리에 머무르지 않고, 통신서비스 안에서의 콘텐츠 향유나 이용자의 행태 — 특히, 적극적인 행태 — 에 관심을 두게 되는 변화를 주목할 필요가 있다. 이에 따라 그 안에서 서비스 제공의 수단으로 활용하게 되는 인공지능 시스템과 같은 범용기술에 대한 규율이 이용자 보호 차원에서 더욱 중요해진 것이다.[2]

일반적인 인공지능 시스템뿐만 아니라 생성형 AI를 기반으로 한 시스템 역시 이용자 보호의 대상이 된다. 생성형 AI의 대두는 사실 통신 기술, 클라우드 및 하드웨어의 발달에 기인한 것으로 추론할 수 있다. 생성형 AI의 기반인 대규모언어모델을 훈련시키기 위해 필요한 물적 부담이 기술의 발달로 인해 대폭 감소한 것이다. 따라서 대규모의 데이터 처리와 학습 모델의 축적이 생성형 AI를 가능케 하였고, 기존에는 상상 속에서만 머물렀고 현실화되리라고 생각하기 어려웠던 이른바 인공일반지능(Artificial General Intelligence: AGI)으로의 발전 가능성 역시 운위되고 있는 상황이다.[3] 네트워크 기술의 발전에 따라 방송통신의 이용자들은 과거보다 개별화되고 능동적인 태도를 보여주고 있다.[4] 그러나 생성형 AI의 경우 이용자의 개별성은 더욱 강화될 수 있지만, 이용자의 능동성을 강화한다고

---

1) 전기통신사업법 제1조는 동법의 목적 중 하나로 "이용자의 편의를 도모함"을 들고 있으며, 이용약관의 신고(제28조) 및 이용자 보호(제32조)에 대한 직접적인 규정을 두고 있다. 뿐만 아니라 금지행위(제50조)가 성립하기 위한 요건 중 하나로 "이용자의 이익을 해치거나 해칠 우려가 있는" 경우를 들고 있기도 하다. 요컨대 통신서비스에서의 이용자 보호는 전기통신사업법의 주요 목적 중 하나이며, 우리나라 통신법의 규율의 한 축이라고 평가할 수 있다. 통신서비스에 대한 지금까지의 이용자 보호 논의에 대해서는 김태오, 통신시장 이용자보호의 쟁점과 과제, 경제규제와 법 제2권 제1호, 서울대학교 공익산업법센터, 2009, 257쪽 이하; 박종수, ICT융합에 따른 방송통신 이용자보호 법제의 합리적 개선방안, 법제연구 제44호, 한국법제연구원, 2013, 108쪽 이하; 최경진/정경오/이종관, 통신시장 이용자 보호를 위한 법제분석 – 효율적인 피해구제 방안을 중심으로 –, 한국법제연구원, 2015, 25쪽 이하.

2) 이원태 외, 지능정보화 이용자 기반 보호 환경 조성 총괄보고서, 정보통신정책연구원, 2018, 15쪽 이하.

3) AI 연구의 방향성이 AGI로 옮겨가고 있다는 분석이 있다. Susnjak et. al., Over the Edge of Chaos? Excess Complexity as a Roadblock to Artificial General Intelligence, arXiv:2407.03652 [cs.AI], 2024. "생성형 AI의 한계를 극복하고, 인간의 지능을 전반적으로 구현할 수 있는 AGI의 개발이 필요"하다고 말하는 문헌도 있다. 이화민, 생성형 AI를 넘어 AGI 시대로, 전기의 세계 제72권 제12호, 대한전기학회, 2023, 단, 생성형 AI 혹은 기반모델을 본격적으로 겨냥한 법적·윤리적 규범은 아직까지 본격적으로 제시되지 않고 있는 것으로 보인다.

4) 문화형성자 및 생산자로서의 이용자 성격의 변화에 대해서는 이원태 외, 앞의 보고서, 16쪽 참조.

말할 수 있을 것인지는 의문이다. 우선 신기술의 도입 초기 단계로서 이용자의 경험과 역량이 성숙되지 않았다는 문제도 있을 것이다. 그러나 무엇보다도 데이터에 대한 심층적인 분석을 기반으로 이용자 개개인에게 밀착한 서비스를 제공한다는 관점에서 이용자의 능동성의 필요성 자체를 제거한다고도 볼 수 있다. 이에 따라 이용자 스스로의 오남용으로 인해 발생하는 해악 역시 서비스 제공자의 이용자 보호 의무의 범주에서 다루어지는 문제점으로 발생할 수 있을 것이다.

이용자 보호의 관점에서도 중요한 것은 이용자가 해당 서비스를 지속 가능하게 이용할 수 있도록 하는 것이다. 따라서 인공지능 시스템의 이용이 야기할 수 있는 리스크를 적절히 제어하여, 인공지능 관련 생태계가 인류에게 도움이 되는 방향으로 발전할 수 있는 방향성을 제시하는 일이 중요하다. 우리나라에서도 지난 제21대 국회에서 다수의 인공지능 관련 법안이 발의된 바 있으며[5], 제22대 국회에서도 마찬가지의 입법 시도가 있을 것으로 보인다.[6] 그러나 명확한 목적성을 가지고 인공지능의 리스크 제어와 인공지능 기반 산업 진흥 양면에서 모순점이 없는 입법을 하기 위해서는 충분한 논의와 고찰이 필요하다. 종합적으로 우리 법질서가 인공지능에 대해 어떠한 태도를 취해야 할지 고찰하기 위해서는, 기존의 규범들을 살펴보는 일이 유용할 것이다. 이하에서 각국에서 입법하거나 발의 또는 논의 중인 기존의 인공지능 관련 규범들과 윤리적 의제들을 통해 인공지능 시스템과 관련한 이용자 보호 문제에 적용 가능한 법리들을 살펴보고, 우리나라에서의 규범 논의를 위한 시사점을 도출하고자 한다.

