

인공지능 안전성 국제 동향 및 국제표준화 협력에 관한 연구

이혜진

katehjee@tta.or.kr

A Study on International Trends in AI Safety and International Standardization Cooperation

Hye Jin Lee

Telecommunications Technology Association (TTA)

요약

본 논문은 인공지능(AI) 안전성에 관한 국제 동향과 표준화 협력의 중요성을 탐구한다. 주요 국가들의 AI 안전성 관련 정책과 국제표준화 노력을 분석하고, 최근 영국과 한국에서 개최된 AI Safety Summit의 주요 내용을 포함하여 AI 안전성의 현재와 미래를 조망한다. 이를 통해 안전하고 신뢰할 수 있는 AI 기술 개발과 시스템 구축을 위한 국제표준화 협력의 필요성을 강조하고자 하였다.

I. 서론

인공지능 기술의 발전과 함께 AI 시스템의 안전성 확보는 중요한 과제가 되었다. 왜냐하면 AI는 다양한 분야에서 혁신을 이끌고 있지만, 동시에 그 잠재적 위험도 커지고 있기 때문이다. 각국 정부와 국제기구는 AI의 잠재적 위험을 최소화하고 안전한 AI 활용을 촉진하기 위해 다양한 정책과 표준화 작업을 추진하고 있다. 본 연구는 AI 안전성에 대한 국제적 동향을 파악하고, 표준화 협력의 현황과 향후 방향성에 대한 고려 사항을 제시해 보고자 한다.

II. 본론

1. AI 안전성 국제 동향

1.1 주요 국가별 동향

미국, 유럽 연합, 중국 등 주요 국가들은 AI 안전성을 보장할 수 있도록 AI 규제 측면에서의 다양한 정책을 수립·제시하고 있다. 미국은 2023년 AI의 안전성을 확보하고 기술혁신을 도모하기 위한 행정명령을 공개했다. 이 행정명령은 특히, AI 시스템을 출시하기 전 상무부 산하 국립표준기술연구소(NIST)의 안전성 평가를 의무적으로 거쳐야 한다는 내용을 포함하여 AI 기술이 대중들에게 공개되기 전 다방면에서의 안전성에 대한 사전 심사를 하도록 했다[1]. 유럽 연합(EU)은 2024년 3월, 세계 최초로 "AI 규제법"을 통과시켜, 빠른 속도로 발전하는 기술에 따른 위험으로부터 시민을 보호하는 동시에 혁신을 촉진할 것이라고 밝혔다[2]. 중국은 AI를 전체적으로 규제하기보다는 새로운 AI 제품이 등장할 때마다 개별 법안을 발표하는 형태로 단편적으로 세분화된 규제를 시행하고 있으나, 2023년 포괄적인 AI 법을 입법 의제로 삼겠다고 발표함에 따라 AI 기술의 윤리적 사용과 안전성 확보를 위한 법적 틀을 포괄적으로 마련하고자 노력하고 있다[3].

1.2 산업계 동향

AI 기술 개발을 이끌어어나가고 있는 글로벌 빅테크 기업들은 AI 안전성과

신뢰성을 확보하기 위하여 기업 별도의 윤리 원칙과 거버넌스를 구축하고, 관련한 R&D를 수행하거나 이해관계자들과의 협의 등을 통해 다양한 노력을 추구하고 있다. 또한, 국가 정책과는 별개로 안전성·신뢰성을 보장하는 AI 분야의 주도권을 확보하기 위한 연합을 결성하기도 했다. 특히 폐쇄형 AI 진영인 OpenAI, Google, Microsoft는 2023년 7월, 대규모 머신러닝 모델의 안전하고 책임감 있는 개발을 보장할 기관인 Frontier Model Forum의 결성을 발표했다. 이는, AI 안전 연구 발전, 모범 사례 규명, 다방면에 걸친 협업, AI가 사회의 가장 큰 과제를 해결할 수 있도록 지원하며, 안전·신뢰 AI 분야 연구를 촉진하기 위한 AI 안전기금을 조성했다. 개방형 AI 진영인 Meta와 IBM도 2023년 12월, 개방적이고 안전하며 책임감 있는 AI를 발전시키는 것을 목표로 50개 이상의 조직을 구성한 국제 연합 AI Alliance를 출범시켰다[4].

