안전하고 신뢰할 수 있는 인공지능을 위한 주요국의 정책 및 국제협력 동향과 시사점

서민경, 임진국*

정보통신기획평가원(IITP)

mkseo@iitp.kr, *limjk@iitp.kr

Major Countries' Policies and International Cooperation Trends for Safe and Trustworthy AI: Implications and Insights

Seo Minkyung, Lim Jinkuk*

Institute of Information & Communications Technology Planning & Evaluation

요 약

인공지능 기술의 급격한 발전은 새로운 사회문제를 야기했으며, 주요국은 '신뢰할 수 있고 안전한 인공지능' 보장을 위하여 규제·제도·정책 등을 발표하고 있다. 또한, 인공지능 통제의 필요성에 대하여 국제적 공감대가 형성되어 주요국을 중심으로 국제협력 움직임과 합의가 있었다. 본 논문에서는 미국, EU, 중국, 영국, 한국 등 주요국이 '인공지능 안전'을 위하여 추진 중인 정책과 국제협력의 동향을 분석했다. 각국은 처한 환경에 따라 근본적인 접근 방식에 차이가 있지만, 미국, 영국, 한국은 법적 구속력 없는 자율규제를 채택하며, EU와 중국은 법을 제정하여 패널티를 부과하고 있다.

Ⅰ. 서 론

대표적인 생성형 인공지능인 오픈AI의 챗GPT가 2022년 11월 공개되어 사회에 큰 파문을 불러왔고, 단 2개월 만에 월간사용자수(MAU) 1억 명을 돌파하는 기록을 세웠다. 그리고 불과 17개월 지난 2024년 5월에는 구글의 '제미나이(Gemini)', 메타의 '이미지바인드(ImageBInd)', 네이버의 '하이퍼클로바X' 등 빅테크 기업의 멀티모달(Multi Modal) 경쟁 중이다. 인공지능 기술은 인류 역사상 전례 없는 속도로 빠르게 진보하고 있으며, 향후 몇 년 이내에는 인간처럼 다중지능과 자율행동이 가능한 AGI(Artificial General Intelligence)로의 진화가 예고되어 있다. 인공지능은 사회, 경제, 문화 등 전방위적으로 큰 영향을 미쳤고, 이 과정에서 생산성, 효율성 및 편의성 향상과 같은 긍정적인 영향도 있었다. 하지만 짧은 시간 동안 인공지능의 급격한 발달은 부작용을 불러왔고, 인공지능의 무분별한 사용, 알고리즘에 의한 차별과 배제의 확대·재생산, 개인정보 침해, 신구 서비스 간의 갈등, 인공지능의 저작권 인정 등과 같은 새로운 문제를 야기했다.[1]

이에 따라, 주요국은 인공지능 혁신을 지원하는 한편, 본격적으로 '안전성과 신뢰성 확보'를 위한 규제 등 정책을 발표 및 추진하기 시작했다. 2023년은 인공지능 규제 정책의 원년으로 핵심 정책을 비롯하여 토대가 되는 정책들이 발표되었으며, 2024년에도 그 기조가 이어지고 있다. 또한, 국제적 공감대가 형성되어, 표준 마련과 질서 정립 등을 위한 협력을 추진 중이다. 본 논문에서는 '인공지능 안전'과 관련된 미국, EU, 중국, 영국, 한국 등 주요국의 정책과 국제협력 동향, 각국에서 추구하는 정책의 방향성을 분석했다. 또한, 우리나라 정책과의 비교를 통해 시사점을 도출했다.