---

5) 영역과 주제별로 대표적인 법안들을 꼽자면 「인공지능교육진흥법안」, 「한국인공지능·반도체공과대학교법안」, 「인공지능 책임 및 규제법안」, 「알고리즘 및 인공지능에 관한 법률안」, 「인공지능산업 육성 및 신뢰 확보에 관한 법률안」등이 있다.

6) 제22대 국회가 개원한지 세 달도 채 되지 않은 2024년 8월 31일 현재 이미 아홉 건의 인공지능 관련 법안이 발의되었다. 아래는 국회 의안정보시스템(https://likms.assembly.go.kr/)에서 "인공지능"이라는 검색어로 계류의안을 검색한 결과이다.

의안명	제안자명	제안일자
인공지능 발전 진흥과 사회적 책임에 관한 법률안	배준영의원 등 10인	2024-08-28
인공지능책임법안	황희의원 등 10인	2024-08-27
인공지능 기본법안	한민수의원 등 10인	2024-08-22
인공지능 개발 및 이용 등에 관한 법률안	권칠승의원 등 15인	2024-07-04
인공지능기술 기본법안	민형배의원 등 13인	2024-06-28
인공지능산업 육성 및 신뢰 확보에 관한 법률안	김성원의원 등 11인	2024-06-19
인공지능산업 육성 및 신뢰 확보에 관한 법률안	조인철의원 등 19인	2024-06-19
인공지능 발전과 신뢰 기반 조성 등에 관한 법률안	정점식의원 등 108인	2024-06-17
인공지능 산업 육성 및 신뢰 확보에 관한 법률안	안철수의원 등 12인	2024-05-31

Ⅱ. 해외에서의 인공지능 시스템에 대한 규율 사례

1. 법질서가 인공지능을 바라보는 시각

생성형 AI에 대해서는 아직 제도적인 대응이 불충분한 상황이지만, 기존의 인공지능 관련 규율 논의들의 추이를 통해 생성형 AI를 포함한 인공지능 시스템에 대한 규율 방향성을 예측할 수 있다. 인공지능 관련 규율은 크게 두 가지의 방향성으로 대별할 수 있다. 첫 번째는 인공지능을 경제 발전의 동력으로 인식하고, 인공지능을 사회적으로 수용함에 있어서 장애가 되는 리스크들을 제거하기 위해 공권력과 법제도를 활용하는 것이다. 두 번째 시각은 인공지능이 가져오는 리스크가 인류에게 미치는 해악을 방지하는 것을 국가와 법제도의 가장 중요한 임무로 인식하는 것이다. 기술의 보유 정도나 원천 기술에 대한 접근성의 유무 및 기존 사회에서의 인공지능의 사회적 수용성의 정도에 따라 주체별로 다른 태도를 가지는 점을 발견할 수 있다.

최근 주목받고 있는 생성형 AI에 대해서도 이와 같은 방향성 분화는 법제에 따라 나타날 것으로 보인다. 여전히 생성형 AI를 통해 산업 발전과 기술 개발을 촉진하고자 하는 시각에서는 생성형 AI가 기존의 인공지능에 비해 가지는 특수성에 주목하여, 그 활용 가능성을 높이려는 시도를 법제도에 담을 가능성이 높다. 즉, 대규모 언어모델의 활용 폭을 높이기 위해 필요한 데이터 공유의 허용성 등을 더욱 확대할 가능성이 높다. 반면에 생성형 AI의 윤리적 문제를 중시하는 입장에서라면 기존 인공지능 시스템과는 달리 생성형 AI 시스템에서 나타날 수 있는 새로운 리스크들을 제어하기 위한 수단으로서 법제도를 이용하려고 할 것이다. 이하에서 유럽연합, 국제기구들 및 미국을 포함한 주요국의 최근 인공지능 규율 동향에 대해 살펴본다.

2. 유럽연합의 윤리적 접근에서 법적 규제로 변화

유럽연합의 인공지능 관련 규범은 Jean-Claude Juncker 집행위원장 시기(2014년~2019년)에 주로 윤리적 내지는 정책적 수단으로 접근하던 기조에서, Ursula von der Leyen 집행위원장 시기(2019년 11월~현재)에 들어서 법적 규제로 변화한 것을 발견할 수 있다. 2019년 집행위원회가 전문가들을 초청하여 구성한 "인공지능 최고 전문가그룹(High-Level Expert Group on AI: HLEG)"은 "신뢰할 수 있는 인공지능을 위한 윤리 가이드라인(Ethics guidelines for trustworthy AI)"[7]을 발표한다. 여기에는 윤리적 원칙과 관련된 가

치[8]를 비롯하여 인공지능에 대한 신뢰를 위해 필요한 7가지 핵심 요건[9]과 신뢰성 있는 AI 평가 목록을 담고 있다.

