일반논문 「행정논총」 제63권 제2호(2025. 6): 109~147 http://dx.doi.org/10.24145/KJPA.63.2.4

AI 규범의 국가간 확산 흐름에 관한 연구: 주요 국제기구 및 EU, 미국, 한국을 중심으로*

박준희**

〈目 次〉

〈요 약〉

본 연구는 AI에 관한 국내외 규범들의 국가간 확산에 있어서의 흐름을 살펴보기 위해 그 규범들에 포함된 AI 원칙들을 기준으로 분석하였다. 주요 국제기구와 주요 행위자로 상정한 EU, 미국, 한국에서 2016년부터 2024년까지 발표한 AI 관련 24개 규범들을 분석대상으로 하였으며, OECD, UN 등의 국제기구에서 공통적으로 제시한 6가지 윤리원칙들을 분석의 기준으로 삼았다. AI 규범의 확산 과정을 분석한 결과, 시간의 흐름에 따라 각 규범들이 강조한 AI 원칙들과 함께 각 주체들의 흐름 변화와 선도적 국가로서 가지는 분야를 확인할 수 있었다. 행위자별 AI 규범들의 분석결과를 통해 EU는 AI의 안전성과 책임성을 담보하는 데 있어서 주도적 역할을 해온 반면, 미국은 AI 기술의 개발, 활용 및 시장의 촉진 측면에서 국제사회의 리더적 지위를 유지해왔고, 한국은 이들의 정책을 따르던 후발국가에서 현재는 국제사회에서 주요한 국제회의를 주도하는 흐름으로 나아가는 과정에 있음을 밝혔다. 이러한 연구결과는 AI 거버넌스에 관한 후속 연구에 있어서 주체와 원칙 및 내용의 체계화에 기여할 수 있다.

【주제어: AI 규범, AI 원칙, 확산 흐름】

* 이 논문은 연세대학교 바른ICT연구소의 지원을 받아 수행된 연구임.
** 연세대학교 바른ICT연구소 연구교수(j_hpark@yonsei.ac.kr, elin31@snu.ac.kr)
논문접수일(2025.4.28), 수정일(2025.6.10), 게재확정일(2025.6.16)

Ⅰ. 서론

국제사회는 OECD, UN 등 국제기구와 EU 및 미국을 필두로 인공지능에 대해 규범적으로 규율하려는 움직임으로 재편되고 있다. 최근 UN총회가 AI에 관한 결의안을 채택하였고, OECD에서는 2019년의 AI 원칙을 개정하였으며, EU가 2021년 최초로 제안한 AI 법안은 3년 만에 법률로 확정되었다. 특히 EU의 AI 법은 AI 기술이 시민의 안전과 권리를 보호하는 방식으로 규제하는 최초의 포괄적 법률이라는 의미를 가진다. OECD, UN, UNESCO 등 주요 국제기구들의 AI 규범에서는 인간의 기본적 권리와 존엄성 존중, 개인정보 보호, AI 시스템의 안전성과 투명성, 공정성 등을 공통적으로 제시하고 있다. 이는 AI 기술의 실현과 활용이 인간의 존엄을 지키고 보다 바람직한 사회를 구현하기 위한 방향으로 나아가고자 하는 정책적 지향을 의미한다. AI에 관한 원칙과 규범들은 EU와 같은 초국가적 기구 뿐만 아니라 개별 국가들 차원에서도 권고, 지침, 법률 및 행정명령 등의 형식으로 제시되고 있으며, 그들의 참여국이거나 리더적 지위에 있는 국가들을 따르는 국가들의 후속조치도 지속적으로 나타나고 있다. 한국 역시 현재 AI Safety Summit, GPAI, G20에 참여하면서 그 규범들의 영향 하에 놓여있다.

이러한 국제 정세 속에서 본 연구는 국제사회에서 주요 행위자들에 의해 제정 및 발표된 AI에 관한 규범들이 언제, 누구에 의해 제기되었고, 그 확산의 흐름이 어떠하였는지에 대하여 살펴보기로 한다. 이를 위하여 주요 국가들과 국제기구들이 제시한 24개의 AI에 관한 법령 및 규범들을 대상으로 하는 분석을 시도하였으며, 국가간 정책확산에 관한 이론적 논의와 AI 원칙들을 분석틀과 기준으로 활용하였다. 이를 통해 AI 규범 논의에 있어서 국제사회에서의 주도적 행위자의 흐름을 파악하고, 향후 한국의 선도자로서의 가능성과 방향에 대하여도 살펴보고자 한다.

현재 AI에 관한 규율은 AI 라는 기술 자체 만큼이나 전세계적 이슈이며 AI 거버넌스 중 하나에 해당한다. 본 연구에서는 AI 규범들의 확산 현상을 정책확산에 관한 이론적 모형을 통해 분석하고 AI 윤리원칙을 기준으로 한 질적 분석을 정량화하여 나타내고 있으며, 그러한 확산을 시간적 흐름 속에서 살펴보고 시각화하고 있다는 점에서 기존 연구들과 크게 차별화된다. 나아가 이 AI 규범에 관한 연구는 AI에 관한 국가전략 형성에 관한 실무적 측면과 확산 및 AI 거버넌스 연구와 관련한 학문적 차원 모두에서 중요한 의미가 있다.

Ⅱ. 선행연구 검토 및 이론적 배경

1. AI에 관한 입법정책적 동향에 관한 선행연구

AI가 혁신 기술로서 사회에 커다란 영향을 미치기 시작하면서 그 입법정책적 동향을 보고하는 연구들이 폭발적으로 나타났다. 국내 연구들에서는 주요 국가들의 AI 입법과 정책들을 검토하거나(성욱준 외, 2023), EU의 AI 법의 주요 쟁점들을 밝히거나(양천수, 2024; 김법연, 2024; 라기원, 2024a) AI에 관한 국제기구의 윤리적 지침 또는 권고들을 분석하는 연구들(유재흥 외, 2021; 박기갑, 2022; 양천수, 2024)이 주로 제시되었고, 최근에는 주요 국가들의 거버넌스 형성에 관한 동향을 파악하는 단계로 나아가고 있다. 라기원(2024b)이 EU의 인공지능법을 가장 대표적인 거버넌스 규범화의 사례로 보면서 인공지능 거버넌스를 "AI 시스템의 잠재적 위험으로부터 사회를 보호하는 것을 목적으로 인공지능 기술의 개발 및 사용 단계에서의 규칙을 제공하고 위험성을 평가하며 감독하는 체계"라고 정의한 데 반해, 양천수(2024)의 경우에는 당해 법이 윤리적 규율에서 경성규범(regulation)을 통한 규율로 이행하였다는 의미를 가진다고 평가하였다. 그 밖에 미국의 AI 관련 행정명령이나 영국의 AI 국가전략, EU의 디지털 서비스법(DSA) 및 디지털 시장법(DMA) 등 개별 법령들에 대해 분석하거나 쟁점을 논의하고 있는 연구들(NIA, 2021; 이병준, 2021; 국가안보전략연구원, 2023)도 꾸준히 제시되어 왔다.

해외에서도 EU의 GDPR의 한계로서 자동화된 의사결정과 위험으로부터 보호하지 못하는 문제점을 지적하거나(Wachter, Mittelstadt, & Floridi, 2017; Van & Vrabec, 2019; Vanberg, 2020; Chawki, 2024) EU의 AI 법안에 대해 비교분석하는 연구 및 당해 법률에서 설명하는 사항들을 비판적으로 분석하는 연구들(Pavlidis, 2024; Ziller, 2024)도 나타났다. 세부적으로는 미국, 유럽, 호주 등의 개인정보보호 법률 및 시스템을 비교하거나(Souza, Carlos Affonso, et al., 2021; Egan, 2022; Phillips, 2024) AI의 설명가능성 또는 투명성을 비판적으로 검토하는 연구(Karwatzki, Sabrina, et al., 2017; Chowdhury, Soumyadeb, et al., 2022; Long, Yonghao, et al., 2023) 뿐만 아니라 공정성 및 책임성, 투명성, 개인정보보호 등 AI에 관한 일반 원칙들을 검토하는 연구들(Mittelstadt, 2019; Tsamados, Andreas, et al., 2021; Siddiqui, 2024)이 나타나고 있다.

2. AI 거버넌스에 관한 선행연구

AI의 대두와 함께 AI 관련 개별법이나 정책에 관한 연구들 못지않게 각 국가들의 AI 거버

넌스를 비교하는 연구들도 등장하였다. Radu(2021)는 2017년에서 2019년 사이의 12개 국가들의 AI 국가전략들을 비교분석하면서 이들을 공공과 민간의 경계가 뒤섞인(hybrid) 거버넌스 구조로 이해하고 그 특징들을 도출하였다. 그에 따르면, 이들 전략들은 대부분 윤리중심적(ethics-oriented)이며, 역할과 책임이 명확하지 않은 기능적 불확정성(functional indetermination)을 특징으로 한다. 윤리 원칙을 중심으로 AI 규제정책들을 비교분석한 연구들에는 Walter(2024)와 Shukla(2024)의 연구가 있다. Walter(2024)의 연구는 미국과 EU, 아시아 및 아프리카 등 주요 지역의 AI 전략 및 실행방식을 평면적으로 비교분석하면서 규제의 거버넌스 방식과 AI 평가항목을 제시하고 있으나, 이론적 분석틀은 제시되지 않았다. Shukla(2024) 역시 'AI 시스템이 어떠한 윤리원칙에 의해 개발 및 확산되어야 하는가?'라는 질문에 답하도록 하여 각 국가들이 제정한 AI 규제 정책들을 투명성, 프라이버시, 공정성, 책임성, 사회적 영향 등의 윤리원칙들을 기준으로 분석하였으며, 여기서 윤리원칙은 AI 시스템 개발 및 배포의 메커니즘으로 보았다.

대부분의 AI 거버넌스 연구들은 각국의 정책 및 문헌들에 대해 질적 내용분석 방법을 활용한 기능적 비교방식(functional comparison)을 사용하고 있다. Kashefi et al.(2024)의 경우 EU, 미국과 중국, 일본, 한국 등 아시아 국가들의 AI 규제들을 대상으로 ⅰ)강제력, ⅱ)주체의 성격, ⅲ)윤리원칙의 반영, ⅳ)규제와 혁신간 균형, ⅴ)국제적 규범과의 조화 등을 분석틀로 삼아 비교하였다. 특히 동일한 성격의 법규범이 아닌 다양한 수준과 유형의 AI 규범 문서(즉, 법제, 가이드라인, 전략, 계획 등)를 내용, 목표, 적용방식을 기준으로 비교하였다. Schiff et al.(2020)의 경우에도 공공, 민간, NGO의 25개의 윤리 및 정책들을 대상으로 법령을 포함하여 가이드라인, 윤리원칙, 전략 등에 대해 기능적 비교방식으로 분석하였다. 이 연구에서는 AI 윤리의 제정 동기를 국제기구의 사회적 책임(social responsibility), 개별 국가들의 경쟁이익(Competition Advantage), 전략 계획 및 개입, 사회적 책임, 선도국으로서의 입지 확보라는 리더십 신호(leadership signaling) 등으로 설명하였다. Xu et al.(2024)의 경우에는 아시아 국가들이 EU, 미국 및 중국의 영향을 받을 것으로 예상하면서, AI 규제 거버넌스로서 12개 국가전략 문서를 선정하여 키워드를 활용하여 관련 구절을 탐색하는 질적 내용분석 방법으로 분석하였다.

유사한 방식의 이론중심적 개념적 연구로서 Wirtz et al.(2019)은 179개 문헌과 사례를 통해 거버넌스, 기술, 윤리, 운영에 관한 11개 AI 프레임워크를 제안하였으며, Moon(2022) 역시 포용적(inclusive) AI를 실현하기 위한 참여적 거버넌스 개념에 가치기반 윤리지침 및 법제적 규제, 범정부적 거버넌스 체제, AI 영향평가, 독립적 개인정보보호 기구, UN SDGs와 연계한 AI 기반 글로벌 협력 구상들을 구성하여 제안하였다. John et al.(2024)은 미국 바이

든 행정부의 AI 행정명령(2023), EU의 AI법, 중국의 AI Regulation in China, 싱가포르의 Model AI Governance Framework 등 상대적으로 제한된 범위의 문헌들에 대해 인권 요소를 중심으로 비교분석하였다.

이러한 AI 규제 또는 거버넌스에 대한 비교연구들과 달리, 정량적 토픽모델링과 정성적 내용분석을 사용하여 동남아시아 국가들과 인공지능 글로벌 파트너쉽(GPAI) 회원국인 호주와 싱가폴의 AI 거버넌스 정책을 비교하는 방법을 사용한 연구(Keith, 2024)도 나타나고 있지만, 대부분의 연구들은 윤리원칙이나 정책 및 기능들을 평면적으로 비교하는 데 그치는 한계를 가진다.

이에 비해 본 연구는 국제기구에서 공통적으로 제시되고 있는 AI 원칙을 기준으로 국가 및 국제적 AI 규범들에 나타난 윤리원칙들의 시계열적 변화가 어떻게 나타나는지를 주체별로 분석함으로써 AI 규범의 확산 현상을 분석하고 있다. 즉, 시간의 흐름에 따라 주요 국가들에서 강조되고 있는 AI 원칙들과, 주체별로 어떠한 원칙들을 강조하면서 규범의 확산이 나타나고 있는지를 살펴볼 수 있다는 점에서 기존 연구들과 차별점을 가진다고 할 수 있다.

3. 정책확산에 관한 이론적 논의

1) 정부간 또는 국가간 정책확산에 관한 선행연구

확산에 관한 선구자로서 Walker(1969)는 더 많은 정부들이 채택하는 정책일수록 정당성을 얻으며, 이를 획득한 혁신은 그 자체의 추진력을 가진다고 하였다. 특히, 혁신의 확산은 선도적 정부의 정책 채택이나 이웃 정부의 영향을 받기도 하지만, 강력한 권위와 영향력을 가진 국제기구나 상위정부로부터의 수직적 압력을 받아 이루어지기도 하고, 때로는 그 영향을 받은 정부가 다시 확산을 주도하기도 한다. 이러한 흐름은 국제사회에서 다른 사회적, 경제적 수준을 가진 국가들 사이에서도 나타나며 정책확산에 관한 이론적 틀로 분석될 수 있다.

확산의 유형은 수준과 권한이 동일한 정부간 수평확산(horizontal diffusion)과 상하위 정부간 수직확산(vertical diffusion)으로 구분된다. 수평확산에는 인접지역(neighboring state)간 확산과 인접하지 않은 동위의 정부들간 확산이 포함되며, 수직확산에는 국제기구에서 국가 차원으로의 도입(Brune et al., 2004; Schimmelfennig & Sedelmeier, 2004), 연방정부에서 주정부로의 확산(Allen et al, 2004; Shipan & Volden, 2006; Karch, 2006), 중앙 또는 주정부에서 지방정부로의 확산 또는 상위정부로부터 하위정부로의 확산(Karch, 2008; Welch & Thompson, 1980; 남궁근, 1994; 이승종, 2004; 최상한, 2010; 정명은,

2012; 이석환, 2013)이 포함된다.