Ⅱ. 본론

미국은 국가 인공지능 정책의 시작점이라고 할 수 있는 「국가 인공지능 연구개발 전략계획('16.10)」에서 연구개발 4순위에 '인공지능 시스템의 안전과 보안 확보'를 제시하여 초기부터 투자해 왔지만, 본격적인 것은 「인공지능 권리장전 청사진('22.10)」부터라고 할 수 있다. 권리장전을 기반으로 하여 이후로, 백악관의 「기업의 자발적인 인공지능 안전서약('23.7, '23.9)」을 발표했다.[2] 마지막으로, 미국 정부는 역대 가장 포괄적인 인공지능 관련 행정명령인 '안전, 보안, 신뢰 기반의 인공지능 행정명령('23.10)'을 발효했다.[3] 인공지능 신뢰성을 위한 표준 마련뿐만 아니라, 보안 및 안보, 연구개발 혁신, 국제협력, 정보의 인공지능 활용 등 각 부문에 대하여 세분화 된 접근방식을 채택했으며, 관리예산국(OMB), 상무부(DOC) 등에 관련 정부 부처 및 기관에 이행을 위한 명령을 하달했다. 2024년 2월에는 본 행정명령 발효 이후 90일간의 주요 달성 현황에 대하여 발표했으며, 국가안보와 관련하여 개발자의 안전 테스트 결과 보고 의무 부과, 클라우드 기업의 정보 공유 의무 부과 등의 성과를 제시했다.

EU는 회원국의 「인공지능 협력('18.4)」 선언을 기반으로, 최초의 전략인 「유럽을 위한 인공지능 전략('18.4)」을 발표했다. 선언 직후에는 '인공지능 고위급 그룹(AI High-Level Expert Group, '18.6)'을 조직하여 '신뢰할 수 있는 인공지능 윤리 지침('19.4)'과 '신뢰할 수 있는 인공지능 평가 목록('20.7)'을 발표했다. 또한, 「인공지능 백서: 탁월성과 신뢰에 대한 유럽의 접근 방식('20.2)」으로 방향성을 제시했고, 이를 기반으로 최근까지 정책 기조의 기틀이 되는 「인공지능에 대한 유럽의 접근 방식: 인공지능에 대한 법적 프레임워크 및 인공지능 조정계획 개정('21.4)」을 발표했다. 즉, EU는 초기부터 인공지능의 신뢰성 확보를 추구해 왔으며, 이러한 일관적인 정책 방향성이 세계 최초의 인공지능 포괄적 규제 법안인 「인공지능법('23.12)」 제정으로 이어졌다고 할 수 있다. 「인공지능법」은 위험기반 접근방식을 채택하고 있으며, 인공지능 시스템 개발자, 배포자 및 사용자에게 명확한 요구사항과 의무를 제공하여 법 집행의 일관성을 유지하여 투자와 혁신, 단일 시장 개발을 촉진하는 것을 목적으로 제정되었다.[4]

중국은 인공지능을 국가 미래를 선도할 전략기술로 주목하여, 국무원, 국가발전계획위원회 등 최상위 기관에서 전략을 발표했다. 국가발전계획위의 「인터넷 플러스 인공지능 3개년 실천방안('16.5)」, 국무원의 「차세대 인공지능 발전규획('17.7)」, 「14차 5개년 규획('21~'25)('21.3)」, 「인공지능 대형 모델 산업 혁신발전 3개년 행동계획('23.11)」 등 이다. 인공지능

윤리 관련으로는 「차세대 인공지능 거버넌스 원칙(책임있는 AI 개발)('19.6)」, 「차세대 AI 윤리규범('21.9)」 등을 발표하여 인공지능에 대한 통제를 강화했다. 이후, 법령을 통해 규제를 강화했으며, 「생성형 인공지능 서비스 잠정관리방법('23.7)」이 8월부터 시행되었다.[5] 이를 통하여 보안평가, 데이터 훈련·라벨링, 개인정보보호, 운영상 규제 등 생성형 인공지능 서비스 제공·이용 전반에 대한 의무사항을 명확하게 제시했다. 시행된 법은 의견수렴초안에 비하여 '생성형 인공지능 제공자'에 대한 책임을 일부 삭제하여, 규제 수준을 낮추어 혁신을 강조했다고 할 수 있다. 또한, 최근 발표된 「인공지능+ 이니셔티브('24.3)」에서 인공지능의 산업 경쟁력 제고 방안과 동시에 '인공지능의 안전과 감독'에도 주목했다.[6]