2019년 12월에 취임한 Von der Leyen 집행위원장은 디지털 분야에서의 적극적인 법적 규율 의지를 천명하고, 법적 규율의 준비 작업으로서 우선 「인공지능 백서(White Paper on Artificial Intelligence)」(2020년)를 발간한다. 이 백서에서 집행위원회는 인공지능 분야에서 유럽연합 시민의 가치와 권리를 존중한다는 궁극적인 목표 아래, 윤리적이고 신뢰할 수 있는 AI의 개발 및 활용과 더불어 AI 분야 유럽의 혁신 역량 증진을 위한 정책 방향을 제시하고 있다. 아울러 인간 중심 인공지능(Human-Centric AI), 즉, 인공지능의 개발과 사용을 인간의 복리 추구에 이바지하는 방향으로 유도하는 것 역시 정책 목표의 하나로 제시하고 있다. 또한 유럽 내 단일 인공지능 생태계 조성을 위해 i) 우수한 생태계(ecosystem of excellence) 조성을 위한 프레임워크, ii) 신뢰의 생태계 조성을 위한 규제 프레임워크를 제시하고 있다. 이 규제 프레임워크에서는 학습데이터에 대한 관리에서 시스템 자체의 견고성과 안정성 및 정확성의 보장 및 인간의 감시 및 개입 보장에 이르기까지의 규제 체계를 제안한다. 특히, 인공지능을 활용한 원격 생체 인식 기술 적용과 같이 리스크가 큰 시스템의 사용에 대해서는 엄격하게 제한할 것을 제안한다. 유럽연합 데이터 보호 규정(GDPR)에 따라 생체(biometric) 정보의 처리는 원칙적으로 금지되며, 예외적으로 법률의 근거에 따라 상당한 공익상(substantial public interest)의 이유가 있는 경우에 한하여 처리를 허용할 수 있으나, 인공지능 시스템을 통한 활용은 보다 엄격하게 규제하고자 했던 것이다. 이러한 규제 프레임워크의 적용 대상으로는 기본적으로 사전 적합성평가를 통해

---

7) 해당 가이드라인의 전문은 다음 사이트에서 확인할 수 있다. Ethics guidelines for trustworthy AI | Shaping Europe's digital future 〈https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai〉(2024. 8. 31. 최종 접속). 가이드라인의 내용과 그에 대한 비평에 대해서는 선지원, 유럽 HLEG 인공지능 윤리 가이드라인과 지능정보사회 이용자보호 정책의 비교, The Digital Ethics vol. 3, 2019, 59쪽 이하를 참조하라.

8) 시스템 전주기에 걸쳐 반드시 지켜져야 하는 세 가지 구성 요소로 (1) 합법적이어야 하며 모든 관련 법규를 준수해야 하고, (2) 윤리적이어야 하며 윤리적 원칙과 가치를 준수해야 하며, (3) 좋은 의도로 AI 시스템이 의도하지 않은 결과를 초래할 수 있기 때문에 기술 및 사회적 관점 모두에서 견고해야 함을 요구하고 있다.

9) '신뢰 가능한 AI(Trustworthy AI)'를 구현하기 위한 요건은 다음과 같다.
- 인간 자율성 및 감독(Human agency and oversight)
- 기술적 견고함 및 안전(Technical robustness and safety)
- 프라이버시와 데이터 거버넌스(Privacy and data governance)
- 투명성(Transparency)
- 다양성, 차별금지, 공평성(Diversity, non-discrimination and fairness)
- 사회 및 환경 복지(Soceital and environmental wellbeing)
- 책임성(Accountability) 원칙

판명된 고(高) 리스크 AI 애플리케이션을 제시하고 있으나, 그밖의 애플리케이션에 대해서도 자발적 Labeling을 유도한다는 방침을 밝히고 있다.

유럽연합은 이 백서의 내용에 기반한 회원국 및 이해관계자들의 의견 수렴을 거쳐 이른바 "AI Act"[10]를 입법한다.[11] 2021년에 발의되어 2024년에 의회와 평의회의 의결을 마친 AI Act는 인공지능 개발 및 사용을 위한 공통된 조건을 만들어 유럽 단일 시장의 기능을 보장하기 위한 목적에서 인공지능에 대한 리스크 기반의 규제를 골자로 하는 규제입법이다. 인공지능에 대한 포괄적인 규율을 하는 입법례로서는 세계 최초의 사례로 볼 수 있다. AI Act는 1) 안전하고 기존 EU 법률을 존중하며, 2) 투자를 촉진하기 위해 법적 안정성을 보장하고, 3) 유럽법 차원의 기본권과 안전 요건 보장을 위해 규제를 효과적으로 집행할 수 있는 구조를 조성하고, 4) 합법적이고 안전하며 신뢰할 수 있는 AI 애플리케이션을 위한 단일 시장의 성장을 촉진할 것을 목적으로 하는 제2차법[12]에 해당하는 규정이다. 그 핵심 내용은 이미 위의 AI 백서에서 밝혔던 대로 리스크 기반의 통제라고 할 수 있다. 즉, 인공지능 시스템에 대해 리스크별로 분류한 후, 리스크의 정도에 따라 규제의 정도를 달리하는 것이다. 우선 허용할 수 없는 정도의 리스크(unacceptable risk)를 안고 있는 인공지능 시스템(예컨대 잠재의식을 조작하는 시스템, 인간의 사회적 지위 등을 포괄적으로 평가하는 시스템, 실시간으로 원격 생체정보를 식별하는 시스템 등)은 전면적으로 이용을 금지한다. 높은 리스크(High Risk)를 가진 시스템(예: 신용평가, 건강 및 생명보험, 투표 집계 시스템, 의료 분야에서의 환자 분류 등)은 엄격한 규제를 통해 제한된 조건 아래에서만 활용이 가능하도록 하고, 활용 이후 사후 규제 역시 강화하는 것이다. 높은 리스크를 가진 인공지능 시스템에 대해서는 기본권영향평가 및 제3자 혹은 자율 적합성 평가를 거쳐야만 시장에 출시가 가능하도록 규율하고 있으며, 출시 후에도 다양한 의무들을 통해 리스크를 관리할 것을 요구하고 있다. 반면 제한된 리스크(Limited Risk)만 가지고 있는 시스템에 대해서는