2. AI Safety Summit

최근 영국과 한국에서 개최된 AI Safety Summit은 AI 안전성에 관한 국제협력을 강화하는 중요한 행사였다. 2023년 11월 영국에서 열린 AI Safety Summit에서는 28개국 정부, 주요 AI 기업, 연구소 등에서 150여 명이 영국 블레츨리 파크에 모여 혁신적이면서도 위험성이 내포된 범용 AI를 프론티어(Frontier) AI로 정의하고 이의 위험성을 대비하기 위한 글로벌 공동연대의 약속을 담은 '블레츨리 선언(Bletchley Declaration)'을 발표했다[5]. 그리고 올해 5월 21일부터 이틀 동안 서울에서 각국의 정상들과 AI 담당 정부 고위 관료 등이 참석하는 'AI 서울 정상회의(AI Seoul Summit)'가 개최됐다. 작년 회의가 AI의 안전성에 집중했다면, 이번에는 AI 안전성뿐만 아니라 혁신과 포용성까지도 확대한 것으로 알려져 있다. 즉 안전한 AI를 만들기 위한 노력은 물론이고 생성 AI가 경제와 사회발전에 기여할 혁신의 가치에 대해서도 인정하고, 이러한 안전한 AI가 가져올 새로운 기회를 다양성과 포용성의 존중을 통해 소수의 특정 국가나 기업만이 아닌 전 세계 누구도 소외되지 않도록 함으로써 미래 AI 시대를 선도하는 아젠다를 논의했으며[6], 그 결과로 한국, 미국, 영국 등 10개국과 유럽 연합(EU)의 정상들은 서울 선언을 채택했다[7].

3. 우리나라의 현황

한국은 세계 주요국들의 AI 관련 정책 발표를 주시하는 한편, 국제 협의체들의 논의와 더불어 다양한 방면에서 노력을 기울이고 있다. 2024년 2월 발표한 “2024년 과기정통부 주요 정책 추진계획”에 따르면, AI-디지털 대전환 선도를 필두로 ‘인공지능법’, ‘AI안전연구소’ 등 AI 공존 시대를 선도하고, 국내에서 발표한 ‘디지털 권리장전’의 글로벌 확산을 위한 다양한 세부 정책을 추진하겠다고 밝혔다[8]. 또한, 2024년 5월 16일 자 전자신문 기사에 따르면, 한국 정부는 미국, 영국, 일본 등에 이어 AI 안전을 위한 연구 전담 조직을 구축할 예정이라고 밝혔다. AI의 순기능뿐 아니라 AI의 발달로 대두되는 다양한 문제점과 특히, 윤리적 AI 기술과 안전하고 신뢰성 높은 AI 기술의 개발과 발전에 대한 우려에 대한 해소를 위한 노력이 강화되고 있다. AI 안전연구소 연내 설립을 통해 해외 안전연구소들과 국제 협력을 위한 네트워크를 구축하고, AI법, 안전하고 신뢰 있는 AI 기술과 시스템 개발 및 발전을 위한 활동을 본격적으로 추진할 계획이라고 밝혔다[9]. 한국은 AI 분야에서의 글로벌리더십을 강화하고, AI 기술의 신뢰성을 확보하기 위해 AI 안전연구소를 설립하고 적극적인 대응과 후속 조치를 마련하고 있다.

4. 국제표준화 동향

국제 표준화 기구인 ISO와 IEC는 AI 기술과 관련한 표준을 개발하고 있다. 이러한 표준은 AI 기술의 안전한 사용을 촉진하고, 글로벌 시장에서의 상호운용성을 보장하는 역할을 한다. 특히, ISO/IEC JTC 1/SC 42는 AI의 안전성과 윤리에 관한 표준을 제정하고 있으며, 이는 전 세계적으로 일관된 AI 안전성 기준을 마련하는 데 중요한 역할을 하고 있다. SC42는 생성형 AI 윤리 문제, AI 신뢰성 평가, AI 시스템 개발조직의 능력을 평가하는 AI 성숙도 모델, AI 경영시스템 가이드 등과 관련한 표준을 논의하고 개발하고 있다[10]. 이외에도 IEEE, ITU 등 다양한 국제기구들이 AI 안전성과 관련된 표준을 개발하고 있으며, 이러한 표준들은 AI 시스템의 개발 및 활용 과정에서 중요한 가이드라인으로 작용하고 있다. IEEE는 2023년 초, 글로벌 ‘신뢰할 수 있는 AI(Trustworthy AI)’를 위한 지침과 고려 사항을 제공하는 인공지능 윤리 및 거버넌스의 글로벌 사회 기술 표준에 대한 무료 프로그램을 발표했다[11]. ITU도 2023년 7월, ‘선한 인공지능을 위한 글로벌 서밋(AI for Good Global Summit)’을 주최하여 인공지능 거버넌스 문제를 놓고 국제적인 협력과 노력이 필요하고 목소리를 높이는 등 AI 개발과 활용 측면에서 제기되고 있는 AI에 대한 우려에 대한 국제기구 차원의 노력을 기울이고 있다[12].