영국은 인공지능 기술을 산업 및 경제 발전의 원동력으로 설정하여, 국가 경쟁력을 제고하기 위한 정책을 추진해 왔으며, 인공지능의 윤리 및 안전 관련하여서는 주로 공공 부문에 초점을 맞춰왔다. '공공 부문에서의 인공지능을 사용하기 위한 지침('19.6)', '인공지능 조달 지침('20.6)', '자동화된 의사결정을 위한 윤리, 투명성, 책임성 프레임워크('21.5)' 등이다. 최근에는 「국가 인공지능 전략('21.9)」 지원 등 글로벌 리더십 유지를 위한 관점에서 접근하고 있다. '인공지능 규제를 위한 친(親)혁신 접근법 구축('22.7)' 정책서 제출을 시작으로, 「인공지능 규제 백서: 인공지능 규제에 대한 친혁신 접근법('23.3)」 발표를 통하여 명확하고 친혁신적이며 유연한 접근 방식을 수립했다.[7] 인공지능 산업 육성을 위한 환경 조성에 있어서 '규제'의 역할에 주목하여, 책임 있는 혁신을 가속화 하기 위해 규제의 명확성과 일관성을 부여하기 위하여 프레임워크를 최대한 유연하게 설계한 것이 특징이다.

한국은 「신뢰할 수 있는 인공지능 실현 전략('21.5)」를 통해 실천 방안을 구체화하였다. 특히, 디지털 심화시대의 신질서 및 규범을 재정립하기 위한 「디지털 권리장전('23.9)」을 발표했으며, 이를 기반으로 글로벌 질서 정립을 위한 정책을 추진하고 있다. 2023년에는 「인공지능 일상화 및 고도화 계획('23.1)」, 「초거대AI 경쟁력 강화 방안('23.4)」, 「전국민 AI 일상화 실행계획('23.9)」을 연달아 발표했으며, 각 전략에서 인공지능의 신뢰성 및 윤리 확보를 위한 방안을 제시했다. 우리나라는 기본적으로 민간의 자율규제를 채택하여 「사전적정성 검토제(안)('24.1)」 제도 마련 중이며, 2024년 초에 고시 제정으로 추진 예정이다. 또한, 인공지능 혁신과 규제를 포괄하는 「인공지능 기본법」 제정을 추진 중이다.

한편, '인공지능 안전성과 신뢰성' 확보는 단일 국가가 홀로 달성할 수 없으며, 국제협력의 필요성에 대하여 공감대가 형성되었다. 이에 따라 의미있는 국제적 합의가 도출되었으며, 특히 EU는 양자 간의 협력에서도 강조하고 있다. 첫 번째로, 영국이 '인공지능 안전' 분야의 글로벌 리더로 발돋움하기 위하여 2023년 11월 '제1회 글로벌 인공지능 안전성 써밋(AI Safety Summit)'을 런던의 블레츨리에서 개최했다. 이를 계기로, 영국 외 28개국이 '블레츨리 선언(Bletchley Declaration)'에 합의했으며, 이에 따라 인공지능을 안전하고, 인간 중심적이며, 신뢰할 수 있고, 책임 있게 설계·개발·배치할 것이며, 이를 위해 글로벌 대화를 지속할 것을 결의했다.[8] 제2회 회의를 2024년 5월에 서울에서 개최하여 논의를 지속할 예정이다.