---

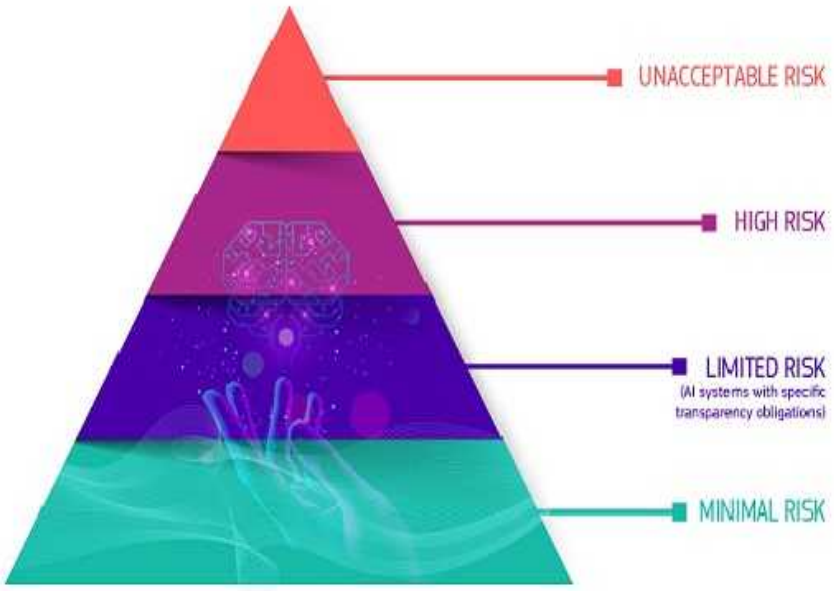
10) Proposal for Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence(Artificial Intelligence Act) and amending certain Union legislative acts, Brussels, 26 January 2024.

11) 법안의 내용에 대해 최종 의회 채택안을 기준으로 설명하고 있는 국내 문헌으로 라기원, 유럽연합 인공지능법(EU AI ACT)의 특성과 쟁점 – 우리나라 인공지능 입법에 대한 시사, 규제혁신법제연구 이슈페이퍼 24-20-3, 한국법제연구원, 2024.

12) 유럽연합의 제2차법(Secondary Law)이라 함은 제1차법(Primary Law)에 해당하는 조약에 근거하여, 유럽연합의 입법자인 유럽 의회와 유럽연합 평의회(Council of the European Union)가 발령하는 규범들을 의미한다. 대표적인 것으로 회원국에게 해당 내용의 내국법 입법 의무를 부여하는 지침(Directive), 회원국의 입법을 거치지 않고 회원국 국민에게 직접 적용되는 명령(Regulation), 개별·구체적인 사안에 대해 적용되는 결정(Decision) 등이 있다. Oppermann/Classen/Nettesheim, Europarecht 7. Auflage, 2016, 118 ff.

사고 발생 시 사후 대처가 가능하도록 하는 투명성 의무(관련 정보에 대한 고지나 공개)를 부과하고, 최소의 리스크를 가졌거나 리스크가 없는 것으로 평가되는(Minimal or no Risk) 인공지능 시스템에 대해서는 자유로운 사용을 보장한다. 그밖에도 범용인공지능에 대해서는 그 모델의 제공자에 대해 학습 과정 등에 대한 기술적 문서를 작성할 의무를 포함한 특수한 의무를 부여하고 있다.

그림 1 AI Act의 리스크 관리 체계도



자료: AI Act | Shaping Europe's digital future 홈페이지 (https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai) 그림 인용

AI법이 예정하는 인공지능 규율 거버넌스는 유럽 AI 위원회(European Artificial Intelligence Board), 독립한 전문가 패널 및 회원국 내국의 경쟁당국으로 삼분된다. 아울러 2024년 2월에 이미 집행위원회 산하에 유럽 인공지능청(European AI Office)을 설치하여 규칙의 집행과 적용을 준비하고 있다.

유럽연합은 여전히 인공지능을 장래의 성장 동력으로 삼으려는 정책 기조를 유지하면서도, 더욱 안전하고 더 신뢰할 수 있는 인공지능 구현을 위해 향후에도 그 오용으로 인한 리스크에 적극 대처하겠다는 입장을 보이고 있다. 기존의 인공지능법뿐만 아니라 향후 산업에서의 AI 활용에 관한 이니셔티브를 담은 입법 움직임이 있을 것으로 보이며, 인공지능 연구 이사회(European AI Research Council) 등의 거버넌스도 집행위원회 차원에서 준비

하고 있다고 공표하였다.[13] 명시적인 언급은 없지만 생성형 인공지능에 대한 대응 역시 인공지능 기술 발전에 발맞추어 더 적극적으로 이루어질 것으로 보인다.

단, 유럽연합의 입법 기조에 대해서는 리스크 기반의 포괄적이고 종합적인 접근법이 인공지능 기술의 이용 형태와 부합하지 않는다는 비판이 가능하다. 즉, 애플리케이션의 특성을 고려하지 않고 인공지능 시스템 자체의 리스크에 주목한 나머지, 실제로 해당 기술을 사용했을 때 나타나는 리스크의 정도나 모습이 영역별로 달라질 수 있다는 점을 간과할 수 있는 것이다. 예컨대 같은 대규모 언어모델 기반으로 웹상의 텍스트로부터 정보를 취합하여 결론을 내는 형태의 시스템이라 하더라도, 이를 단순히 대화를 위한 챗봇에 적용하느냐, 법률서비스에 적용하느냐에 따라 야기할 수 있는 리스크가 달라질 수 있다. 이러한 차이에 대한 고려 없이 시스템을 중심으로 리스크를 분류할 경우 애플리케이션에 따라 모순된 규제가 발생할 가능성도 있어 보인다. 향후의 인공지능 입법 논의에서 충분히 고려해야 하는 사안으로 보인다.