정책확산에 관한 분석틀로서 Berry and Berry는 내부적, 외부적 접근을 하나의 모형으로 통합하여 제시한 후, 후속연구에서 외부적 모형을 국가간 상호작용 모형(national interaction model), 지역적 모형(regional model), 선도자-후발자 모형(leader-laggard model), 동형화 모형(isomorphism model), 수직적 영향 모형(vertical influence model)으로 구분하여 정부간 확산이 어떻게 이루어지는지에 대해 설명한 바 있다(Berry and Berry, 2007). 특히, 선도자-후발자 모형(leader-laggard model)은 선도자 역할을 하는 특정 국가의 정책을 다른 국가들이 모방하는 형태로서(Walker, 1969; Berry & Berry, 2007), 지리적 위치에 관계없이 선도적 정부 또는 리더 국가가 채택하면 다른 정부들의 채택률이 증가한다고 설명한다. 수직적 영향력 모형의 경우 단일 국가의 선도적 역할을 제시한다는 점에서 선도자-후발자 모형과 유사한 면이 있지만, 후발자의 재량(discretion)을 제거하는 대신 이들에게 인센티브를 제공하거나 특정한 활동을 강제하는 방식을 통해 확산을 가속화시킨다는 점에서 차이가 있으며, 확산 국가들의 지리적 군집성을 설명하지 못한다는 한계를 가진다. 지역적 확산 모델(regional diffusion model)의 경우에는 유사한 환경과 사회적, 경제적 문제를 가지는 인접 국가들의 수평적 확산을 잘 설명할 수 있지만, 인접하지 않은 국가간 확산을 설명하는 데는 한계가 있다. 지리적 인접성 여부와 상관없이 국가간 확산에 선도적 행위자의 역할을 설명할 수 있는 이론적 틀로는 선도자-후발자 모형이 적절하다고 볼 수 있다.

최근 정책 확산과 혁신(Policy Diffusion and Innovation)에 관한 연구를 다룬 de Oliveira et al.(2023)에 따르면 확산의 대상에는 법, 규제 및 프로그램 등 정책 자체뿐만 아니라 그 기반의 사고방식이나 가치가 포함되며, 그것이 국가간, 지방정부간 및 국제기구에서 국내로 확산될 수 있다고 하면서, 확산의 경로로서 다른 나라의 정책 성공 및 실패 사례를 통한 학습(learning), 인정받는 정책을 단순히 따라하는 모방(imitation), 경제적, 정치적 이익을 위한 경쟁적 도입(competition), 국제기구나 상위 정부의 압력(coercion), 권력적 역학 또는 제도적 저항 등이 작용할 수 있다고 설명하였다.

정책확산에 관한 선행연구들(2004-2019)을 체계적으로 검토하여 이론적 흐름과 연구 경향 및 방법론 등을 정리한 Danaeefard et al.(2022)의 연구에서도 Berry and Berry(2007)와 선행연구들로부터 National Interaction model, Regional Diffusion model, Leader-Laggard model, Convergence model, Vertical Penetration model 등의 다섯가지 확산 모델과 학습, 모방, 경쟁, 강제의 4가지 메커니즘(Shipan & Volden, 2008)을 제시하였다. 이들은 대다수의 연구가 지방 정부간 또는 이웃 국가간 확산에 초점을 맞추고 있어 국가간 확산에 관한 연구와 통계적 연구가 부족함을 지적하고, 향후 국제정치적 맥락이 반영된 연구와

정량적, 비교 방법론적 확대, 행위자 중심의 분석과 이론통합적 접근이 필요하다는 제안을 하고 있다.

중국의 정책확산에 관한 연구로서 하향확산과 상향확산을 분석한 Zhang & Zhu(2019)의 연구와 중국 노인수당(OAAP) 정책의 확산에서 성급 정부의 중개자적 역할을 강조한 Lou et al.(2023)의 연구의 경우 한국과 유사한 지방구조를 가진 중국 내 정부간 확산을 분석하였다는 점에서 의미를 가진다. 중국의 공공자원 거래 플랫폼 혁신에 대한 법령 규정을 대상으로 상향 확산과 하향 확산을 분석한 Wang et al.(2020)의 연구의 경우에는 확산 모델의 진화를 주장하면서 단계마다 다른 모델을 적용한다는 점에서 차별성을 가지나, 계량적 검증 절차상의 한계가 나타난다.

국가간 정책확산에 있어서는 UNASUR과 같은 지역적 제도(regional institutions)의 역할을 강조한 Agostinis(2019)의 연구와 국제기구나 초국적 네트워크와 같은 외부적 압력이 인권규범의 확산에 미치는 영향을 검토한 Ring(2014)의 연구가 있다. Ring은 강압은 강대국들이 다른 행위자들의 변화를 촉진하기 위해 자신의 힘을 사용하는 하향식 과정이고, 경쟁은 약소국들이 국제적 인정을 받기 위해 경쟁하는 하향식 과정이며, 모방은 약소국들이 강대국을 모델로 삼는 하향식 과정이라고 보면서, 학습 역시 아직 경험하지 못한 정보를 적극적으로 찾는 하향식 메커니즘으로 가정하고 있다.

2) 사회구성주의와 규범적 정당성 접근

국제관계이론(IR theory)으로서 사회구성주의는 국가간 상호작용을 설명하는 이론으로서, 정책확산, 규범의 전파, 국가정체성 형성 등을 설명한다. Wendt(1999)에 따르면 국가간 관계는 사회적으로 구성된 행위자 간 상호작용의 산물로서, 특정 정책의 도입 역시 '당연히 그래야 마땅하다'는 규범적 압력에 의해 움직인다. Finnemore(1996)도 정당성 권위자(legitimacy authorities)인 국제기구가 정상국가(normal state)의 정책 모델을 규범으로 제시하고 국가들은 이를 국제적 정당성 확보의 수단으로 받아들이며 정책을 자발적으로 수용한다고 설명한 바 있다. 즉, 규범의 확산은 힘에 의해서가 아니라 '어떻게 행동하는 것이 옳은가'라는 규범적 정당성 인식이 작용한다. 이러한 설명은 국제기구의 보편적 규범 창출이 개별 국가들에 영향을 미치는 과정을 설명할 수 있다. 국가간 확산에 규범적 정당성 접근을 시도한 연구로서, 이론적 측면에서 외부적 확산에 관한 분석틀을 제시한 Weyland(2005)는[1] 외부적 압력 접근(external pressure framework)이 국제기구들의 수직적 수단(vertical

1) 그의 이론적 분석틀은 크게 외부적 압력(external pressure), 상징적 또는 규범적 모방(normative imitation), 합리적 학습(rational learning), 인지적 휴리스틱스(cognitive heuristics)로 구분된다.

imposition)으로 인해 빠른 속도의 확산과 함께 서로 다른 국가들간에 나타나는 공통점(commonality amid diversity)을 설명할 수 있지만, 인접 국가간의 군집효과(neighborhood effect)를 쉽게 설명하지 못하는 한계가 있는 데 반해, 규범적 모방 접근(normative imitation approach)은 다른 국가들간 공통성이 나타나는 이유를 국제적 정당성 획득과 새로운 국제 규범에 대한 순응으로 설명할 수 있다고 하였다. 이러한 주장은 혁신(정책)을 현대성(modernity)의 상징이자 매력적이고 규범적으로 적절한 모델로 전환시킴으로써 채택의 정당성을 강화하는 일종의 구성주의적 접근이기도 하다. 특히 이 접근은 국제사회에서 느슨한 연계를 가지는 회원국간 관계와 다양한 국가들간에 같은 정책이 채택되거나 빠른 속도의 혁신이 나타나는 현상을 잘 설명할 수 있다. 다른 국가들이 많이 채택한 정책일수록 매력적이며, 합리적 학습이 요구하는 평가적 정보를 얻기 전에 빠르게 국제사회의 트랜드를 따라가려는 동기가 후발자로 하여금 밴드웨곤(bandwagon)에 올라타도록 촉구한다.

3) 확산 메커니즘(Diffusion Mechanisms)

확산의 메커니즘은 구성주의와 규범적 정당성 접근에서 설명하는 모방(imitation)과 강압(Coercion) 기제 외에 학습(learning)과 경쟁(competition) 기제를 포함한다(Simmons et al., 2006). 전자가 권위적 국제기구에 의한 확산을 설명할 수 있다면, 동위의 정부간 확산에 대해서는 학습과 경쟁 기제에 대한 설명을 요한다. 확산 기제에 대한 선행연구들로부터 이들 각각의 개념을 정의한다면, 학습(learning)이란 정책문제를 해결하기 위해 성공적으로 판명된 다른 정부의 정책을 도입함으로써 정책확산이 일어나도록 하는 메커니즘(Shipan & Volden, 2008; 이석환, 2022)으로서 정책결정자가 다른 정부의 경험 및 정보를 이용하여 합리적으로 결정하는 행위라 할 수 있다. 모방(emulation)은 '상징적 차원에서 정책의 정당성 확보나 사회문화적 유사성에 동조하기 위하여 단순히 다른 정부의 정책을 따르는 메커니즘'이라고 정의할 수 있다(DiMaggio & Powell, 1983; Dobbin et al, 2007; 장석준, 2013: 257). 경쟁(competition)에 대해서는 국가 또는 정부간 관계에서 경제적, 정치적 이익을 얻거나 잃지 않기 위하여 정책을 채택하는 과정에서 확산이 나타나는 메커니즘으로 설명한다(김대진, 2010; 조일형 외, 2014: 10). 강압(Coercion)은 권위나 권력을 가진 상위 정부의 규제나 재정적 압력을 통해 정책이 확산되는 수직적 메커니즘이라 할 수 있으며, 주로 국제기구나 강력한 국가에 의한 정책 도입이 나타난다(Braun & Gilardi, 2006).

확산에 관한 다수의 연구들은 확산 모형과 이 네 가지 기제를 통해 분석하기도 한다. 대표적으로 Shipan & Volden(2008)은 미국 내 675개 도시의 금연 정책의 유형을 조사하여 확산에 영향을 미치는 네 메커니즘에 대해 도시 간, 하향 확산에 초점을 두고 분석하였으며,

Suárez(2012)는 미국과 영국간 정책확산을 확산 기제를 통한 질적 분석방법으로 분석하였다. Gilardi & Wasserfallen(2019) 역시 정책확산의 배경에 정치가 있음을 설명하면서 네 가지 유형의 확산기제를 제시한 바 있다.

4. 선행연구의 한계와 본 연구의 차별점

국제사회에서 AI에 관한 규범들이 확산되는 데에는 그러한 규범의 형성에 주도적 역할을 하는 선도자(leader)와 이들을 따라 규범을 형성하는 후발국가(follower)들의 조치가 있어야 한다. 본 연구는 국제사회에서 AI에 관한 규범들에 선도적인 국가 및 국제기구들을 중심으로 정책확산에 관한 외부적 모형을 활용하여 AI 규범들에 관한 질적 분석을 시간적 흐름 속에서 시도하였다.

기존의 연구들은 각국의 AI에 관한 입법정책적 동향을 개괄적으로 살펴보거나 개별법이나 소수 법제들을 단순비교하거나 법적 쟁점 및 원칙에 한정하여 논의하는 경향을 보여왔으나, AI에 관한 정책 또는 규범들을 계량적으로 분석하거나 국가간 상호작용 및 정책적 흐름을 분석한 연구는 보이지 않는다. AI 규제 또는 거버넌스에 대한 비교 연구들도 주체별 윤리원칙이나 기능에 대한 평면적 비교에 그치고 있으며, 시계열적 확산의 흐름을 제시한 연구는 아직까지 나오고 있지 않다. 본 연구는 국제기구에서 공통적으로 제시되고 있는 AI 원칙을 기준으로 주요 국가들 및 국제적 AI 규범들에 나타난 윤리원칙들의 시계열적 변화가 어떻게 나타나는지를 주체별로 분석함으로써 AI 규범의 확산 양상의 특성을 분석하고자 하였다. 즉, 시간의 흐름에 따라 주요 국가들에서 강조되고 있는 AI 원칙들과, 주체별로 어떠한 원칙들을 강조하면서 규범의 확산이 나타났는지의 흐름을 살펴볼 수 있다는 점에서 기존 연구들과 차별점을 가진다고 할 수 있다.

Ⅲ. 분석적 틀

1. 분석대상 및 시기적 범위

질적 분석방법을 통한 확산 연구에서 적절한 대상 선택은 중요한 의미를 가지며, 여기에는 선택 편향(selection bias)이 불가피하게 나타나는 특징을 보인다(Meseguer & Gilardi, 2009; Marsha & Sharman, 2009). 본 연구의 분석대상은 2016년부터 2024년 현재(12월)까지 OECD, UN, UNESCO 등 주요 국제기구에서 발표한 규범들과 EU, 미국, 한국의 등 개별

국가들의 AI에 관한 규범들로서 법률 및 선언, 지침 등을 포함한다.[2] 주요 국제기구들은 AI에 관한 국제적 선언이나 지침을 발표함으로써 개별 국가들의 규범 형성에 강력한 영향을 미칠 수 있는 위치에 있고, EU의 경우 AI에 관한 최초의 규제법을 제정하는 등 국제사회에서 AI에 관한 규제 논의를 주도하였으며, 미국은 AI를 포함한 주요 기술의 개발 및 산업 분야에 있어서 글로벌 리더 국가로서의 역할을 해왔기 때문이다. 한국의 경우[3] 이들에 비해 후발자로서의 위치에서 출발하였지만 최근 '서울 AI Summit'(2024)을 개최하며 서울선언을 발표하는 등 국제사회에서 AI에 관한 규범적 논의를 시작하였다. 이에 따라 본 연구에서는 EU, 미국, 한국과 주요 국제기구들이 선언한 총 24개 규범들에 대하여 AI에 관한 공통된 6가지 윤리원칙을 기준으로 분석하였다.

〈표 1〉 분석대상 총 24개 AI 규범

주체	시기	규범	코드
EU(7)	2016	EU 유럽 일반 개인정보보호법 [GDPR: General Data Protection Regulation]	EU1
	2018	EU AI 전략 및 유럽 인공지능 전략 [Communication: Artificial. Intelligence for Europe]	EU2
	2019	신뢰할 수 있는 AI를 위한 윤리지침 [EU Ethics Guidelines for Trustworthy AI]	EU3
	2021	인공지능법 초안 [EU Artificial Intelligence Act]	EU4
	2022	디지털 시장법(DMA) [Digital Markets Act]	EU5
	2022	디지털 서비스법(DSA) [Digital Services Act)]	EU6
	2024	인공지능법 최종안 [Artificial Intelligence Act]	EU7
OECD	2019	OECD AI 권고안 [OECD Recommendation of the Council on Artificial Intelligence]	OE1
UNESCO	2021	유네스코 인공지능 윤리 권고 [UNESCO Recommendation on the Ethics of Artificial Intelligence]	UNE
UK(2)	2021	(영국) 국가 AI 전략 [National AI Strategy]	UK1
	2023	제1회 AI 안전성 정상 회의 (블레츨리 선언) [AI Safety Summit (Bletchley Declaration)]	UK2
UN	2024	UN 총회 AI 결의안 [United Nations General Assembly AI Resolution]	UN
OECD	2024	OECD AI 권고안 개정안 [OECD Recommendation of the Council on Artificial Intelligence]	OE2
EU-US-UK	2024	AI와 인권, 민주주의 및 법치에 관한 기본 협약[Council of Europe Framework Convention on artificial intelligence and human rights, democracy, and the rule of law]	EUK

2) 다만, 분석대상의 시기적 범위에 한국의 AI 기본법은 해당하지 않아 분석대상에서 제외되었다.