III. 결론

AI 기술의 발전과 함께 AI 안전성의 중요성은 날로 증가하고 있다. 각국의 정책과 국제표준화 노력은 AI 시스템의 안전성을 보장하고, 신뢰할 수 있는 AI 생태계를 조성하는 데 기여하고 있다. 특히, 최근 개최된 AI 서울 정상회의(두 번째 AI Safety Summit)는 AI 안전성에 대한 국제협력의 중요성을 재확인하고, 향후 발전 방향을 제시하는 중요한 계기가 될 것으로 예상된다. 이뿐만 아니라 각국의 AI 안전연구소 설립과 국제 네트워크 구축 및 공동의 노력은 AI 안전성 확보를 위한 국제적 협력을 강화하고자 하는 의지를 반영하고 있다고 볼 수 있다. 그러나 이러한 다양한 협력 활동과 정책 발표는 국가별, 지역별로 다양화되고, 세분화되고 있어 오히려 전 세계가 공통으로 적용하고 개발자나 사용자(이용자)들이 참고할 수 있는 AI 안전성과 신뢰성에 관한 국제표준 개발은 상대적으로 덜 활성화되어 있거나 더디게 움직이고 있는 것도 현실이다. 이는 국제 정세를 고려할 때, AI 기술은 국가 간 경쟁력의 중요한 요소로 작용하고 있기 때문이기도

하다. 하지만 또 한편으로는 미국과 중국을 비롯한 주요 국가들은 AI 기술 선점을 위해 치열한 경쟁이 오히려 AI 안전성에 대한 국제적 협력과 표준화의 필요성을 더욱 부각시키고 있기도 하다. 한국은 이러한 국제 정세 속에서 AI 안전성 연구와 국제협력에 적극 나서야 하며, 주요 AI 기술 선진국들과의 전략적 관계 형성을 기반으로 가속화되고 있는 AI 안전성과 신뢰성을 위한 국제표준 개발을 주도해 나가야 한다. 이를 통해 전 세계가 안전성과 신뢰성을 기반으로 한 AI 기술과 시스템의 다양성 강화를 위해 글로벌 협의체와 함께 노력하고, 글로벌 AI 연대의 리더로서 자리 잡기 위해 노력해야 한다.

ACKNOWLEDGMENT

본 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2024-00348454, ICT 표준 글로벌 협력체계 구축).

참 고 문 헌

- [1] 머니투데이, “바이든, AI 행정명령 공개...”AI 안전평가 의무화”, 2023. (<https://news.mt.co.kr/mtview.php?no=2023103022090258326>)
- [2] 동아사이언스, “EU, 세계 최초 ‘AI 규제법’ 통과...위반시 매출 7%까지 과징금”, 2024 (<https://www.dongascience.com/news.php?idx=64266>)
- [3] KISTEP, “EU 인공지능(AI) 규제 현황과 시사점” pp. 4-5, 2024
- [4] KISTEP, “안전·신뢰 AI” pp. 9-10, 2024
- [5] UK Government, “The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023”, 2023 (<https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>)
- [6] 머니투데이, “[기고] ‘AI 서울 정상회의’ 개최, 그 의미와 기회, 2024 (<https://news.mt.co.kr/mtview.php?no=2024050713384014182>)
- [7] 매일경제, “尹 “AI 안전연구소 설립 동참” AI 정상회의 ‘서울 선언’ 채택, 2024 (<https://www.mk.co.kr/news/it/11021403>)
- [8] 과학기술정보통신부 보도자료, “2024년 과기정통부 주요 정책 추진계획 발표”, 2024 (<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3184072>)
- [9] 전자신문, “국내 ‘AI 안전연구소’ 설립 핵심 의제로” 2024 (<https://www.etnews.com/20240516000007>)
- [10] 전자신문, “세계 AI 전문가 한국 집결...국표원, ‘AI 국제표준화 총회’ 개최”, 2024 (<https://www.etnews.com/20240422000014>)
- [11] 인공지능신문, “IEEE, ‘신뢰할 수 있는 AI’ 위해 나섰다... ‘인공지능 윤리 및 거버넌스 인증’ 글로벌 무료 프로그램 시작”, 2023 (<https://www.aitimes.kr/news/articleView.html?idxno=27177>)
- [12] 로봇신문, “인공지능(AI) 거버넌스, 국제적인 협의 필요하다”, 2023 (<https://www.irobotnews.com/news/articleView.html?idxno=32087>)