3. 주요 국제기구 및 단체들의 연성규범을 통한 접근

주요 국제기구들과 단체들 역시 구속력이 있는 법규가 아닌 권고안 등의 연성규범을 통해 국제사회 혹은 회원국에게 인공지능 정책의 방향성을 제시하고 있다.

경제협력개발기구(OECD)는 디지털 변혁의 시대에서 지속가능한 발전과 인공지능 활용을 위해 OECD 회원국 간의 논의를 거쳐 2019년 5월에 「OECD 인공지능 이사회 권고」(OECD Recommendation of the Council on Artificial Intelligence)를 발표하였다.[14] 향후 회원국들의 경제개발정책의 새로운 성장 동력으로서 인공지능의 중요성을 인정하고, 역기능을 최소화한 채로 인공지능을 최대한 활용하기 위한 전략을 제시하는 내용이다. OECD는 인공지능을 통한 포괄적 성장, 지속가능한 발전 및 삶의 질(웰빙)(Inclusive growth, sustainable development and well-being)을 달성하기 위해 몇 가지 개별 원칙들을 제시하고 있다. 인간 중심적 가치와 공정성(Human-centered values and fairness), 투명성과 설명가능성(Transparency and explainability), 견고성, 보안 및 안전성(Robustness, security and safety) 및 책임성(Accountability)이 이에 해당한다. OECD는 지난 2024년 5월, 생성형

13) 이는 2024년에서 2029년까지 집행위원장 2기 임기를 맞이한 Ursula von der Leyen의 정책 가이드라인에 드러난다. Political Guidelines | Ursula von der Leyen - Candidate for the European Commission President, 18. 7. 2024. p. 10.

14) OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.

인공지능을 비롯한 기술의 발전 사항들을 반영하여 본 권고안의 내용을 개정하여 발표하였다. 여전히 신뢰할 수 있는 인공지능을 위한 책임 있는 관리 임무(responsible stewardship)를 요구한다는 기조를 유지하면서, 권고의 대상을 "인공지능 시스템"으로 변경하고 있다. 또한 생성형 인공지능의 맥락에서는 정보의 무결성을 유지할 것을 요구하고 있다.[15)]

전기전자기술자협회(IEEE)는 지난 2019년 9월, 산하 프로그램 중 하나인 "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems"를 통해 인공지능 윤리에 대해 기술자 및 기술 개발 기업이 준수해야 하는 사항을 담은 "Ethically Aligned Design"이라는 윤리적 설계를 위한 문서를 발간했다. A/IS[16)]가 국제적으로 승인된 인권을 존중하고 고양시키며 보호하기 위한 목적으로 개발 및 운영되어야 한다는 인간 권리에 대한 내용을 포함하여, 인간 생활 웰빙의 개선, 데이터 에이전시, 효율성, 투명성, 책임성, 오남용의 인식을 설계의 목표로 삼아야 한다는 점을 제시하고 있다. 또한 개발자들이 A/IS의 안전하고 효율적인 운용을 위해 필요한 지식과 기술을 명시할 의무도 부여하고 있다.

국제연합교육과학문화기구(United Nations Educational Scientific and Cultural Organization: UNESCO) 역시 국제적인 교육, 과학 및 문화의 가치를 인공지능의 개발과 사용에 대해 관철할 목적으로 권고안(The Recommendation on the Ethics of Artificial Intelligence)을 2020년에 발표한다.[17)] 유네스코가 추구하는 가치에 맞추어 인공지능 윤리를 위한 가치로 인간의 존엄성, 인권 및 근본적 자유, 소외된 사람이 없을 것, 조화로운 삶, 신뢰가능성 및 환경보호를 제시하고 있다. 이를 달성하기 위한 실천 규범은 두 가지 범주로 제시한다. 첫 번째 범주는 인간과 인공지능 간의 관계로서, 인간과 인간의 번영, 비례성, 인간의 관리감독 및 결정, 지속가능성, 다양성과 포용성, 개인정보 보호, 인식과 교율 및 다중이해관계자 및 적응형 거버넌스를 세부 원칙으로 제시한다. 두 번째 범주는 인공지능 시스템 자체에 대한 내용으로 공정성, 투명성 및 설명가능성, 안전과 보안 및 책임과 책무성 등을 강조한다.

---

15) 관련 내용에 대해서는 OECD Legal Instruments 〈https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#backgroundInformation〉(2024. 8. 31. 최종 접속)

16) IEEE는 인공지능(Artificial Intelligence)이라는 용어 대신 자동화/지능화 시스템(Autonomous and Intelligent System)이라는 용어를 사용한다.

17) 이 권고안의 전문을 번역하고, 해설을 덧붙여 국내에 출간한 문헌이 있다. 이상욱, 유네스코 인공지능(AI) 윤리 권고 해설서 – 인공지능 윤리 이해하기, 유네스코한국위원회, 2021.

4. 미국의 법안과 정부입법

미국 정부와 의회 역시 다양한 형태의 인공지능 관련 규범을 입법하거나 입법하기 위해 시도하고 있다. 연성규범 및 행정입법 형태로는 몇몇 규범들이 발효되었으나, 유럽연합과 달리 법률 차원의 규범은 아직 제정되지 않은 상태이다. 본래 미국 정부의 기조는 내국 기업들의 우월한 인공지능 관련 기술을 바탕으로 인공지능의 활용성을 최대한 확장하는 것이었으며, 오히려 미국 내 IT 기업과 학계 등이 인공지능 윤리 규범의 정립에 앞장서고 있던 상황이었다.[18] 그러나 최근 생성형 인공지능 등 기술의 확대로 인해 미국 정부와 입법자 역시 인공지능 시스템의 리스크 제어에 어느 정도 시선을 두고 있는 것으로 보인다.