3) 한국의 인공지능 관련 법안은 계류의안 13개(22대 국회)와 처리의안 15개(20-21대 국회, 임기만료폐기)로 총 28개이며, 아직까지 통과된 법안은 없다(https://likms.assembly.go.kr/bill/FinishBill.do, 2024.11.8. 기준). 이에 따라 본 연구에서는 정부의 AI 관련 지침과 윤리 기준 등 규범들을 분석대상으로 삼았다.

미국(6)	2019	미국의 인공지능 선도적 지위 유지 유지를 위한 '인공지능 리더십 유지 행정명령(제13859호) [Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence"]	US1
	2020	연방정부의 신뢰할 수 있는 인공지능 사용 촉진 행정명령(제13960호) [Executive Order, "Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government"]	US2
	2021	국가 인공지능 이니셔티브법 [National Artificial Intelligence Initiative Act of 2020] (H.R 6216)	US3
	2022	인공지능훈련법[Artificial Intelligence Training for the Acquisition Workforce Act]	US4
	2022	인공지능진흥법 [Advancing American AI Act]	US5
	2023	안전하고 보안이 보장되며 신뢰할 수 있는 인공지능의 개발과 사용에 관한 행정명령(제14110호) [Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence]	US6
한국(4)	2019	인공지능 국가전략	KR1
	2020	인공지능 윤리기준	KR2
	2024	서울 정상회의 서울선언 [Seoul Declaration for Safe, Innovative and Inclusive AI]	KR3
	2024	인공지능의 책임있는 군사적 이용에 관한 고위급회의 (REAIM) 선언	KR4

2. 분석틀 및 분석기준

주요 국제기구들에서 제시하고 있는 AI 원칙(Principles)들은 'AI에 관한 규제가 어떠한 원칙과 가치에 근거하여 이루어져야 하는가'에 대하여 책임성, 안전성, 신뢰성, 투명성과 같은 윤리적 기준 및 인간중심의 가치를 핵심적인 요소로 다루고 있다. OECD의 AI 권고안(2019, 2024), UNESCO의 AI 윤리에 관한 권고안(2021), UN의 AI에 관한 결의안(2024), EU와 미국 및 영국간 AI에 관한 협약(2024)에서 제시한 AI에 관한 가치기반 원칙(values-based principles)들을 정리하면 다음과 같다.

〈표 2〉 분석기준 도출을 위해 참조한 국제기구의 AI 원칙들

주체 및 규범	AI에 관한 주요 원칙
OECD, AI 권고안[4] (2019, 2024)	• 포용적 성장과 지속가능한 개발 및 웰빙(inclusive growth, sustainable development and well-being) • 인간 중심 가치와 공정성(human-centred values and fairness)(2019) ▶ 공정성 및 프라이버시를 포함하여 법치, 인권, 민주주의 가치에 대한 존중 (respect for the rule of law, human rights and democratic values, including fairness and privacy)(2024) • 투명성과 설명가능성(transparency and explainability) • 견고함, 보안 및 안전(robustness, security and safety) • 책무성(accountability)

UNESCO, AI 윤리에 관한 권고안 (2021)	가치: • 인권과 기본권적 자유 및 인간 존엄성에 대한 존중, 보호 및 촉진(Respect, protection and promotion of human rights and fundamental freedoms and human dignity) • 환경 및 생태계 제고(Environment and ecosystem flourishing) • 다양성과 포용성 확보(Ensuring diversity and inclusiveness) • 평화롭고 공정하며 상호연결된 사회에서의 삶(Living in peaceful, just and interconnected societies) 원칙: • 비례성과 무해성(Proportionality and Do No Harm) • 안전성과 보안(Safety and security) • 공정성과 비차별(Fairness and non-discrimination) • 지속가능성(Sustainability) • 프라이버시권, 데이터보호(Right to Privacy, and Data Protection) • 인간의 감독과 결정(Human oversight and determination) • 투명성과 설명가능성(Transparency and explainability) • 책임과 책무성(Responsibility and accountability) • 인식 및 문해력(Awareness and literacy) • 다중이해관계자와 적응적 거버넌스 및 협력(Multi-stakeholder and adaptive governance and collaboration)
UN, AI 결의안 (2024)	• 인간중심적(human-centric) • 신뢰할수 있는(reliable) • 설명가능한(explainable) • 윤리적(ethical) • 포용적(inclusive) • 인권 및 국제법(promotion and protection of human rights and international law) • 프라이버시 보호(privacy preserving) • 지속가능한 개발 지향(sustainable development oriented) • 책임성있는(responsible) • 디지털 전환, 평화 촉진, 디지털 디바이드 해소(promote digital transformation, promote peace, overcome digital divides between and within countries) • 인간중심으로서 인권 및 기본적 자유의 향유(promote and protect the enjoyment of human rights and fundamental freedoms for all, keeping the human person at the centre)
EU-미-영, AI에 관한 협약 (2024)[5]	일반의무: • 인권 보호(Protection of human rights) • 민주적 절차 및 법의 지배 존중(Integrity of democratic processes and respect for the rule of law) 원칙들: • 인간 존엄성과 개인의 자율성(Human dignity and individual autonomy) • 투명성과 감시(Transparency and oversight) • 책무성과 책임(Accountability and responsibility) • 형평성과 비차별(Equality and non-discrimination) • 프라이버시와 개인정보보호(privacy and personal data protection) • 신뢰성(reliability) • 안전한 혁신(safe innovation)

본 연구에서는 AI에 관한 규범들을 분석하기 위하여 위의 AI에 관한 권고안, 협약, 지침 등에서 공통적으로 제시하고 있는 주요 원칙들을 도출하여 이를 분석의 기준으로 삼았다. 그 여섯 가지 원칙으로는 1. 인간 존엄성 및 인간중심,[6] 2. 개인정보 보호(프라이버시) 및 데이터 보호, 3. 투명성 또는 설명가능성, 4. 공정성, 5. 안전성, 6. 책임성이다.[7]

〈표 3〉 본 연구의 분석기준으로서 AI 원칙들

기준	개념
Human dignity & rights/human-centric (인간 존엄성, 인권)/ 인간중심)	• AI 시스템은 인간의 자율성과 의사결정을 지지해야 하며 민주적이고 번영되며 평등한 사회에 대한 가능자로서 기본권 신장에 기여해야 함. 인간의 개입과 감독, 감시 및 결정과 관련하여, 의사결정과 행동에서 AI 시스템에 의존할 수는 있으나 인간의 책임을 대체할 수는 없음. 인간의 감독은 인간참여(human-in-the-loop, HITL), 인간지배(human-on-the-loop, HOTL), 인간지휘(human-in-command, HIC) 접근 등의 거버넌스 메커니즘을 통해 성취될 수 있음.
Personal information protection/ Privacy (개인정보보호 /프라이버시)	• AI 시스템의 데이터는 국제법 및 권고사항에 명시된 가치와 원칙에 따라 수집, 사용, 공유, 보관 및 삭제해야 하며, AI 행위자는 AI 시스템 수명주기 전반에 걸쳐 개인정보가 보호되도록 AI 시스템의 설계 및 구현에 책임을 져야 함. AI 시스템은 개인의 선호 뿐 아니라 성적 취향, 연령, 성별, 종교 또는 정치적 견해 등을 추론할 수 있으므로, 이러한 개인의 수집된 데이터가 불법적이거나 부당하게 차별하는 데 사용되지 않도록 해야 함. 데이터 수집시 사회적으로 구성된 편견, 부정확성, 오류 등이 포함될 수 있으므로 데이터의 무결성이 보장되어야 함 (UNESCO, 2021).
Transparency and Explainability (투명성/설명가능성)	• AI 행위자는 AI 시스템에 대한 투명성과 책임있는 공개를 보장하고, AI 시스템의 능력과 한계를 포함한 전반적 이해를 제고하고, 이해관계자들에게 AI 시스템과의 상호작용을 알리며, 데이터 출처, 요소, 프로세스 및 예측/내용/권고 또는 결정에 이르는 로직 등에 대해 이해하기 쉬운 정보를 제공할 것(OECD, 2019, 2024). (본 연구의 기준에서는 알고리즘의 설명가능성과 기술적 투명성을 모두 포함함)

4) OECD(2019, 2024)에서는 프라이버시와 데이터 보호, 공정성에 관하여 '법치, 인권 및 민주적 가치'와 관련한 내용에서 비차별 및 평등, 자유, 존엄성, 개인의 자율성, 다양성, 공정성, 사회정의 및 국제적으로 인정된 노동권 등과 함께 다루고 있다.

5) 이 협약은 2024년 5월 17일에 유럽평의회의 각료위원회에서 채택되었고, 2024년 9월 5일부터 서명이 개시되었다. 협약에 서명한 국가는 안도라, 조지자, 아이슬란드, 노르웨이, 몰도바 공화국, 산마리노, 영국, 이사라엘, 미국, EU이며, EU의 AI Act와 완전히 양립가능하다.

6) EU(2018)는 관련하여 '인간의 개입과 감독'을 이 분류에 포함하고 있다고 보인다.

7) 다만, 이 원칙들은 AI에 대한 규제적 측면에서 공통적으로 지향하고 있는 가치들이라는 점에서, AI의 효율성과 시장 및 산업 육성, 교육 분야 등에 관한 내용은 분석기준에 포함하지 않았다. 또한, 지적재산권 및 저작권 분야는 현재 합의된 국제 규범이 없고, 여러 기구와 국가에서 다양한 관점으로 논의 중이라는 점에서 본 연구에서는 제외하였다.

Ethics/ Fairness/Bias/ non-discrimination (윤리/공정성/ 편향성/ 비차별)	• AI 행위자는 국제법을 준수하고 사회정의의 증진과 공정성 및 차별 금지를 보호해야 함. 다양한 연령과 문화, 장애인, 여성, 취약하고 소외된 사람들의 요구를 고려하여 AI 기술의 이점을 모든 사람들이 동등하게 접근할 수 있도록 하는 포괄적인 접근 방식을 채택해야 함(UNESCO, 2021). AI 시스템에서 사용하는 데이터 셋은 의도하지 않은 불완전성이 포함되어 편견이 지속될 수 있으며, 특정 집단이나 사람들에 대한 편견과 차별, 소외를 악화시킬 수 있음. 따라서, 식별가능하고 차별적인 편견은 가능한 한 수집단계에서 제거되어야 함. 여기에 국가간 및 국가내 디지털 격차의 해소도 포함됨.
Safety, security/ Reliability/Robustness (안전성/신뢰성/견고함)	• AI 시스템이 전체 수명주기 동안 견고하고 안전하며 보안이 강화되어야 하며, 정상적 사용 및 오용, 기타 불리한 조건 하에서도 적절하게 작동하여 안전 또는 보안 위험을 초래하지 않아야 한다(OECD, 2024)
Accountability/ Responsibility (책무성/책임성)	• AI 시스템이 내린 결정과 행동에 대한 윤리적, 법적 책임(ethical responsibility and liability)을 포함(UNESCO, 2021). AI 시스템의 수명주기 동안 이루어진 데이터, 프로세스, 결정과 관련된 추적성(traceability)의 보장, 각 단계의 체계적인 위험관리 접근, AI 시스템 관련 위험*을 해결하기 위한 책임있는 행동이 채택되어야 함(OECD, 2019, 2024). (*이러한 위험에는 유해한 bias, 안전, 보안 및 개인정보를 포함한 인권, 노동 및 지적재산권 관련 위험이 포함됨)

3. 분석방법

본 연구는 실증분석에 기반한 연구결과를 도출하기 위해 다음과 같은 절차를 따랐다. 본 연구에서 AI 규범들에 대한 질적 내용분석을 위한 코딩은 전체 24개 문헌 및 총 1,749개의 구문이라는 방대한 분량을 대상으로 한다. 이를 보다 체계적으로 수행하기 위하여 먼저 연구책임자가 분석대상인 국제 규범들로부터 공통된 AI 윤리원칙을 도출한 후, 분석을 위한 각 원칙별 개념과 분석기준을 정하였다. 이를 토대로 국제학 및 철학 전공인 2명의 연구보조원에게 각 원칙 및 유사어 키워드 추출 방식으로 각 구문들을 원칙별 셀에 구분하여 기입하도록 지시하였다. 또한, 코딩의 신뢰성과 일관성 유지를 위하여 각 보조원이 수행한 업무는 서로 교차검토하도록 하고 매주마다 연구책임자의 확인과 수정 작업을 거쳤다. 전체 구문들의 분석은 2024년 8월 말부터 10월 말까지 수행되었다.

규범들의 내용분석 절차는 다음과 같다. (1) 먼저 AI에 관한 24개 규범들을 각 조항 및 구문별로 분절화하였다. 법률은 조항 및 세부항목별로, 선언이나 지침의 형태의 경우에는 소단락(문단)별로 구분하였다. (2) 각 규범들에는 그 주체에 따라 한국은 'KOR', 유럽연합은 'EU', 미국은 'US', OECD는 'OE', UN은 'UN', UNESCO는 'UNE', 영국은 'UK'라는 국가코드를 부여하고, 제정 순서에 따라 주체별 코드 뒤에 숫자를 부여하였다. (3) 전체 24개 규범들을 조항, 단락 및 문단별로 분절화한 결과, 총 1,749개 구문이 산출되었다. (4) 이들에 대해서는 각 구문별로 해당되는 AI의 원칙들에 0과 1로 코딩하였으며, 복수의 원칙들에 해당하는 경우에는 해당 원칙들에 대해 모두 1로 입력하였다. (5) 각 규범들별로 전체 구문 중 각 원칙

에 해당하는 구문의 비중(%)을 산출하였으며, 이를 원칙들의 수를 고려하여 수치화하였다.