두 번째로, G7에서 '히로시마 인공지능 프로세스'를 통해 강령과 가이드라인을 발표했다. OECD의 인공지능 원칙을 기반으로 개발된 11개의 원칙을 제시했으며, 인공지능 개발자 및 관련자에게 안내를 제공하기 위한 자발적인 행동 지침 및 강령이다. EU도 본 강령을 공식적으로 채택했다.[9]

세 번째, EU는 글로벌 파트너의 디지털 정책을 EU와 일치시키고, 글로벌 표준 마련 등을 위한 국제협력으로 'EU-미국 무역기술위원회(TTC)', 일본, 한국, 싱가포르, 캐나다와의 '디지털 파트너십(Digital Partnership)'을 추진 중이다.[10]

각국은 처한 환경에 따라 규제에 대한 근본적인 접근 방식에서 차이가 있지만, 크게 두 가지 방향성으로 구분할 수 있다. 첫 번째, 인공지능 혁신을 지향하는 민간기업 자율규제를 채택하고 있는 미국, 영국, 한국의 경우이다. 가장 대표적인 국가인 미국은 기업의 '자발적인 서약'을 받아서 자율규제 형식을 취하고는 있지만, 「AI 행정명령」 발표를 통해 중요 테스트 정보를 의무적으로 보고하도록 규제를 발표했으며, 중국을 견제하기 위해 클라우드 기업에 관련 정보를 요구하고 접근을 차단하는 규제를 추가로 준비 중이다. 즉, 국가안보와 관련된 사항에 대해서는 혁신보다는 규제에 초점을 맞췄다. 영국의 경우, 친(親)혁신을 지향하는 「인공지능 규제 백서」의 실현을 위해 '인공지능 표준' 개발을 위해 투자하고 있으며, '인공지능 안전성 써밋'을 주도하는 등 글로벌 리더로서 발돋움 중이다.

두 번째, 관련 법령을 제정하여 규제를 추진하고 있는 EU와 중국의 경우이다. 양국의 규제 방향성과 접근 방식에는 차이가 있지만, 혁신을 촉진하기 위한 규제 환경의 일관성 제공이라는 공통된 목적도 있다. 하지만, 중국의 경우 정부가 요구하면 거의 모든 정보를 제공하게 되어있으며, 관련 규정이 없더라도 서비스 중지 권한을 부여하는 등 규제의 강도 측면에서도 차이가 있다.

Ⅲ. 결론

본 논문에서는 인공지능 분야를 선도하는 미국, EU, 중국, 영국, 한국에서 추진 중인 '인공지능 안전'에 관한 주요 정책 동향과 양자, 다자간의 국제협력 동향을 살펴보았다. 주요국에서는 인공지능의 혁신과 규제 사이에서 균형점을 모색하며, 각국의 상황에 따라 접근법을 달리하고 있다. 또한, 미국, EU, 영국 등은 '인공지능 안전 글로벌 표준'을 위한 기술개발과 '질서 정립'을 위한 국제협력을 강조하고 있다. 우리나라도 신뢰할 수 있는 인공지능 개발 문화 및 인식 변화를 위한 정책적 지원이 필요할 것이다.

참 고 문 헌

[1] S. M. "Artificial Intelligence Policy Trends", IITP SPOT ISSUE, December 2023

[2] White House, "Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI", July 2023

[3] White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence", October 2023

[4] Council of the EU, "AI act: Council and Parliament strike a deal on the first rules for AI in the world", December 2023

[5] Cyberspace Adminstration of China, "Interim Measures for Generative Artificial Intelligence Service Management", July 2023

[6] National Development and Reform Commission of the People's Republic of China, "All kinds of enterprises will have broad space for development in China", March 2024

[7] Department for Science, Innovation and Technology, "AI regulation: a pro-innovation approach", March 2023

[8] GOV. UK, "Chair's Summary of the AI Safety Summit 2023, Bletchley Park", November 2023

[9] EU Commission, "G7 Leaders' Statement on the Hiroshima AI Process", December 2023

[10] European Commission, "Implementation of the Digital Decade objectives and the Digital Rights and Principles", September 2023