우선 제46대 대통령인 Joe Biden 정부 초기인 2020년 12월에 "연방정부의 신뢰할 수 있는 인공지능 사용을 도모하기 위한 행정명령(Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government)"을 발표한 것을 본격적인 인공지능 규범의 시작으로 볼 수 있다. 이 행정명령은 연방정부 내에서의 인공지능 사용에 대한 통제를 목적으로 하는 것으로, 민간 영역에서의 인공지능 사용을 규율하는 것은 아니다. 이 명령에서는 인공지능 사용의 원칙을 a) 합법적이면서 미국의 가치를 존중하는 사용, b) 목적 및 성과 중심의 사용, c) 정확하고 신뢰할 수 있으며 효과적인 사용, d) 안전하고 회복 가능한 사용(Safe, secure, and resilient), e) 합리적인(Understandable) 사용, f) 책임성 있고 추적 가능한 사용(Responsible and traceable), g) 정기적인 모니터링, h) 투명성, i) 결과에 대한 책임(Accountable) 등으로 제시하고 있다. 또한 이 원칙들이 실효성을 가질 수 있도록 하는 감독 기구 설치, Use Case의 개발, 기관 간 협력, 외부(산업 및 학계) 전문가의 참여 등을 제시하고 있다. 정부 내의 인공지능 시스템 사용에 적용되는 규범이지만, 공공 입찰을 통해 참여하는 민간 사업자들에게도 명령을 적용함으로써 일부 우회적인 규제가 이루어진다.

의회에서의 인공지능 관련 입법 시도로는 Ron Wyden과 Cory Booker(상원) 및 Yvette Clarke(하원) 의원이 발의한 알고리즘 책임법안(Algorithmic Accountability Act of 2022)을 들 수 있다. 자동화된 의사결정 시스템에 대한 투명성과 책임성 확보를 목적으로, 기업이 이용자와 규제기관에 '자동화된 의사결정 프로세스'가 어떻게 활용되는지를 인식할 수 있도록 관련 정보를 제공함으로써 투명성을 확보하고, 기업이 그 의사결정의 영향평가를 지속적으로 수행하도록 하는 것이 골자이다. 적용 대상은 자동화된 의사결정시스템(Automated

18) 2023년 7월에도 구글, 메타, 마이크로소프트, 아마존, OpenAI, Anthropic 및 Inflection의 7개 주요 AI 관련 선도 기업들이 안전하고 책임 있는 AI 개발을 위한 노력을 약속하는 "Voluntary Commitments"를 발표한 바 있다.

Decision System; ADS)과 증강된 중요 의사결정 프로세스(Augmented Critical Decision Process; ACDP)로서 사람이 최종적인 의사결정을 하는 경우에도 그 의사결정이 인공지능·알고리즘 활용 결과를 근거로 할 때는 적용 대상이 된다.

"인공지능 권리장전 청사진(Blueprint for an AI Bill of Rights)"(2022년)은 백악관 과학기술정책국(OSTP)이 인공지능 시스템 설계·이용·배포 과정에서의 시민 권리 보호를 위한 지침서로 발간한 것이다. 이 권리장전에서는 시민 권리 및 민주적 가치 보호를 위한 5대 원칙을 밝히고 있다. 안전하고 효과적인 시스템(Safe and Effective Systems), 알고리즘 차별 금지(Algorithmic Discrimination Protections), 개인정보보호(Data Privacy), 고지 및 설명(Notice and Explanation), 인간 대안, 비상 시 대책(Human Alternatives, Consideration and Fallback) 등이 그 내용이다. 아울러 5대 원칙의 실행을 위한 별도 핸드북도 발간하고 있다. 이 핸드북에서는 산업, 정부, 지자체 및 기타 관계자들이 위의 보호 원칙을 정책, 관행 및 기술 프로세스에 관철하기 위해 취해야 할 예시를 제공한다.

생성형 인공지능이 본격적으로 대두하기 시작한 2023년 10월에 연방정부는 다시 "안전하고 신뢰할 수 있는 인공지능의 개발과 사용을 위한 행정명령(Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)"을 발령한다. 앞선 2020년의 행정명령과는 달리 공공부문에서의 인공지능 시스템 사용뿐만 아니라, 국가 안보, 건강, 안전을 위협하는 AI 기술 개발과 이용을 규제하겠다는 의도를 담고 있다. 주요 내용으로는 AI 개발기업의 안전성 평가 의무화, AI 도구의 안전성 표준 수립, 가짜뉴스 및 딥페이크 검증을 위한 콘텐츠 인증 표준 수립 및 개인정보 보호 등이 있다. 즉, 행정명령의 적용 영역에 속하는 인공지능 시스템 개발자는 정부의 전문가팀(이른바 AI Red Team)이 수행하는 안전 검사를 받아야 하는 의무를 지게 되며, 그 결과를 정부에 제출해야 한다. 개인정보와 관련해서는 개별 기업들이 인공지능 개발 훈련 과정에서 개인정보의 불법 사용을 규제하는 지침을 마련하도록 요구하고 있다. 기업들에 대한 의무 부여 외에도 정부 기관들에 대해서도 인공지능의 리스크 제어와 관련한 조치를 권고하고 있다. 인공지능 기술 안전성을 보장하기 위한 최고 수준의 표준을 국립표준기술연구소에 권고하고 있으며, 연방 상무부에 대해서는 인공지능 생성 콘텐츠 식별을 위한 워터마크 의무화를 제도화하도록 권고하고 있다. 이 행정명령은 적용 대상을 미국의 국가 안보와 미국 국민의 건강과 안전을 위협하는 AI 기술 개발과 이용에 국한시키고 있지만, 인공지능 시스템이 야기하는 문제들이 대부분 안전 문제로 귀결된다는 점에서 인공지능에 대해 포괄적으로 규율하고 있는 규범으로 평가할 수 있을 것이다.