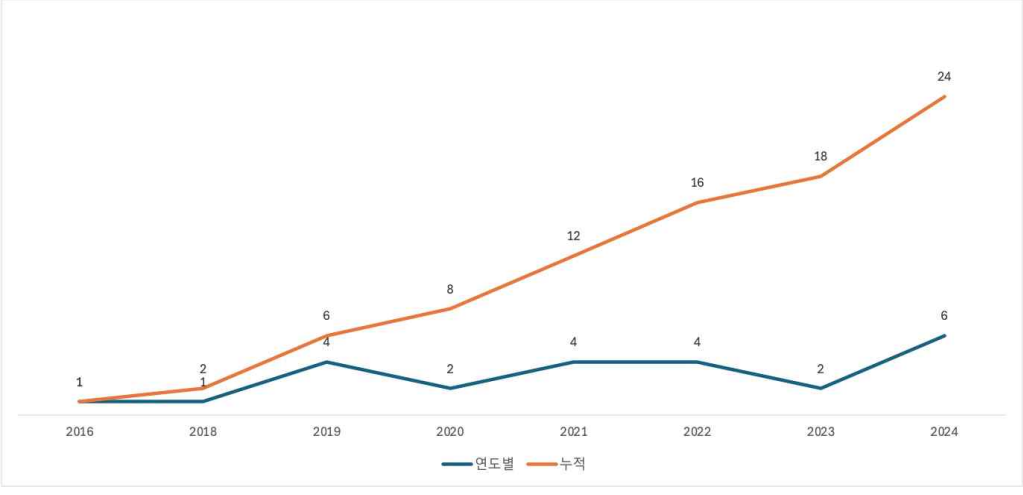
Ⅳ. 분석결과

1. 전체 분석대상에 대한 분석

1) AI 규범들의 개괄적 흐름

먼저, 전체 24개 AI 규범들이 제정된 개괄적 흐름을 살펴보면, 2016년에는 EU가 'GDPR(유럽 일반 개인정보보호법)'을 제정하였고 2018년에는 'AI 전략'을 제시하였으며, 2019년에는 미국의 '인공지능 선도적 지위 유지를 위한 인공지능 리더십 유지 행정명령(제13859호)'과 EU의 '신뢰할 수 있는 AI를 위한 윤리지침', 그리고 'OECD AI 권고안'이 제시되었다. 한국도 2019년 '인공지능 국가전략'을 수립하였다. 2020년에는 미국이 '연방정부의 신뢰할 수 있는 인공지능 사용 촉진을 위한 행정명령(제13960호)'을 발표하였고, 한국은 '인공지능 윤리기준'을 제정하였다. 2021년에는 미국의 '국가인공지능계획법'과 EU의 'AI Act' 초안, 영국의 '인공지능 국가전략'이 제정되었고, 유네스코에서는 '인공지능 윤리 권고'를 발표하였다. 2022년 EU는 개별법인 '디지털시장법(Digital Market Act)'과 '디지털서비스법(Digital Services Act)'을 제정하였고, 미국은 '인공지능훈련법'과 '인공지능진흥법'을 제정하였는 바, 여기서 규제에 초점을 둔 EU와 AI 산업의 육성과 촉진에 중점을 둔 미국간 태도의 차이가 드러난다. 2023년에는 미국은 '안전하고 보안이 보장되며 신뢰할 수 있는 인공지능의 개발과 사용에 관한 행정명령(제14110호)'을, 영국은 'AI Safety Summit(안전성 정상회의)' 개최를 통해 '블레츨리 선언'을 발표하였다. 2024년이 되어 EU는 마침내 AI Act을 법률로 통과시켰고, UN총회의 'AI에 관한 결의안'과 'OECD AI 권고' 개정안, EU-미국-영국의 'AI에 관한 협약'(Convention on Artificial Intelligence and Human Rights Democracy and the Rule of Law)이 발표되었다. 한국의 경우 AI 안전성 회의의 차기 개최지로서 2024년 5월 'Seoul AI Summit'을 개최하여 서울 선언을 발표하는 한편, '인공지능의 책임있는 군사적 이용에 관한 고위급회의(REAIM)' 개최를 통해 그 청사진을 제시하였다.

〈그림 1〉 AI 규범의 제정 현황



2) AI 원칙들의 흐름

AI 규범들에 나타난 원칙들을 살펴보면, 먼저 '인간 존엄성 및 인간중심 가치'는 UNESCO의 인공지능 윤리 권고(2021)에서 강조된 후 미국의 인공지능진흥법(2022), 영국의 AI 안전성 정상회의(2023), EU-미국-영국의 AI에 관한 협약(2024)에서 강조되었고, '프라이버시 또는 개인정보보호' 원칙의 경우에는 EU의 GDPR(2016)에서 강조된 후 미국에서 연방정부의 신뢰할 수 있는 인공지능 사용 촉진을 위한 행정명령(2020) 및 인공지능진흥법(2022), 영국의 AI 안전성 정상회의(2023)에서 강조되었다. '투명성' 원칙의 경우에는 EU의 GDPR(2016)에서 강조된 이후 크게 부각되지 않다가 디지털서비스법(2022)과 영국의 AI 안전성 정상회의(2023)에서 다시 강조되었다. 특히 디지털 서비스법은 온라인 플랫폼 상에서 개인정보들의 사용과 보호에 관한 엄격한 기준들을 정하고 있기에 이들 조항들의 비중이 크게 나타난 것으로 보인다. '안전성' 원칙은 2023년 영국의 AI 안전성 정상회의 이전에는 크게 부각되지 않았다가 당해 블레츨리 선언에서 강조(14.1%)되고 이후의 규범에서는 다른 원칙들과 유사한 비중을 이루었다. '책임성' 원칙의 경우에는 GDPR(2016)에서 강조된 후 2022년 EU에서 제정된 디지털시장법과 디지털서비스법에서 다시 비중이 늘었으며, 영국의 AI 안전성 정상회의(2023)와 EU의 AI 법(2024)에서 부각되었다. '공평성' 원칙의 경우 OECD의 AI에 관한 권고안(2019)에서 그 중요성을 드러낸 후(5.2%) 한국의 인공지능 윤리기준(2020)에서도 인권(6.9%) 다음으로 높은 비중을 차지하였다(5.9%). 이 원칙은 한국 뿐 아니라 UNESCO(2021), UN(2024) 등 국제기구에서 제안하는 규범들에서도 강조되고 있다.

이를 종합하면, 전체 1,749개 구문 중에서 인권 및 인간중심 원칙에 해당하는 구문이 차지하는 비중은 2.7%(285개 구문)로 나타났으며, 프라이버시 또는 개인정보보호 원칙은 2.2%(235개 구문), 투명성 원칙은 2.1%(220개 구문), 공평성 원칙은 2.4%(254개 구문), 안전성 원칙은 2.3%(246개 구문), 책임성 원칙은 3.8%(395개 구문)로 나타났다. 분석대상이 AI에 관한 규제와 기술 촉진의 내용을 포함하고 있다는 점에서, 책임성 원칙의 비중(3.8%)이 가장 높게 나타나면서도 AI의 효율성과 혁신성 또한 강조되었다고 볼 수 있다. 또한 전체 원칙들 중 인권의 비중(2.7%)이 책임성 다음으로 크다는 것은 AI가 궁극적으로는 인간 중심으로 설계되고 활용되어야 한다는 정책적 지향성이 나타난 것으로 볼 수 있다.

〈표 4〉 AI 규범의 원칙별 비중

(단위: 해당기준 구문 수, %)

연도	코드	AI 규범	인권	프라이버시	투명성	공평성	안전성	책임성	전체 구문
2016	EU1	EU 유럽 일반 개인정보보호법 [GDPR]	13 (2.2)	53 (8.9)	37 (6.2)	3 (0.5)	28 (4.7)	59 (9.9)	99
2018	EU2	EU AI 전략	12 (3.1)	7 (1.8)	4 (1.0)	12 (3.1)	7 (1.8)	8 (2.1)	64
2019	US1	미국의 인공지능 선도적 지위 유지를 위한 인공지능 리더십 유지 행정명령 제13859호	3 (1.4)	5 (2.4)	0 (0.0)	0 (0.0)	6 (2.9)	0 (0.0)	35
2019	EU3	EU 신뢰할 수 있는 AI를 위한 윤리 지침	31 (4.1)	7 (0.9)	12 (1.6)	24 (3.2)	18 (2.4)	29 (3.8)	126
2019	OE1	OECD AI 권고안	4 (4.2)	4 (4.2)	1 (1.0)	5 (5.2)	3 (3.1)	3 (3.1)	16
2019	KR1	한국 인공지능 국가전략	6 (1.2)	4 (0.8)	0 (0.0)	7 (1.4)	4 (0.8)	2 (0.4)	83
2020	US2	연방정부의 신뢰할 수 있는 인공지능 사용 촉진을 위한 행정명령 제13960호	2 (3.3)	3 (5.0)	1 (1.7)	0 (0.0)	1 (1.7)	1 (1.7)	10
2020	KR2	한국 인공지능(AI) 윤리기준	7 (6.9)	2 (2.0)	1 (1.0)	6 (5.9)	2 (2.0)	2 (2.0)	17
2021	US3	미국 국가인공지능계획법	0 (0.0)	4 (1.0)	0 (0.0)	8 (2.0)	5 (1.3)	2 (0.5)	66
2021	EU4	EU AI Act 초안	35 (3.4)	32 (3.1)	20 (1.9)	20 (1.9)	20 (1.9)	37 (3.5)	174
2021	UK1	영국 국가인공지능전략	11 (1.0)	11 (1.0)	6 (0.5)	30 (2.7)	32 (2.9)	30 (2.7)	185
2021	UNE	유네스코 인공지능 윤리 권고	46 (5.0)	20 (2.2)	16 (1.8)	47 (5.2)	17 (1.9)	28 (3.1)	152

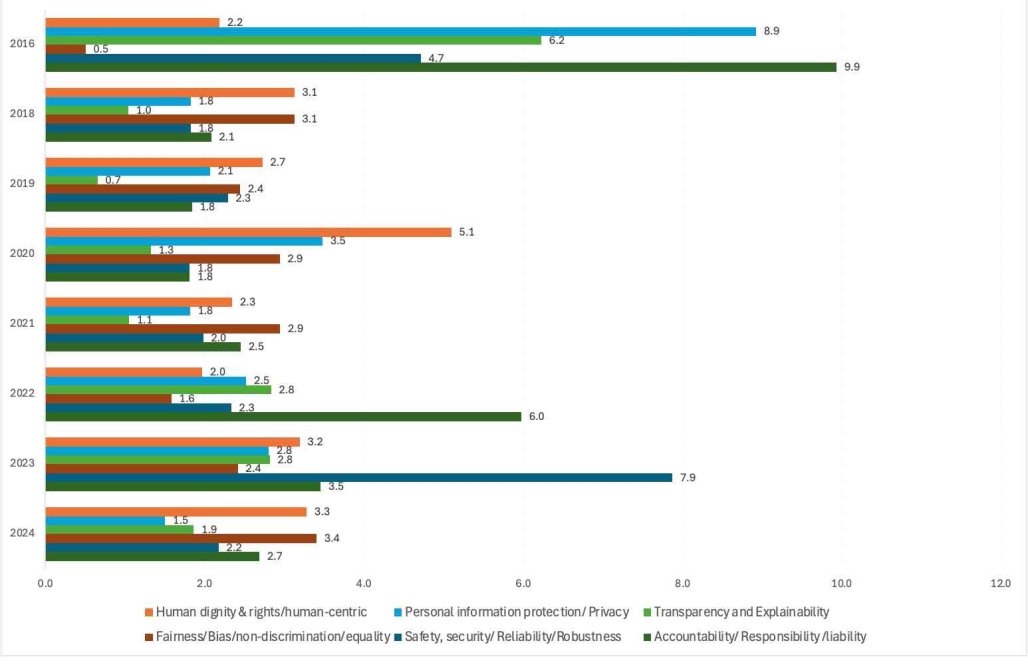
2022	EU5	디지털시장법 (Digital Markets Act)	3 (0.9)	3 (0.9)	11 (3.4)	4 (1.2)	3 (0.9)	40 (12.3)	54
2022	US4	미국 인공지능훈련법	0 (0.0)	1 (1.3)	1 (1.3)	2 (2.6)	3 (3.8)	1 (1.3)	13
2022	EU6	디지털서비스법 (Digital Services Act)	13 (2.3)	13 (2.3)	32 (5.7)	9 (1.6)	10 (1.8)	47 (8.4)	93
2022	US5	미국 인공지능진흥법	5 (4.6)	6 (5.6)	1 (0.9)	1 (0.9)	3 (2.8)	2 (1.9)	18
2023	US6	안전하고 보안이 보장되며 신뢰할 수 있는 인공지능의 개발과 사용에 관한 행정명령 제14110호	10 (1.3)	14 (1.8)	4 (0.5)	18 (2.3)	13 (1.6)	4 (0.5)	132
2023	UK2	제1회 AI 안전성 정상회의 '블레츨리 선언'	4 (5.1)	3 (3.8)	4 (5.1)	2 (2.6)	11 (14.1)	5 (6.4)	13
2024	EU7	EU AI Act 최종안	57 (3.2)	36 (2.0)	60 (3.4)	34 (1.9)	48 (2.7)	77 (4.4)	293
2024	UN	UN총회 AI 결의안	2 (2.4)	2 (2.4)	3 (3.6)	6 (7.1)	3 (3.6)	1 (1.2)	14
2024	OE2	OECD AI 권고 개정안	3 (2.5)	2 (1.7)	3 (2.5)	4 (3.3)	2 (1.7)	6 (5.0)	20
2024	EUK	Convention on Artificial Intelligence and Human Rights Democracy and the Rule of Law	13 (6.0)	1 (0.5)	1 (0.5)	4 (1.9)	3 (1.4)	8 (3.7)	36
2024	KR3	서울 정상회의 서울선언	2 (3.7)	1 (1.9)	0 (0.0)	1 (1.9)	1 (1.9)	0 (0.0)	9
2024	KR4	REAIM 행동을 위한 청사진	3 (1.9)	1 (0.6)	2 (1.2)	7 (4.3)	3 (1.9)	3 (1.9)	27
전체			285 (2.7)	235 (2.2)	220 (2.1)	254 (2.4)	246 (2.3)	395 (3.8)	1,749

이에 따라, 2016년에는 개인정보보호(프라이버시)(8.9%)와 책임성(9.9%)의 비중이 가장 컸고, 2020년에는 인권(5.1%)이 강조되었으며, 2022년에는 책임성(6.0%), 2023년에는 안전성(7.9%)이 두드러지게 나타났다. 이러한 초기 흐름의 배경에는 EU의 2016년 유럽 일반 개인정보보호법(GDPR: General Data Protection Regulation) 제정이 기여한 바가 컸다고 볼 수 있다. 본 법은 EU 회원국 뿐 아니라 EU에 법인 또는 지점이 있는 외국 기업들까지 개인정보보호 및 데이터 관리에 관하여 준수해야 하는 의무를 가지며, 그에 따라 책임성 담보와 개인정보 처리에 관한 조항들이 많은 비중을 차지할 수밖에 없기 때문이다. 그러나 이후의 EU의 'AI 전략' 및 '신뢰할 수 있는 AI를 위한 윤리지침', 미국의 '선도적 지위 유지를 위한 인

공지능 리더십 유지를 위한 행정명령', OECD의 'AI 권고안' 등에서는 6가지 원칙들이 대체로 균형있는 비중으로 나타났다. 그러다가 2023년이 되면서 그간 AI 산업 및 시장 육성, 혁신을 강조해오던 미국에서도 점차 AI의 위험과 안전성, 인권 및 민주주의 가치에 대한 인식을 AI 규범에 담기 시작하였다. 2023년 바이든 대통령은 AI에 관한 첫 행정명령(제14110호)에서 '안전하고 신뢰할 수 있는 AI의 개발과 사용'을 강조하기 시작했다. 이는 영국 블레츨리 파크에서 개최된 '제1회 안전성 정상회의'와 함께 이후의 국제사회 규범들과 AI에 관한 국제협약들에서 '안전'이 가장 핵심적 가치로 포함되는 분기점을 형성하였다.

〈그림 2〉 연도별 AI 원칙들의 비중

(%: 복수 기준)



2. 주체별 분석

1) 원칙별 분석에 따른 정책적 지향

24개 AI에 관한 규범들에 대하여 주요 행위자별로 중요하게 다룬 원칙들에 대해 살펴본 결과, 먼저 EU의 경우에는 책임성(5.5%)이 가장 높게 나타났고, 투명성(3.3%), 인권(3.0%), 프라이버시(2.8%), 그리고 안전성(2.5%) 및 공평성(2.0%) 순으로 나타났다. 이는 미국의 경우 프라이버시(2.0%), 안전성(1.9%), 공평성(1.8%) 기준이 상대적으로 높고, 인권(1.2%), 책임성

(0.6%)과 투명성(0.4%) 기준이 낮게 나타난 점과 대비된다. OECD, UNESCO, UN 등 주요 국제기구들의 규범에서는 공평성(3.8%)이 가장 강조되었고, 인권(3.2%) 및 책임성(3.1%) 기준이 높게 나타난 가운데 안전성(2.7%), 프라이버시(1.7%), 투명성(1.3%) 순으로 강조되었다. 한국의 경우에는 주요 국제기구와 같이 공평성(2.6%)과 인권(2.2%)이 가장 강조되고 안전성(1.2%), 프라이버시(1.0%), 책임성(0.9%), 투명성(0.4%) 순으로 비중이 높게 나타났다.

〈표 5〉 주체별 AI 원칙들의 비중

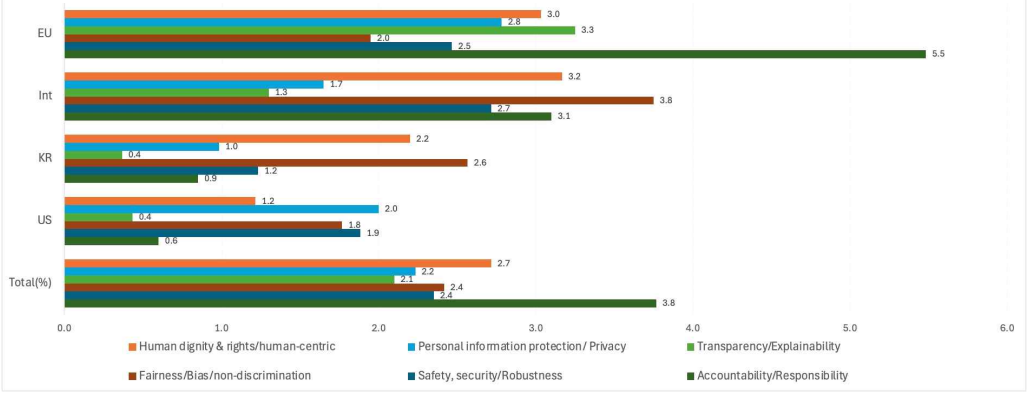
(%: 전체 구문 중 해당 기준)

주체 (규범수)	인권/ 인간중심	프라이버시	투명성	공평성	안전성	책임성	전체 구문수(%)
EU(7)	164 (3.0)	151 (2.8)	176 (3.3)	106 (2.0)	134 (2.5)	297 (5.5)	903 (100)
국제기구(7)	83 (3.2)	43 (1.7)	34 (1.3)	98 (3.8)	71 (2.7)	81 (3.1)	436 (100)
한국(4)	18 (2.2)	8 (1.0)	3 (0.4)	21 (2.6)	10 (1.2)	7 (0.9)	136 (100)
미국(6)	20 (1.2)	33 (2.0)	7 (0.4)	29 (1.8)	31 (1.9)	10 (0.6)	274 (100)
전체(24)	285 (2.7)	235 (2.2)	220 (2.1)	254 (2.4)	246 (2.4)	395 (3.8)	1,749 (100)

주: 각 기준별 비중은 법안 또는 규범들을 구분하여 전체 구문 중 각 기준에 해당하는 비중이며, 해당 기준에 중복되는 경우를 포함하였으므로 기준 수(6)로 나누어 값을 산출함.

〈그림 3〉 주체별 AI 원칙들의 비중

(%: 복수 기준)



AI 규범의 주체별로 이처럼 다른 비중의 분석결과가 나타난 것은 정책이나 규범에 그 국

가나 주체의 맥락적 특징이 반영된다는 점에 기인한다. 동일한 대상에 대한 규제라 하더라도 그 성격은 각 국가들의 역사적, 사회경제적, 정치적 및 제도적 환경과 기술발달 단계에 따라 다르게 나타날 수 있기 때문이다. EU의 경우 시민의 헌법상 권리를 중시하며 엄격한 규제체계로 이를 담보하려는 법철학이 AI 법에 반영되었다면, 미국의 규범들에는 민간주도의 혁신과 기술발전을 지원하면서도 기본권 보호를 위해 최소한의 규제를 도입하려는 전략적 방향성이 담겨있다고 볼 수 있다. 한국은 이들의 영향을 받으면서 인공지능 산업의 육성과 진흥이라는 종래 목표와 인간의 기본적 권리와 존엄성 보호라는 가치를 지키면서도 형평성을 고려한 AI 규범체계를 지향하고 있는 것으로 나타난다.

2) 시간 흐름에 따른 분석

(1) EU

가장 먼저 EU는 2016년 유럽 일반개인정보보호법(GDPR)을 제정하면서 국제사회에서 개인정보보호 분야에서 선두적 위치로 자리잡기 시작하였다. 본 법은 28개 EU 회원국 전체에 적용되고 유럽경제지역(EEA) 국가(노르웨이, 아이슬란드 등)와 영국 등이 참여하고 있으며, 법 제정 이후에는 일본과 한국, 중국이 각각 개인정보보호법을 제·개정(APPI, 2020; PIPA, 2020, 2024; PIPL, 2021)하는 등의 영향이 나타났다. 이러한 파급효과는 아시아 지역뿐만 아니라 이스라엘(2017년), 케냐(2019년), 나이지리아(2019년), 남아프리카공화국(2020년)의 포괄적 데이터 보호법 및 브라질의 일반 개인정보보호법(LGPD) 제정에도 미쳤다고 할 수 있다. 이러한 개별 국가들의 후속적 움직임을 통해 EU의 GDPR은 전세계적인 데이터보호법의 모델로서 일명 '브뤼셀 효과'를 일으켰다고 평가되고 있다.

EU가 2018년에 제시한 인공지능 전략(Communication: Artificial, Intelligence for Europe)[8] 역시 유럽 내 프랑스(2018) 및 독일(2018)과 한국(2019)과 영국(2021)의 인공지능 국가전략 형성에 영향을 미쳤다고 볼 수 있다. 유럽은 이후 Horizon Europe 및 Digital Europe 프로그램을 통해 AI에 관한 투자를 확대하는 한편 그 위험에 대비한 규제 마련을 위한 논의도 본격화하고 있다.

AI에 관한 일반적 규범 마련에 있어서 EU의 선도적 역할이 두드러진 것은 2019년 '신뢰할 수 있는 AI를 위한 윤리지침(Ethics Guidelines for Trustworthy AI)'과 2021년 'AI 법' 초안의 발표로 나타난다. 특히 이 윤리지침에 나타난 주요 원칙과 가치들은 이후 국제사회의

8) 캐나다는 국가단위로는 세계 최초로 '인공지능 국가전략(National AI Strategy)'을 발표하였으며, 국가연구소로서 Edmonton에 Amii, Montreal에 Mila, Toronto에 Vector 연구소 등 세 개의 연구소가 있다(https://cifar.ca/ai/).

AI에 관한 규제적 논의와 규범들의 토대를 이룬다.

EU는 또한 2022년 디지털시장법(DMA)과 디지털서비스법(DSA) 제정을 통해 독점적 디지털 서비스 회사(gatekeeper) 및 대규모 온라인 플랫폼(VLOP)을 규제하고 이용자의 데이터보호를 위한 글로벌 규제의 표준을 형성함에 있어서 선도적 역할을 하였다. 법에서 게이트키퍼로 지정된 기업들(Alphabet, Amazon, Apple, ByteDance, Meta, and Microsoft)은 DMA의 모든 규정을 준수해야 하므로, 독일이나 브라질 등 다른 국가들에서도 그에 해당하는 규정을 마련하는 등 EU의 경로를 따르게 되었으며, 영국 역시 이와 유사한 성격의 DMCC 법안을 준비하게 되었다.[9] DSA의 경우 2024년 2월부터 유럽 내 모든 플랫폼에 적용되게 되면서 DSA의 후속조치로서 디지털서비스 조정관(DSC)을 지정한 국가들이 나타났으며, 유럽 국가들 중에는 오스트리아, 불가리아, 덴마크, 핀란드, 헝가리, 이탈리아, 룩셈부르크, 루마니아, 스페인 등이 있다.[10]

AI에 관한 규제적 규범에서 EU가 남긴 가장 선도적 자취는 범용 AI에 관한 규제를 포함한 'AI 법'이 2024년에 최종적으로 결정된 것이라 할 수 있다. 확정된 법률은 AI 기술의 위험성 수준에 따른 접근(risk-based approach)과 생성형 AI를 포함한 범용 AI에 대한 규제를 포함하고, 법률 위반시 제재조치를 명시하는 등 강한 법적 책임성을 담보하고 있다. 이러한 규제적 흐름과 함께 EU는 2024년 5월, 영국, 미국 등과 함께 'AI와 인권, 민주주의, 법치에 관한 유럽평의회 기본협약'을 채택함으로써 국제적 프레임워크를 만드는 경로로 나아가고 있다.[11]

(2) 국제기구

EU가 개인정보보호 및 윤리적 가이드라인, 포괄적 규제법에서 선도적 역할을 주도하고 있지만, OECD, UNESCO, UN 등 국제기구 차원에서는 보다 일반적이고 거시적으로 AI 거버넌스의 필요성에 대한 합의가 형성되기 시작하였다. OECD는 2019년 인공지능에 관한 최초의 정부간 표준이라 할 수 있는 '인공지능에 관한 권고안(Recommendation on AI)'을 제시하였다.[12] 이 권고안에서는 인권과 민주적 가치에 대한 존중을 보장하면서 신뢰할 수

9) https://itif.org/publications/2024/03/07/the-brussels-effect-how-the-digital-markets-act-projects-european-influence/

10) https://www.euronews.com/next/2024/02/07/platform-rules-to-apply-this-month-as-member-states-lag-on-oversight-personnel

11) 이 협약은 2024년 5월 17일에 유럽평의회의 각료위원회에서 채택되었고, 2024년 9월 5일부터 서명이 개시되었다. 협약에 서명한 국가는 안도라, 조지자, 아이슬란드, 노르웨이, 몰도바 공화국, 산마리노, 영국, 이사라엘, 미국, EU이며, EU의 AI Act와 완전히 양립가능하다(https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence).

12) 여기에 OECD 36개 회원국과 아르헨티나, 브라질, 콜롬비아, 코스타리카, 페루, 루마니아 등이 'OECD

있는 AI의 책임 있는 관리를 촉진함으로써 AI에 대한 혁신과 신뢰를 육성하는 것을 목표로 제시하고 있다. 같은 해 6월, G20은 그 후속조치인 '오사카 정상회의'를 개최하여 AI에 대한 논의를 심화하였으며, 2020년에는 '인공지능에 대한 글로벌 파트너십(GPAI: Global Partnership on AI)'이 창립되어 책임성있는 AI의 발전과 활용을 위한 국제적 이니셔티브 공유 협의체가 구성되었고 한국은 그 창립국으로서 국제사회에 주도자로서 나서기 시작하였다.[13] UNESCO는 2021년 제41차 총회에서 'AI 윤리권고(Recommendation on the Ethics of AI)'를 제시하였는 바, 여기에 인권, 근본적 자유, 인간 존엄성의 존중과 보호 및 증진, 환경 및 생태계의 번영, 다양성과 포용성의 보장, 평화롭고 정의로운 상호연결된 삶 등의 4대 가치를 포함하고 있다. 이 권고안은 193개 회원국의 만장일치로 채택되었다.

2023년에는 영국이 AI의 안전성과 관련된 최초의 국제 협약이라 할 수 있는 '제1회 AI Safety Summit'을 개최하여 AI의 안전성을 확보하기 위한 국제적 공동 대응을 모색하는 '브레츨리(Bletchley) 선언'을 발표하였고, 한국, 호주, 브라질, 캐나다, EU, 프랑스, 독일, 일본, 미국 등 28개국이 이를 채택하였다. 그 후속조치 상황을 중간점검하는 미니 회의('서울 정상회의')는 후년인 2024년 5월에 한국이 영국과 공동으로 개최하기로 정하였다.[14] AI에 관한 안전성(Safety) 원칙을 강조하는 이러한 흐름은 2024년 3월 UN총회의 'AI 결의안' 채택으로 이어졌다. 이는 EU의 AI 법 이후 나온 세계 최초의 글로벌 결의안이자[15] 그동안의 AI에 관한 국제적 규범들에서 확인된 주요 원칙들에 대해 공동체적 합의가 이루어졌음을 의미한다. 미국 주도 하에 120개국 이상이 공동발의한 이 결의안은 인권 및 데이터 보호와 인공지능의 위험을 모니터링할 것을 촉구하는 것을 주요 내용으로 한다.

OECD는 2019년 권고안 이후 워킹페이퍼(2023)를 통해 영국, 캐나다, 미국, EU, 브라질, 중국, 주요 국제기구의 AI 전략 및 지침에 관한 후속적 동향을 보고[16]한 이후, 2024년에는 'AI 권고안(Recommendation of the Council on AI)'을 개정하여 기존의 원칙들에서의 '인간 중심 가치와 공정성' 부분을 '법치, 인권, 민주주의 가치에 대한 존중'에 대한 강조와 함께

Principles on AI'에 서명했다.

13) GPAI는 인권, 포용성, 다양성, 혁신 및 경제성장에 근거하여 다양한 인공지능 관련 이슈를 다루는 국제적 다중이해관계자적 협의체로서, 15개 창립회원국은 한국, 프랑스, 캐나다, 호주, 미국, EU, 독일, 이탈리아, 일본, 뉴질랜드, 싱가폴, 영국, 슬로베니아, 멕시코, 인도이며, 인권, 근본적 자유와 민주적 가치에 기반하여 '책임성 있고 인간중심적인 인공지능의 발전과 활용'을 지지하는 공동선언문을 발표하였다.

14) 이 회의를 통해 'Seoul Declaration for Safe, Innovative, and Inclusive AI'를 채택하였다.

15) 다만, 구속력없는 발의안이라는 점에서 한계를 가진다.

16) OECD(2023), The State of Implementation of the OECD AI Principles four years on. (https://www.oecd.org/en/publications/the-state-of-implementation-of-the-oecd-ai-principles-four-years-on_835641c9-en.html).

여기에 공정성과 프라이버시를 포함하도록 변경하였다. 합법성과 민주적 질서를 중시하면서도 AI가 생성한 편향된 결과로 인해 인간의 기본적 자유와 평등에 대한 기본적 권리가 침해되지 않도록 AI 규범이 지켜야 할 기본적 원칙을 명확히 하였다고 볼 수 있다. 또한, 할루시네이션(Hallucination) 이슈를 통해 제기된 바와 같이 허위 정보에 대한 처리 및 관리와 목적 외 또는 의도적, 비의도적 오용으로부터의 보호와 정보의 무결성, 투명성과 책임 있는 공개를 보장하기 위해 AI 행위자들이 AI 시스템과 관련하여 제공해야 하는 정보를 명확히 하도록 제안하고 있다.