한편 미국의 연방 각 주 역시 인공지능의 리스크를 방지하고 제어하기 위한 것이거나, 인

공지능의 기술 개발을 촉진하기 위한 내용을 담고 있는 법률을 제정하거나, 관련 입법들을 시도하고 있다. Utah주는 지난 2024년 5월부터 이른바 "인공지능 수정법(Artificail Intelligence Amendments)"[19]을 시행한다. 리스크 제어의 측면에서는 생성형 AI의 사용 여부에 대해 사전 고지할 의무를 기업에게 부여하고 있다. 산업 진흥 측면에서는 주 상무부에 AI 정책사무소를 설치 및 운영할 것을 요구하고 있다. 또한 인공지능 기술 기반 사업을 실행하기 위한 "Testbed"로서의 프로그램 운영 역시 규정한다. California주는 "첨단 AI 시스템의 보안 및 안전 혁신법"의 입법을 추진하고 있다. 인공지능 개발자에게는 안전성 테스트를 실시하도록 하고 있으며, 이를 감독할 거버넌스를 구축할 의무를 주에 부여하고 있다.

5. 영국의 정책 입안 사례

영국은 어느 영역에서건 경성의 규범을 통한 규율보다는 표준화 또는 거버넌스를 통한 접근을 선호해 왔다. 또한 이러한 접근이 산업 영역별로 이루어지는 경우가 흔하다. 인공지능과 관련해서도 2021년에 발간한 "국가 알고리즘 투명성 표준(Algorithmic Transparency Data Standard: ATRS)"이 일정한 규범적 역할을 수행한다. 이 표준은 데이터윤리혁신센터(CDEI)의 권고에 따른 후속조치로서, 공공 부문의 조직이 알고리즘(algorithmic tool)을 사용하는 경우 투명성 의무를 부여하여 투명성 보고서를 발간하게끔 하는 것이다. 투명성 보고는 3단계로 나뉘어 의무화되어 있다. 먼저 제1단계에서는 표준의 적용 대상인지의 여부에 대해 검증하고 공개 대상 정보를 선정한다(원칙적으로 전부 공개하므로 사실상 예외 선정). 제2단계는 알고리즘 투명성 보고서의 작성의 단계이다. 알고리즘 기능과 의사결정 절차, 알고리즘의 개발과 적용에 대한 정보, 결정 이유에 대한 설명, 기술적 특수성과 데이터, 리스크 및 영향분석 결과를 보고서에 싣는다. 이러한 보고서를 공개하는 것이 제3단계의 과업이다.

영국 과학혁신기술부가 발간한 "AI 규제에 대한 혁신적 접근(A prominent approach to AI Regulation)" 역시 부문별 규제로서 영국 AI 규율의 방향성을 제시하고 있다. 이 문서에서는 AI 규제의 목표로 1) 책임 있는 혁신 및 규제 불확설성의 감소를 통한 AI 성장과 번영, 2) 리스크 규율 및 기본가치 보호를 통해 AI에 대한 대중 신뢰 제공, 3) AI 주도국으로서 영국의 입지 강화를 들고 있다. 또한 규제는 부문별 및 상황별로 이루어지도록 함으로써, 일률적인 리스크 기반 접근의 문제점을 해소하고 있다.

---

19) 이는 단일한 법률명이 아니라 Utah주 법전의 일부 내용을 일컫는 말이다. 동법을 통해 Utah주 법전의 제13편 제2장 제12조 및 제13편 제70장의 내용이 개정 혹은 신설되었다.

Ⅲ. 맺음말

앞서 제시한 인공지능 관련 규범들을 인공지능 시스템의 이용자 보호 정책과 향후의 입법에 참고하기 위해서는 우선 각각의 법제 및 윤리원칙의 목적과 특징에 대한 분석이 필요하다. 주요국 및 국제기구들은 인공지능과 관련하여 자국 혹은 해당 단체의 입장을 반영하고자 노력하였다. 따라서 위에서 언급한 대로 정도의 차이는 있지만 인공지능 관련 규율의 방향성을 어떻게 선택했는지에 따라 규율의 내용이 상이해지는 것이다. 선택한 방향성에 따라 발표 주체별로 주안점을 삼는 내용이 달라지며, 규율의 범위와 시각 역시 차이를 보이게 된다.

인공지능 시스템에 대한 이용자 보호 정책을 연구한다는 차원에서 위 내용들을 어떻게 활용할 수 있을지 고찰이 필요할 것이다. 우선 기존의 국내외 인공지능 규범의 내용을 그대로 수용하기보다는, 인공지능 시스템의 유형이 어떤 것인지(예컨대 생성형 인공지능에 초점을 맞출 것인지), 이용자 보호의 목적성이 있다는 점 및 인공지능 관련 산업 생태계와 문화에 차이가 있다는 점을 고려할 필요가 있다. 따라서 법적·사실적 상황에 부합하는 원칙과 정책으로 해외의 사례를 번안할 필요가 있다. 또한 규범의 내용뿐만 아니라, 규범을 효과적으로 이행하기 위하여 필요한 구조와 실행 방식에 대한 탐구 역시 필요하다. 인공지능에 대한 윤리적 원칙을 포함한 많은 규범들이 실행 구조를 통해 뒷받침되지 않는다면, 공허한 구호에 불과할 것이기 때문이다. 인공지능 시스템 기반 애플리케이션의 제공자뿐만 아니라, 다양한 당사자들이 공동으로 역할을 수행할 수 있는 법제도적 방안도 연구할 필요가 있다. 예컨대 최종 이용자의 역할과 의무에 대해서도 어느 정도 고려해야 할 것이다. 관련 거버넌스에서 정책 집행자가 참여하는 정도와 방식에 대한 논의도 이루어져야 할 것이다.