가장 최근에 이루어진 다자간 AI 규범 논의는 EU, 미국, 영국이 2024년 9월 체결한 'AI와 인권, 민주주의 및 법치에 관한 기본 협약'(The Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law)이라 할 수 있다. 이는 AI 기술이 인권과 민주주의 가치를 준수하고 기술발전에 도움이 되도록 보장하는 세계 최초의 법적 구속력 있는 AI 조약에 해당한다. 당해 조약은 AI 시스템의 전 주기를 관리 감독하는 것을 목표로, AI 시스템의 영향을 받는 사람들의 인권 보호, 민주주의 제도와 절차, 사법부 독립 등 권력분립 원칙, 인간 존엄성과 자율성, 투명성과 감시, 책임성, 형평성, 프라이버시, 신뢰성 및 안전성에 초점을 두고 있다.[17]

(3) 미국

미국은 국가의 미래 경제와 안보를 위한 AI의 전략적 중요성을 인식한 이후, 첫 국가 AI 전략을 발행하는 등 인공지능에 관한 미국의 리더십을 강화하기 위한 노력을 기울여왔다고 볼 수 있다. 미국은 2019년 AI 연구개발 및 활용 분야에서 선도적 지위를 유지하고자 관련 기술의 개발과 적용에 대한 촉진과 신뢰를 형성하도록 하는 '인공지능 리더십 유지 행정명령 제13859호(Maintaining American Leadership in Artificial Intelligence)'를 발표하고, 여기에 나열된 국가안보전략을 실행하기 위해 '국가 AI 연구개발 전략계획 개정안'(The National Artificial Intelligence Research & Development Strategic Plan: 2019 Update)을 발표하였다.

미국의 국가안보 및 산업경제 중심의 방향성은 2020년 '연방정부의 신뢰할 수 있는 인공지능 사용 촉진(Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government)을 위한 행정명령(제13960호)'에서 다소 변화를 보였다. 미국은 이 행정명령을 통해 연방정부에서 AI를 설계, 개발, 취득하거나 사용할 시 준수해야 할 원칙을 규정하였는

17) 이는 그동안 미국이 주지해 온 가치들에 해당하는 '인권보호, 민주적 절차와 법치의 존중, 책임성, 비차별, 프라이버시, 신뢰성' 등의 가치에서 투명성과 안전성, 인간 또는 관리당국의 감독 등을 포함하는 보다 넓은 범위로 나아갔다고 볼 수 있다.

바, 그 주요 원칙으로는 a. 합법적이며, 미국의 가치를 존중할 것, b. 목적과 성과 중심일 것, c. 정확하고 안정적이며 효과적일 것, d. 안전성과 보안성 및 회복력이 있을 것, e. 이해할 수 있을 것, f. 인간의 책임을 명확히 하고 추적이 가능할 것, g. 정기적으로 모니터링할 것, h. 투명할 것, I. 책임을 부담할 것 등을 명시하고 있으나, 여전히 미국의 AI 규범에서 중요한 비중은 AI의 개발, 활용의 촉진이 차지하고 있었다. 이러한 흐름은 2021년 미국의 '인공지능 이니셔티브법(National Artificial Intelligence Initiative Act of 2020)'[18]과 '인공지능진흥법(Advancing American AI Act)' 및 '인공지능훈련법(Artificial Intelligence Training for the Acquisition Workforce Act)'[19]을 통해 AI 분야와 산업에서 미국의 리더십을 보장하기 위한 노력의 가속화로 나타났다.

이처럼 미국은 AI 개발 및 활용, 산업의 촉진 측면에서는 선도적 지위를 점하고자 하였으나, 책임성 및 개인정보보호 분야에 있어서의 법제적 규율에는 성공하지 못하였다는 특징을 보인다. '알고리즘책임법 또는 인공지능책임법'(Algorithmic Accountability Act)'의 경우 2019년과 2022년에 발의되었다가 모두 부결되었고,[20] 2023년에 다시 발의된 법안 역시 의회를 통과하지 못했다. 개인정보보호 분야에 있어서도 2022년 '연방 개인정보보호법(안)(ADPPA, American Data Privacy and Protection Act)'이 발의되었으나 임기만료로 폐기되었으며, 자국민의 개인정보를 무분별하게 수집·활용하는 것을 막기 위한 연방 차원의 포괄적 개인정보보호법인 '미국 프라이버시보호법(안)(American Privacy Rights Act, APRA)'이 2024년 4월 다시 발의되었으나, 여전히 의회 계류 중이다(2024년 12월 기준).

국제사회에서 AI에 관한 안전성이 강조되면서 미국에서도 관련 법령들에 변화를 보이기 시작하였다. 기존 트럼프 행정부의 행정명령들(제13859호 및 제13960호)과 이니셔티브 법(National AI Initiative Act of 2020)이 AI의 개발과 활용 및 촉진에 비중을 두고 있었다면,

---

18) 이 법안은 원래 2020년 3월 12일에 하원에 소개되었으나, 최종적으로 법률로 확정된 것은 2021년 1월 1일 하원과 상원이 대통령의 거부권을 무효화하고 법안을 통과시켰다. 이 법안은 2021년 회계연도 국방수권법(National Defense Authorization Act, DNAA)의 일부로 포함되어 통과되었다.

19) '미국 인공지능 진흥법(Advancing American AI Act)'(2022.12.23.)은 인공지능 분야에서 미국의 경쟁력을 강화하기 위해 인공지능 관련 프로그램 및 시책을 장려하는 것을 목적으로 하고 있으며, '인공지능 훈련법(Artificial Intelligence Training for the Acquisition Workforce Act)'(2022.10.17.)은 연방정부 행정기관 직원들의 지식과 역량 향상을 위한 목적으로 하고 있다.

20) 2019년 4월 10일에 미국 의회에 제출된 이 법안은, 상원에서는 Cory Booker 상원의원과 Ron Wyden 상원의원이 발의(S.1108)하였고, 하원에서는 Yvette Clarke 하원의원이 동일한 내용의 법안을 발의(H.R.2231)하였다. 2022년에는 민주당의 Ron Wyden, Cory Booker 상원의원, Yvette Clark 하원의원에 의해 발의되었다. 이 법안은 대규모 기술 기업들이 고용, 금융서비스, 의료, 주택 및 법률 서비스 등 다양한 분야에서 중요한 결정을 내리는 자동화된 의사결정 시스템에 대해 편향성 영향평가를 수행하도록 요구하는 내용을 담고 있다.

바이든 행정부의 행정명령(제14110호)에서는 보다 '안전하고 보안이 보장되며 신뢰할 수 있는 인공지능의 개발과 사용(Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)'을 내세우면서 연방정부와 기업의 책임을 강화하는 규제적 성격의 비중이 높아졌다. 바이든 행정부의 이 행정명령은 AI 개발자의 안전성 평가 의무화 및 AI 도구 표준 마련, 개인정보보호 강화 등을 주요 내용으로 하고 있다. 이후 미국은 EU, 영국 등과 함께 'AI와 인권, 민주주의 및 법치에 관한 기본 협약'(2024)을 체결함으로써 AI에 관한 규범의 초점이 혁신과 촉진에서 신뢰와 안전성 지향의 방향으로 옮겨가고 있다고 보인다.

(4) 한국

한국은 앞서 논의한 국제행위자들에 비해 상대적으로 후발자적 위치에서 AI에 관한 규범을 이행하고 있다. 한국 정부는 2016년에 '제4차 산업혁명 대응 계획'을 발표하면서 AI 기술 개발과 활용을 핵심 과제로 설정하였으나, 그에 관한 원칙이나 규범, 구체적 전략은 없었다는 점에서 AI 거버넌스에 대한 논의로는 부족하였다고 보인다. 그러나 OECD의 AI 원칙(2019)을 반영하여 '인공지능 국가전략'을 발표하고(2019년 12월), 이후 2020년에 OECD, EU 등 국제기구의 권고에 따라 'AI 윤리기준'을 마련하면서 보다 본격적으로 AI 거버넌스 형성의 추진이 이루어졌다고 보인다.[21] 이에 따라 마련된 한국의 AI 윤리기준의 3대 기본원칙은 인공지능의 개발 및 활용 과정에서의 1) 인간의 존엄성 원칙, 2) 사회의 공공선 원칙, 3) 기술의 합목적성 원칙이며, 핵심 요건은 인권 보장, 프라이버시 보호, 다양성 존중, 침해 금지, 공공성, 연대성, 데이터 관리, 책임성, 안전성, 투명성 등이다.

이처럼 한국은 국제사회의 규범을 따르는 형태를 취하면서도 법제 측면에서는 AI에 의한 변화에 적극적으로 대처하려는 태도를 보였다. AI로 인해 새로운 처분 개념이 가능해지면서 2023년 행정기본법 제정을 통해 자동적 처분의 입법 기준을 마련하였고,[22] 개정된 개인정보보호법 제37조의2 제1항에서 자동화된 결정이 권리의무에 중대한 영향을 미치는 경우 해당 결정을 거부할 수 있는 권리를 규정하였다. 이는 AI 기술로 인한 새로운 위험에 법률로 대처했다는 점에서 긍정적으로 평가할 수 있다. AI로 인한 공정성 저해를 방지하기 위해 공직선

21) 2020년 12월 23일 과기정통부 보도자료, '사람이 중심이 되는「인공지능(AI) 윤리기준」마련'.

22) 2023년 3월 23일 제정된 행정기본법 제20조에서는 '행정청은 법률로 정하는 바에 따라 완전히 자동화된 시스템(인공지능 기술을 적용한 시스템 포함)으로 처분을 할 수 있다'고 정함으로써 자동적 처분으로 인한 기본권 침해 소지가 없도록 법률에 근거를 두도록 하고 있다. 다만 이는 모든 처분이 아니라 요건충족적 행정행위인 경우(즉, 법령상 요건이 충족되면 행정청이 어떠한 행위를 해야 하는 행정처분)에만 그러하고, 사람의 판단이 개입되어야 하는 재량적 처분인 경우에는 그러하지 않다. 또한, 개인정보보호법 제37조의2 제1항에서는 자동화된 결정이 자신의 권리 또는 의무에 중대한 영향을 미치는 경우에는 해당 개인정보처리자에 대하여 해당 결정을 거부할 수 있는 권리를 가진다'라고 규정하고 있다.

거법 제82조의8에 '선거운동에 있어서 인공지능을 이용한 이른바 딥페이크 기술을 활용한 게시물을 금지'하는 조항을 새로 규정한 것 역시 적실성 있는 제도적 대처라 할 수 있다. 이러한 인공지능 콘텐츠의 이용자에게 혼선을 방지하고 신뢰성과 책임성을 강화하기 위해 인공지능 기술로 제작된 콘텐츠임을 표시하도록 하는 개별 법률들을 추진 중이기도 하다.[23]

한국이 AI에 관한 국제적 규범을 국내 법령에 적응시키는 과정을 거치면서, 국제사회의 주도적 행위자로서의 행보를 보이기 시작한 것은 2020년 이후라 할 수 있다. 한국은 당해 6월 GPAI[24]의 창립회원국으로서 AI에 관한 국제적 논의의 주체가 되었고, 2023년 영국에서 개최된 'AI 안전 서밋(AI Safety Summit)'에 차기개최국 자격으로 참여하면서 선도자로서의 가능성을 드러내기 시작했다. 또한 제1차 안전성 정상회의의 후속회의인 'AI 서울 정상회의(AI Seoul Summit)'를 2024년 5월 영국과 공동개최하였는 바, 당해 회의에서는 AI 거버넌스에 대한 논의와 함께 안전하고 보안성과 신뢰성을 갖춘 AI의 설계·개발·배치·사용을 보장하고 AI로부터의 혜택을 극대화하는 한편 그로 인해 야기되는 위험들에 대응하기 위한 AI 거버넌스 체계들 간 상호운용성이 중요함을 확인하였다. 이 회의를 통해 한국은 '서울 정상회의 서울선언(Seoul Declaration for Safe, Innovative and Inclusive AI)'과 '프론티어 AI 안전 서약'을 발표하고,[25] '안전, 혁신, 포용'이라는 AI 거버넌스 3대 원칙을 국제사회에 제시하였다. 또한, 당해 9월에는 한국 국방부와 외교부의 공동 주관으로 '인공지능의 책임있는 군사적 이용에 관한 고위급회의(REAIM)'를 개최하였으며, 여기에는 90여개국 정부대표가 참석한 군사 AI 분야의 국제회의로서 공동주최국으로는 네덜란드, 싱가포르, 케냐, 영국이 참여하였다. 이 회의에서 군사분야 AI 규범 마련을 위한 청사진('REAIM Blueprint for Action')을 선언하고 채택하였다. 여기에 명시된 조치들은 AI의 위험성을 완화하면서 AI의 이점을 활용할 수 있도록 국제적 책임의 틀을 구축하는 것으로서, 당해 선언문은 '무기 시스템을 비롯한 모든 군사적 AI 능력의 개발과 전개 시 인간의 감독을 보장'하는 내용을 포함하고 있다.

---

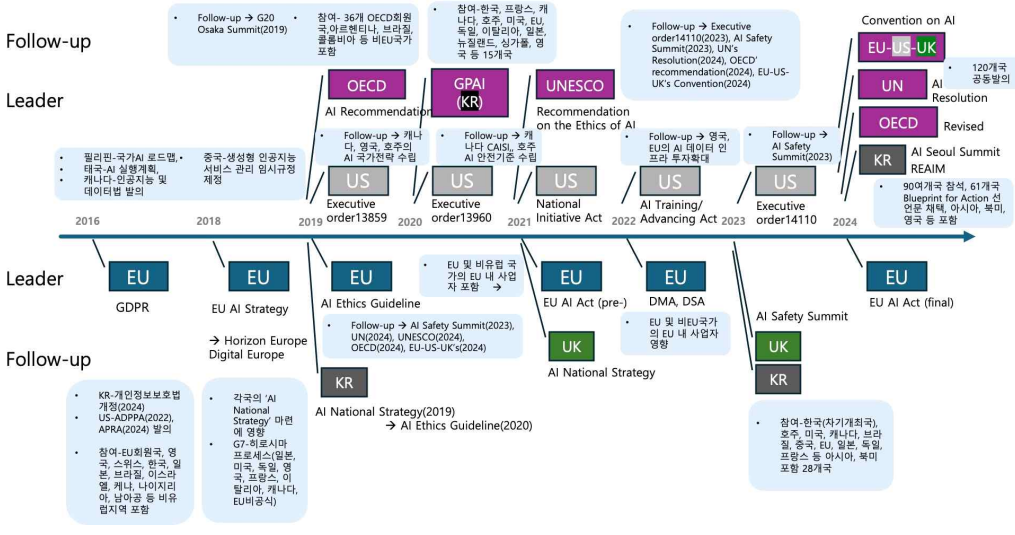
23) '콘텐츠산업 진흥법 개정안', '정보통신망법 일부개정법률안', '지능정보화 기본법 일부개정법률안', '민법 일부개정법률안' 등 딥페이크에 대한 규제 입법이 추진되었다.

24) OECD AI 권고안에 반영된 인권, 자유, 민주적 가치를 추구하는 책임성 있는 AI 개발, 사용 지침을 위한 국제협의체로서, 2019년 G7 정상회의에서 창설을 승인하였다. 현재는 본문에 언급된 15개 국가 외에 14개 회원국이 추가되었으며(네덜란드, 스페인, 폴란드, 브라질, 벨기에, 체코, 덴마크, 아일랜드, 이스라엘, 스웨덴, 아르헨티나, 세네갈, 세르비아, 튀르키에), OECD, UNESCO 등은 옵저버로 참여하고 있다.

25) '프론티어 AI 안전 서약'은 2024년 5월 'AI 서울 정상회의'에서 발표된 이니셔티브로서, Amazon, Google, Microsoft, OpenAI, Samsung, Naver 등 16개의 글로벌 AI 기업들이 서명했다. 이 서약은 위험 식별 및 관리, 위험 임계값 설정, 위험 완화 투자, 투명성과 책임성, 외부 이해관계자들의 참여 등을 핵심 내용으로 한다. 이는, 기업이 안전한 AI 개발을 약속한 첫 국제합의라 할 수 있다.

이러한 주요 행위자별 AI 규범의 확산의 흐름은 아래의 그림과 같이 정리해볼 수 있다.

〈그림 4〉 AI 규범의 확산 흐름



3. 확산 모형과 메커니즘을 통한 접근

AI에 관한 규범의 형성과 확산에 있어서 OECD, UNESCO, UN 등의 국제기구들은 Berry & Berry(2007)의 선도자-후발자 모형에서 상정하는 선도자(Leader)이자 우월적 지위에 있는 정당성 권위자(legitimacy authorities)에 해당한다. 이들은 AI 기술의 발달과 함께 그에 대한 규제적 필요성에 직면하면서 국제사회에서 이에 대한 규범 형성에 주도적 역할을 하였으며, 이들이 발표한 권고안 및 지침, 원칙, 선언들은 이후의 국제 협약이나 개별국가의 규범 형성에도 영향을 미쳐왔다. 기본적으로 이러한 과정은 확산 메커니즘 중 강압(coercion) 또는 하향적 확산으로 설명될 수 있다.

그러나 보다 심층적으로 각 규범들에서의 AI 원칙의 흐름을 통해 분석한 결과, 2019년 OECD의 AI에 관한 권고안에 나타난 주요 AI 원칙들(인간중심 가치, 공정성, 투명성, 안전성 등)은 UNESCO의 인공지능 윤리권고(2021)와 UN총회의 결의안(United Nations General Assembly AI Resolution, 2024)에서 인권과 민주주의 가치, 법치에 대한 강조로 나타났으며, 한국의 '인공지능 국가전략'(2019) 및 '인공지능 윤리기준'(2020)에서도 핵심적 가치로 제시되었다. OECD(2019)와 UNESCO(2021)의 권고안에 나타난 안전과 신뢰, 인권에 대한 강조는 한국 뿐만 아니라 이전까지 AI 기술의 개발과 혁신의 촉진에 중점을 두어 왔던 미국의 '안

전하고 보안이 보장되며 신뢰할 수 있는 인공지능 개발과 사용에 관한 행정명령'(2023)의 발표로 나타났다. 특히 안전성 가치는 2023년 영국의 'AI 안전성 정상회의'와 2024년 한국의 'AI 서울 정상회의' 개최에도 직접적 관련성을 갖는다. 이처럼 AI 규범은 당해 규범에서 강조된 가치를 통해 개별국가들의 규범의 내용에도 영향을 미친다고 할 수 있다. 이는 최근의 확산 연구에서 제시한 바와 같이, 확산의 대상이 단순한 정책 뿐 아니라 그 기저에 있는 가치를 포함함을 의미하며, 사회구성주의에서 정책의 도입을 정상국가의 정책 모델을 규범을 제시함에 따른 정당성 확보의 수단으로 이해하는 것과 맥락을 같이 한다. AI의 발달에 따라 개인정보보호와 정보의 오류나 환각 등 그에 대한 부작용을 규제하여야 할 필요성이 부각되면서 이에 대한 국제적 규제에 대한 규범적 정당성이 제기되었고, 이에 따라 국제기구에서 제정하는 보편적 규범들이 개별 국가들의 규범 형성에 영향을 미치면서 정책이 확산된 것으로 설명할 수 있다.

EU의 GDPR(유럽 일반 개인정보보호법)이나 DSA(디지털 서비스법) 및 DMA(디지털 시장법) 역시 회원국 뿐만 아니라 비회원국에도 책임성과 투명성 의무를 강하게 지우면서 상대국가들의 정책에 영향을 미친 것으로 나타난다. 당해 법들의 제정 이후 당해 조항을 고려하여 한국, 중국, 일본에서는 개인정보보호법을 제·개정하였고, 아프리카 국가들도 개인정보보호 및 포괄적 데이터 보호법을 제정하였으며, 미국의 연방 차원의 개인정보보호법(ADPPA, APRA)의 발의에도 영향을 미쳤다고 볼 수 있다. EU의 개별법 제정에 영향을 받아 회원국들이 관련 법을 제개정한 것은 EU 내 국가들간 경제적, 정치적 이익 추구와 손해를 피하기 위한 경쟁 기제의 작용으로 이해된다. 이는 국제기구에서 국가로의 확산 뿐 아니라 국가연합에서 개별 국가로의 확산을 확인시키는 것이기도 하다.[26] 나아가 EU의 AI 법은 가치기반 원칙들과 강한 규제조항 및 체계적으로 구분된 AI 개념들을 포함하고 있다는 점에서 개별 국가들의 AI 법 제정에 학습 기제로 작용할 가능성이 크다.

미국의 경우 국제사회에서 가장 선도적 지위에 있는 개별 국가라 할 수 있으며, AI 분야에 있어서도 주도권을 유지하고자 행정명령들을 발표하였다. 제13859호 행정명령의 경우 AI의 R&D 투자의 확대와 인력 양성을 강조하였으며, 이는 캐나다, 영국, 호주 등의 AI 국가전략 수립 및 AI에 대한 인프라 확대와 공통 기술표준 연구 및 인력교육을 강화하는 내용에 반영되었다. 이후 제13960호 행정명령에서는 신뢰할 수 있는 AI를 강조하면서 연방기관들에

26) 다만, 유럽연합(EU)의 지위를 개별국가보다 상위의 행위자로 볼 것인지는 유럽 내 행위자와 글로벌 관점에서 다르다고 할 것이다. 유럽연합 내 개별 국가들에게는 상위 기구이지만, 글로벌 관점에서 EU는 OECD 내에서도 개별 국가 회원들과 동등한 지위를 가지는 행위자가 된다. 본문에서 EU 내 국가들이 EU가 제정한 규범의 영향을 받아 개별국가의 법규범을 제정한 경우에는 수직적 확산이 되고, 미국, 한국과 같은 개별국가들에 영향을 미친 경우는 수평적 확산이 될 것이다.

게 AI 활용에 대한 투명성과 프라이버시 보호 등을 위한 기구 마련과 AI Use Inventory의 구축을 요구하였다. 이에 따라 캐나다는 2024년 CAISI(Canadian Artificial Intelligence Safety Institute)를 설립하고 AI 신뢰성 평가 네트워크를 구축하였으며, 호주 역시 자발적 AI 안전기준(Voluntary AI Safety Standard)을 도입하였다. 법적 구속력을 가지는 인공지능 훈련법 및 인공지능진흥법의 경우에도 영국과 EU의 AI 데이터 인프라 투자의 확대에 영향을 미쳤다. 이처럼 행정명령이나 법률은 당해 국가의 정책지향성과 계획의 안정성을 담보하고 있으며, 더구나 그것의 주체가 선도적 지위의 국가인 경우 그와 관련을 맺고 있는 동위 또는 후발국들은 이를 모방 및 학습한 후속조치들을 마련하게 된다.

일반적으로 정책 확산은 유사한 정책적 환경과 수요를 가진 국가들 사이에서 보다 빨리 나타나며, 동일한 권역에 속하거나 인접해 있는 국가간에는 관심있는 정책이나 규범들의 확산이 보다 쉽게 발견된다고 하여 이를 군집성 또는 인접 효과(neighboring effect)로 설명한다. 그러나, 본 연구의 분석에서 AI 규범의 형성은 비단 유럽이나 북미, 아시아 등과 같이 일정한 권역에 집중되어 나타나기보다는 글로벌한 범위의 확산 형태로 나타났다. GPAI의 15개 창립국은 한국, 일본, 싱가폴 등 아시아 뿐 아니라 영국, EU 등 유럽과 캐나다, 미국 등 북미를 포함하고 있으며, UNESCO의 권고안을 채택한 193개 회원국에도 다양한 권역의 국가들이 포함되어 있다. UN 총회의 AI에 관한 결의안에도 120개국 이상이 서명하였으며, 한국이 개최한 'AI Seoul Summit'에는 아시아, 북미, 영국 등 90여개국이 참석하고 61개국이 'Blueprint for Action'에 서명하였다. 이와 같이 AI에 관한 규범의 확산에 있어서는 이웃한 국가간의 군집성이 나타나지 않는다.

본 연구는 이를 규범적 정당성 접근으로 해석하였다. 분석결과, AI라는 혁신적 기술의 개발과 촉진, 산업의 육성 측면의 전략과 이니셔티브가 국가간에 확산되었던 것과 같이, AI에 대한 규제적 규범 역시 일정 지역에 국한되지 않고 전세계적으로 파급되었다. 생성형 AI 기술 자체에 내재하는 편향성의 위험과 차별의 문제, 프라이버시 침해, 알고리즘 투명성, 할루시네이션 오류 등이 실제로 사람들의 일상과 공공 분야에서 위해를 가져오면서 이를 규제해야 할 필요성은 규범적 정당성을 얻게 되었으며, 이에 따른 모방적 도입이 나타났다고 볼 수 있다. 특히, AI 기술의 혁신성과 위험성은 전세계가 서로 연결된 채 빠른 파급효과를 가지기 때문에 그에 대한 불안은 클 수밖에 없다. AI 시스템의 설계와 개발, 활용 등 전 주기에 있어서 안전성, 책임성, 투명성의 담보와, 궁극적으로 인권 및 인간의 존엄성과 인간중심의 기술의 향유로 가야한다는 방향성을 재강조하고 있는 국제기구들의 AI에 관한 윤리적 지침 등은 그것을 따를만한 규범적 정당성을 가진다. 그러한 점에서, 서로 다른 권역에 존재하는 국가들이라 할지라도 동일한 원칙과 성격을 가지는 AI 규범을 형성함으로써 국제사회에서 AI 규

제정책의 확산이 나타나게 되었다고 볼 수 있다. 이는 정책확산 연구에서 설명하는 다양성 가운데 공통성이 나타나는 현상을 잘 반영하고 있다.

V. 결론

본 연구는 국제사회의 주요 행위자들이 제정한 AI에 관한 규범들이 어떠한 흐름으로 확산되었는지를 국가간 정책확산 모형의 이론적 틀과 국제기구의 주요 원칙들을 기준으로 분석하였다. AI에 관한 국제규범들과 그에 담긴 원칙들은 제정 주체들의 정책적 지향을 담고 있기 때문이다. 본 연구의 분석에서 행위자들은 국제기구나 다른 국가들과 상호의존성을 가지고 서로 영향을 받아 규범을 마련하면서도, AI 원칙으로 대표될 수 있는 각 분야별로 선도적 국가로서의 역할을 하고 있는 것으로 나타났다. 즉, EU가 유럽 일반 개인정보보호법과 AI 법 등을 통해 AI의 안전성과 책임성을 담보하는 데 있어서 주도적 역할을 해왔다면, 미국은 AI 기술 개발과 활용 및 산업 촉진 측면에서 국제사회의 리더적 지위를 누려왔다. 다만, 국제사회에서 AI의 위험성 인식과 안전성에 대한 강조의 흐름이 나타나면서 바이든 행정부의 행정명령(2023)에서부터는 보다 AI의 촉진과 규제의 균형을 찾고 있는 것으로 보인다. 이는 미국이 AI에 관한 협약(2024)을 포함한 이후의 규범에서 안정성 및 인권과 민주주의 가치를 강조하는 경로로 변화한 중요한 분기점에 해당하기도 한다. 이에 비해 한국은 국제기구와 미국 및 유럽의 영향 하에서 후발국가로서 이들의 규제 정책을 따르던 흐름에서 AI Seoul Summit 및 REAIM의 개최에 이르기까지 점차 국제사회에서 AI 규범 논의에서 두각을 드러내고 있다. 또한 투명성 및 책임성을 강조한 EU나 산업 촉진과 안정성에 중점을 둔 미국과 달리 한국은 인공지능 국가전략 및 윤리원칙에서 형평성을 고려하는 특징을 보인다. 이는 공공부문에서의 AI 전환과 함께 그로 인한 불평등과 소외가 나타나지 않도록 하려는 한국 정부의 정책적 지향이 담긴 것이라고 판단된다. 한국이 2024년 서울 정상회의(Seoul AI Summit)와 REAIM 등을 개최하면서 규범 논의에서 점차 두각을 나타내고 있기는 하지만, 후발국들의 후속조치를 밝히기는 어렵다는 점에서 아직까지 선도자로서 자리매김하였다고 판단하기는 이르다.

본 연구는 국제기구와 주요 국가들의 AI 규범의 확산되는 과정을 2016년부터 2024년까지의 시간의 흐름 속에서 살펴보았다. 특히 규범들에서 강조된 AI 원칙들을 살펴봄으로써 주체별 정책적 지향을 파악하면서, AI 규범의 선도적 행위자와 후발자들의 움직임을 정책확산에 관한 모형과 메커니즘으로 설명하였다. 이러한 분석은 법, 윤리원칙, 조직체계, 국가전

략 및 기술 인프라 등을 포함한 AI 거버넌스 연구에 있어서도 주요 주체들에 대한 고려와 내용의 체계 및 구체화에 기여할 수 있을 것이다.

그럼에도 불구하고 본 연구는 2024년 12월까지를 분석기간으로 하고 있어 이후의 주요 규범들을 포함하지 못하였고, 법적 구속력이 상이한 규범들을 다룸에 따라 선행연구들에서 나타나는 선택편의 문제를 극복하지 못한 한계를 가진다. 질적 코딩에 있어서 신뢰성 도구를 사용하지 못한 데 따른 주관성 개입 가능성도 내재한다. 또한 본 연구의 분석대상에 미국의 연방 개인정보보호법안(ADPPA) 및 프라이버시보호법안(APRA)이 의회를 통과되지 못함에 따라 분석대상으로 삼지 못한 점은 아쉬우나, 분석 이후 제정된 한국의 AI 기본법에 관해서는 후속 연구를 통해 심층적으로 다룰 예정이다.

AI의 발전은 예측불허의 속도로 시공간을 초월하는 파급효과를 가지고 있고, 그렇기에 이에 대한 통제력이 확보되지 못하면 인간 사회는 심각한 위험에 노출될 수밖에 없다. 기술은 그 스스로 자제하거나 방향을 설정할 수 없다는 점에서 그 통제와 조정은 인간의 몫으로 남아있다. AI 규범은 인간의 존엄성과 인권 보호를 중심 가치로 하면서, 안전하고, 투명하며, 차별적이지 않고, 책임있는 AI에 대한 운용 및 규제 프레임을 형성한다. 따라서 바람직한 AI 규범의 선도자라면 AI에 대한 기술적 이해와 국제적 시야를 가지고 그 원칙과 규범을 제시할 수 있어야 할 것이다. AI에 대한 규제 및 규범은 그 가치를 실현하기 위한 수단으로서의 의미를 가지며, 이는 정책학의 본질과도 통한다. "AI가 중요하기에 오히려 더 규제해야 한다"는 Sundar Pichai의 말[27]은 기술 혁신이 중요해진 현대 사회에서 공공 행정과 정책의 역할이 더욱 중요하다는 점을 역설적으로 보여준다.