특히나 인공지능에 대한 규율은 인류가 일찍이 경험해 보지 못한 사안에 대한 것이라는 점을 유념할 필요가 있다. 예컨대 많은 법적 문제의 실마리는 당사자 간 자유로운 의사표시가 존재한다는 전제 하에 풀 수 있으나, 인공지능이 데이터를 기반으로 내리는 결정에는 자연인의 의사표시가 개입하지 않는 경우가 많다. 인공지능의 의사결정 과정을 고려하지 않은 채 구축된 많은 법과 제도들이 인공지능으로 인해 야기된 문제들을 해소하기에는 어려움이 따르는 것이다.[20] 따라서 인공지능 시스템에 대한 리스크 대응의 관점은 경성의 법적 규범을 동원하는 일보다는 일차적으로는 윤리적 차원에서 사업자 스스로 수립하되 공공

20) 같은 취지의 언급을 하고 있는 최근 문헌으로 김형찬, 올바른 AI 규제로 모두를 위한 AI 만들기, 2024, 17쪽.

기관 및 연구자와 협업하는 방식을 적극 모색할 필요가 있다. 즉, 사업자의 자율규제[21]와 규제권자의 정책 수행 사이에서 끊임 없는 피드백과 토론이 이루어질 때, 이용자 보호를 가장 효과적으로 도모할 수 있을 것이다. 인공지능 시스템의 이용 과정을 살펴볼 때, 인공지능이 야기하는 많은 리스크는 이용자의 역량에 기반하는 경우가 많다. 따라서 이용자 보호와 더불어 적극적인 이용자 역량 제고를 위한 노력도 필요하다. 인공지능 시스템에서의 이용자 보호 정책을 인공지능 윤리 교육 사업과 연계할 경우 효과를 거둘 수 있을 것이다. 결국 우리나라의 법적 환경과 시장 상황을 고려할 때 가장 바람직한 방식의 입법 방향성은 공적 주체는 사업자의 자율규제를 보장하고, 이 자율규제가 원활하게 작동할 수 있도록 지원하는 한편, 필요 시 이용자 보호의 최후의 보루로 기능할 수 있어야 할 것이다.

참고문헌

김태오, 통신시장 이용자보호의 쟁점과 과제, 경제규제와 법 제2권 제1호, 서울대학교 공익산업법센터, 2009.

김형찬, 올바른 AI 규제로 모두를 위한 AI 만들기, 2024.

라기원, 유럽연합 인공지능법(EU AI ACT)의 특성과 쟁점 – 우리나라 인공지능 입법에 대한 시사, 규제혁신법제연구 이슈페이퍼 24-20-3, 한국법제연구원, 2024.

박종수, ICT융합에 따른 방송통신 이용자보호 법제의 합리적 개선방안, 법제연구 제44호, 한국법제연구원, 2013.

선지원, 규제 방법의 진화로서 자율규제의 실질화를 위한 연구, 행정법연구 제64호, 2021.

이상욱, 유네스코 인공지능(AI) 윤리 권고 해설서 – 인공지능 윤리 이해하기, 유네스코한국위원회, 2021.

이원태 외, 지능정보화 이용자 기반 보호 환경 조성 총괄보고서, 정보통신정책연구원, 2018

최경진/정경오/이종관, 통신시장 이용자 보호를 위한 법제분석 – 효율적인 피해구제 방안을 중심으로 –, 한국법제연구원, 2015.

21) 자율규제를 적용하기에 적합한 분야를 선정하는 기준은 1) 추상적인 리스크는 존재하지만 구체적인 리스크는 명확하지 않을 때, 2) 민간 주도의 자율적인 기술 발전이 중요할 때, 3) 시장 행위자의 기술과 산업에 대한 지식 및 전문성이 규제 입안자나 정책 수행자보다 뛰어날 때, 4) 자율성 보장 자체가 하나의 가치로 지켜져야 할 때, 5) 기업의 활동에 시장의 이용자들이 민감하고 적극적으로 반응할 수 있을 때 등으로 요약할 수 있다. 이런 의미에서 인공지능 기반의 서비스 분야는 자율규제가 적합한 영역이라고 평가할 수 있다. 선지원, 규제 방법의 진화로서 자율규제의 실질화를 위한 연구, 행정법연구 제64호, 2021, 110쪽 이하.

Oppermann,Thomas/Classen, Claus Dieter/Nettesheim, Martin, Europarecht 7. Auflage, Mü nchen 2016.

Political Guidelines | Ursula von der Leyen – Candidate for the European Commission President, 18. 7. 2024.

Susnjak et. al., Over the Edge of Chaos? Excess Complexity as a Roadblock to Artificial General Intelligence, arXiv:2407.03652 [cs.AI].

AI Act | Shaping Europe's digital future 홈페이지 (https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai)(2024. 8. 31. 최종 접속).

Ethics guidelines for trustworthy AI | Shaping Europe's digital future 〈https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai〉(2024. 8. 31. 최종 접속).

OECD Legal Instruments 〈https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#backgroundInformation〉(2024. 8. 31. 최종 접속).