참고문헌

고인석. (2021). 인공지능(AI)에 대한 법인격 부여와 법적 책임에 관한 연구-FinTech 산업에서의 인공지능(AI)의 활용과 법적 책임을 중심으로-,「지급결제학회지」, 제13권 제2호.

고학수·박도현·이나래. (2020). 인공지능 윤리규범과 규제 거버넌스의 현황과 과제,「경제규제와 법」, 제13권 제1호.

국가안보전략연구원. (2023). 바이든 행정부의 첫 인공지능(AI) 행정명령과 시사점, 이슈브리프 제480호.

김대인. (2020). 행정기본법과 행정절차법의 관계에 대한 고찰,「법제연구」제59호.

27) Google과 Alphabet의 CEO.

김경민·이용준·강장묵. (2024). 인공지능 발전에 따른 책임과 법적 규제에 대한 연구, 「한국산학기술학회논문지」 제25권 제7호.

김기영. (2016). 로봇규제를 위한 법적 개념과 문제, 「경성법학」 제25권.

김법연. (2024). 생성형 AI의 법적 문제와 규제 논의 동향, 「정보화정책」 제31권 제3호.

김정화·임동민·차호동. (2023). 생성형 인공지능(Generative AI) 기술의 규제 방향에 대한 입법론적 고찰-GhatGPT 등 인공지능 시스템 생성물에 대한 표시·고지의무를 중심으로-, 「형사법의 신동향」 통권 제80호.

남궁근. (1994). 정책혁신으로서의 행정정보공개조례 채택. 「한국정치학회보」, 제28권 제1호.

라기원. (2024a). 유럽연합 인공지능법(EU AI Act)의 특성과 쟁점-우리나라 인공지능 입법에 대한 시사점, 한국법제연구원.

라기원. (2024b). 인공지능 거버넌스-EU, 영국, 미국의 거버넌스 분석, 한국법제연구원.

박기갑. (2022). 국제적 권고·지침 분석에 바탕을 둔 "인공지능(AI) 윤리" 관련 국제조약안의 모색, 「국제법학회논총」, 제67권 제4호.

박성준·김상겸. (2024). AI 창작물의 저작권 보호는 필요한가, 「법학연구」, 제34권 제2호.

박혜란. (2024). 인공지능 기술의 기본권 침해 대응 방안 연구, 「성균관법학」, 제36권 제3호.

설민수. (2017). 머신러닝 인공지능과 인간전문직의 협업의 의미와 법적 쟁점: 의사의 의료과실 책임을 사례로, 「저스티스」 통권 제163호.

성승제. (2023). 인공지능 규제방향에 대한 평가와 전망, 한국법제연구원.

성욱준·최한별. (2023). 인공지능 시대 도래에 따른 AI 입법수요 및 과제 연구, 국회입법조사처.

양천수. (2024). 유럽연합 인공지능법과 인공지능 규제, 국제법리뷰, 제24-1호.

유승익. (2023). 인공지능이 인권과 민주주의에 미치는 영향과 인공지능법안의 쟁점, 「민주법학」 제82호.

유재흥·추형석·강송희. (2021). 유럽(EU)의 인공지능 윤리 정책 현황과 시사점: 원칙에서 실천으로, 소프트웨어정책연구소.

이대웅. (2014). 한국 지방정부의 정책확산에 관한 연구-반부패 신고포상금제도를 중심으로-. 성균관대학교 석사학위논문.

이대웅·권기헌. (2015). 한국 지방정부의 정책확산에 관한 연구-반부패 신고포상금제도를 중심으로-. 「한국정책학회보」, 제24권 제3호.

이병준. (2021). 유럽연합 디지털 서비스법을 통한 플랫폼 규제-디지털 서비스법 초안의 주요내용과 입법방향을 중심으로-, 「소비자법연구」 제7권 제2호.

이석환. (2013). 한국 지방자치단체 출산장려정책의 수평적·수직적 확산. 「한국행정학보」, 제47권 제3호.

이석환. (2022). 정책확산의 기제: 기초자치단체 출산장려정책을 대상으로, 「한국행정학보」, 제56권 제2호.

이성용. (2013). 정보의 자기결정권에 따른 경찰상 정보보호의 입법원리. 「치안정책연구」, 제27권

제2호.
이승종. (2004). 지방 차원의 정책혁신 확산과 시간-지방행정정보공개조례의 사례연구.「한국지방자치학회보」, 제16권 제1호.
이우철. (2024). 인공지능시대 알고리즘 의사결정에 의한 차별과 인권보호방안에 관한 연구-유럽의 법적 규제에 관한 논의를 중심으로-,「세계헌법연구」, 제30권 제2호.
장한일. (2024). AI 딥페이크 제작물이 선거에 끼치는 영향. 2024 봄호 KIPA 규제동향. 한국행정연구원.
정명은. (2012). 지방정부의 경쟁적 세계화: 수직적 확산과 수평적 확산.「한국행정학보」, 제46권 제3호.
장석준. (2013). 정책 유형별 확산 메커니즘의 차별적 영향력에 관한 실증연구,「한국정책학회보」, 제22권 제4호.
정세은. (2023). 인공지능의 개인정보 침해에 대한 책임 귀속 연구, 고려대학교 박사학위논문.
정완. (2024). 인공지능의 입법적 쟁점에 관한 고찰,「경희법학」, 제58권 제4호.
조일형·권기헌·서인석. (2014). 정책학습이 정책확산에 미치는 영향에 관한 연구-출산장려금 정책을 중심으로-,「한국정책학회보」, 제23권 제3호.
최상한. (2010). 지방정부 주민참여예산제도의 확산과 영향요인.「한국행정학보」, 제44권 제3호.
한국지능정보사회진흥원. (2021). 영국 AI 국가전략의 주요 내용 및 시사점, AI Report.
한세억·강지현. (2024). 생성형 인공지능 시대에서 공정성과 법제적 쟁점,「공공정책연구」, 제41권 제1호.
홍남희. (2018). 디지털 플랫폼에 의한 '사적 검열(private censorship), 미디어와 인격권, 통권 제6호.
Agostinis, G. (2019). Regional intergovernmental organizations as catalysts for transnational policy diffusion: the case of UNASUR health. Journal of Common Market Studies, 57(5), 1111-1129.
Allen, M. D., Pettus, C., & Haider-Markel, D. P. (2004). Making the national local: Specifying the conditions for national government influence on state policymaking. State Politics and Policy Quarterly, 4(3), 318-344.
Berry, F. S., & Berry, W. D. (1990). State lottery adoptions as policy innovations: An event history analysis. American political science review, 84(2), 395-415.
Berry, F. S., & Berry, W. D. (2007). Innovation and Diffusion Models in Policy Research, Theories of the policy process, 223-260.
Boehmke, F. J., & Witmer, R. (2004). Disentangling diffusion: The effects of social learning and economic competition on state policy innovation and expansion. Political Research Quarterly, 57(1), 39-51.
Braun, D. & Gilardi, F. (2006). Taking 'Galton's problem' seriously: Towards a theory of policy diffusion. Journal of Theoretical Politics, 18(3), 298-322.

Brune, N., Garrett, G., & Kogut, B. (2004). The International Monetary Fund and the Global Spread of Privatization. IMF Staff Papers, 51(2), 195-219.

Chawki, M. (2024). An effective cloud computing model enhancing privacy in cloud computing. Information Security Journal: A Global Perspective, 33(6), 635–658.

Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2022). Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover. The International Journal of Human Resource Management, 34(14), 2732–2764.

Danaeefard, H., & Mahdizadeh, F. (2022). Public policy diffusion: A scoping review. Public Organization Review, 22(2), 455-477.

de Oliveira, O. P., Romano, G. C., Volden, C., & Karch, A. (2023). Policy diffusion and innovation. In Theories of the policy process (pp. 230-261). Routledge.

DiMaggio, P. J. & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. American Sociological Review, 48(2), 147-160.

Dobbin, Frank, Beth Simmons, & Geoffrey Garrett. (2007). The Global Diffusion of Public Policies: Social Construction, Coercion, Competition, or Learning?, The Annual Review of Sociology, 33: 449-472.

Egan, M. (2022). Privacy boundaries in digital space: an exercise in responsibilisation. Information & Communications Technology Law, 31(3), 301-318.

Finnemore, M. (1996). National interests in international society. Cornell University Press.

Gilardi, F., & Wasserfallen, F. (2019). The politics of policy diffusion. European Journal of Political Research, 58(4), 1245-1256.

John, A. M., & Panachakel, J. T. (2024, October). Navigating AI Policy Landscapes: Insights into Human Rights Considerations Across IEEE Regions. In 2024 IEEE 12th Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1-6). IEEE.

Karch, A. (2006). National intervention and the diffusion of policy innovations. American Politics Research, 34(4), 403-426.

Karwatzki, S., Dytynko, O., Trenz, M., & Veit, D. (2017). Beyond the Personalization–Privacy Paradox: Privacy Valuation, Transparency Features, and Service Personalization. Journal of Management Information Systems, 34(2), 369–400.

Kashefi, P., Kashefi, Y., & Ghafouri Mirsaraei, A. (2024). Shaping the future of AI: balancing innovation and ethics in global regulation. Uniform Law Review, 29(3), 524-548.

Keith, A. J. (2024). Governance of artificial intelligence in Southeast Asia. Global Policy, 15(5), 937-954.

Long, Y., Luo, X., Zhu, Y., Lee, K. P., & Wang, S. J. (2023). Data Transparency Design in Internet of Things: A Systematic Review. International Journal of Human–Computer Interaction, 40(18), 5003–5025.

Lou, S., Sun, Z., & Zhang, Y. (2023). To join the top and the bottom: the role of provincial governments in China's top-down policy diffusion. Journal of Chinese Governance, 8(2), 161-179.

Manners, Ian (2002). Normative Power Europe: A Contradiction in Terms? Journal of Common Market Studies, 40(2), 235–258.

Marsh, D., & Sharman, J. C. (2009). Policy diffusion and policy transfer. Policy studies, 30(3), 269-288.

Meseguer, C., & Gilardi, F. (2009). What is new in the study of policy diffusion?. Review of International Political Economy, 16(3), 527-543.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507.

Moon, M. J. (2023). Searching for inclusive artificial intelligence for social good: Participatory governance and policy recommendations for making AI more inclusive and benign for society. Public Administration Review, 83(6), 1496-1505.

Mooney, C. Z., & Lee, M. H. (1995). Legislative Morality in the American States: The Case of Pre-Roe Abortion Regulation Reform. American Journal of Political Science, 39(3), 599-627.

Pavlidis, G. (2024). Unlocking the black box: analysing the EU artificial intelligence act's framework for explainability in AI. Law, Innovation and Technology, 16(1), 293–308.

Phillips, J. (2024). How does the consent model in Australia's Privacy Act 1988 (Cth) undermine our human right to privacy? Australian Journal of Human Rights, 1–20.

Radu, R. (2021). Steering the governance of artificial intelligence: national strategies in perspective. Policy and society, 40(2), 178-193.

Ring, J. J. (2014). The diffusion of norms in the international system. The University of Iowa.

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (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 (pp. 153-158).

Schimmelfennig, F., & Sedelmeier, U. (2004). Governance by conditionality: EU rule transfer to the candidate countries of Central and Eastern Europe. Journal of European public policy, 11(4), 661-679.

Shipan, C. R., & Volden, C. (2008). The mechanisms of policy diffusion. American Journal of Political Science, 52(4), 840-857.

Shukla, S. (2024). Principles governing ethical development and deployment of AI. Journal of Artificial Intelligence Ethics, 7(1), 45-60.

Siddiqui, M. A. (2024). A Comprehensive Review of AI: Ethical Frameworks, Challenges, and Development. A Journal of Management Sciences, 14, 68-75.

Suárez, S. L. (2012). Reciprocal policy diffusion: the regulation of executive compensation in the UK and the US. Journal of Public Affairs, 12(4), 303-314.

Souza, C. A., De Oliveira, C. C., Perrone, C., & Carneiro, G. (2021). From privacy to data protection: the road ahead for the Inter-American System of human rights. In The Right to Privacy Revisited (pp. 151-180). Routledge.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: key problems and solutions. AI & Society, 36(4), 945-961.

Van Ooijen, I., & Vrabec, H. U. (2019). Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. Journal of consumer policy, 42, 91-107.

Vanberg, A. D. (2021). Informational privacy post GDPR–end of the road or the start of a long journey?. In The Right to Privacy Revisited (pp. 56-82). Routledge.

Volden, Craig. 2006. States as Policy Laboratories: Emulating Success in the Children's Health Insurance Program. American Journal of Political Science 50: 294-312.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76-99.

Walker, J. L. (1969). The Diffusion of Innovation among the American States. American Political Science Review, 63(3), 880-899.

Walter, Y. (2024). Managing the race to the moon: Global policy and governance in artificial intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences. Discover Artificial Intelligence, 4(1), 14.

Wang, C., Li, X., Ma, W., & Wang, X. (2020). Diffusion models over the life cycle of an innovation: A bottom-up and top-down synthesis approach. Public Administration and Development, 40(2), 105-118.

Welch, S., & Thompson, K. (1980). The impact of federal incentives on state policy innovation. American journal of political science, 715-729.

Wendt, A. (1999). Social theory of international politics (Vol. 67). Cambridge university press.

Weyland, K. (2005). Theories of policy diffusion lessons from Latin American pension reform. World politics, 57(2), 262-295.

Wirtz, B. W., & Müller, W. M. (2019). An integrated artificial intelligence framework for public

management. Public Management Review, 21(7), 1076-1100.

Xu, J., Lee, T., & Goggin, G. (2024). AI governance in Asia: policies, praxis and approaches. Communication Research and Practice, 10(3), 275-287.

Zhang, Y., & Zhu, X. (2019). Multiple mechanisms of policy diffusion in China. Public Management Review, 21(4), 495-514.

Ziller, J. (2024). The Council of Europe Framework Convention on Artificial Intelligence vs. The EU Regulation: Two Quite Different Legal Instruments.

ABSTRACT

The Transnational Diffusion of AI Norms: A Focus on Major International Organizations, the EU, the United States, and South Korea

Junhee Park

This study examines the patterns of international diffusion of artificial intelligence (AI) norms by analyzing the principles embedded within them. The analysis is guided by six core ethical principles commonly articulated by global institutions such as the OECD and the United Nations, focusing on 24 AI-related normative documents published between 2016 and 2024 by key international organizations and major actors-namely the European Union (EU), the United States (US), and South Korea. The findings reveal that the emphasis placed on specific AI principles has evolved over time, reflecting shifts in leadership roles and issue-area specialization among the respective actors. The analysis shows that the EU has taken a leading role in ensuring AI safety and accountability, while the US has maintained its position as a global leader in AI development, deployment, and market-driven innovation. South Korea, initially a follower of global norms, is now emerging as an active contributor by hosting and leading major international discussions of AI governance. These findings offer insights into the evolving landscape of global AI norm diffusion and provide a foundation for future research on international AI governance.

【Keywords: AI norms, AI principles, transnational diffusion】