

AI 규제와 원자력을 위한 합의

Dr. Frances Mautner Markhof 오스트리아 국제연구센터 사무총장¹



현재, Chat GPT나 GPT-4와 딥러닝 모델로 구축된 인공지능(AI)들의 안정성, 신뢰성, 투명성 등을 보장하기 위해 국제적인 AI 규제가 필수적이라는 생각이 널리 퍼지고 있다. 이러한 국제적인 규칙과 규제의 개발에는 모든 관련 국가 및 기관이 참여해야 한다.

지금까지의 AI 규제

2023년 12월 9일, EU는 인공지능에 관한 포괄적인 법적 틀인 “EU AI Act”에 합의했다. 이 법률은 2024년 초 공식적으로 발효될 예정이며, 다양한 AI 위험에 관한 규정들이 포함되어 있다. 프랑스, 독일, 이탈리아는 2023년 11월 18일 EU 당국과 함께 미래의 AI 규제에 관한 합의를 도출했다. 푸틴 러시아 대통령은 “핵 비확산 규칙과 유사한 인공지능 통제 의 공통원칙”을 제안한 바 있다.

2023년 11월 1일부터 2일까지 영국에서 개최된 “AI 안전성 정상회의”에서는 안전하고 신뢰할 수 있는 AI 및 AI 규제, AI 위험 통제에 대한 공감대가 형성되었다. 이 회의에서 EU, 미국, 중국을 포함한

많은 국가들은 “AI의 위험을 공동으로 관리”하기로 합의했다. UN은 선언을 승인하였으며, 후속 회의는 2024년에 프랑스에서 개최될 예정이다.

2023년 10월 30일, 바이든 미국 대통령은 AI 규제에 관한 행정명령을 발표하여 AI 응용 프로그램의 안전과 보안을 보장하고, 소비자를 보호하며, 혁신과 경쟁을 촉진하기로 결정했다. 미국의 일부 주에서는 이미 AI 사용을 제한하는 법률이 통과되기도 했다.

2023년 10월 26일에는 UN 사무총장이 개발도상국과 선진국을 포함한 39명의 회원으로 구성된 “인공지능에 관한 고위급 자문기구”를 임명했다. 이 자문기구는 AI의 위험, 기회 및 국제적 거버넌스 등을 다루게 될 것이다.

2023년 9월 7일에는 G7 국가들이 향상된 AI 시스템을 위한 지침과 기업을 위한 국제 행동강령 개발을 약속했다. 향상된 AI 시스템을 개발하는 기업을 위한 국제 지침 원칙의 초안은 안전, 보안, 신뢰성, 투



명성, 위험 관리 및 국제 표준의 필요성을 강조한다. EU는 현재 이러한 지침 원칙들을 검토하고 있다.

2023년 7월에는 중국의 인터넷 규제 기관이 생성적 AI에 관한 최종 규칙을 발표했다. “생성형 인공지능 서비스 관리 방안”으로 알려진 이 규정은 서양에서 논의되고 있는 여러 원칙들을 포함하고 있다. AI 모델의 지적 재산권 침해, AI에 의한 개인정보 침해, AI 알고리즘에 의한 차별성, 그리고 데이터 및 교육에 관한 AI 기업의 투명성 등이 그것이다.

즉, 최고위 정치 차원에서는 AI 규제의 중요성과 긴급성에 대해 명확한 인식이 확산되고 있다. 원자력 확산을 통제했던 경험이 여기에 도움이 될 수 있을 것으로 기대된다.

지금 이 순간, 국제적 AI 규제가 필요하다

국제원자력기구(IAEA, International Atomic Energy Agency)는 원자력에너지의 평화적인 사용을 촉진하고 비평화적인 활동을 금지, 통제하기 위해 효과적이고 국제적으로 수용되는 시스템

을 구축하기 위해 설립되었다. 이와 동일한 이유로 국제인공지능기구(IAIA, International Artificial Intelligence Agency)가 설립되어야 한다. 이는 인공지능의 규제를 비롯해 인공지능의 안전성, 투명성 및 유익한 발전을 보장하기 위한 것이다.

이것은 특히 개별 국가나 집단이 AI 규제를 위해 서로 다르고 상충할 가능성이 있는 자체적 시스템을 개발하는 것과 달리 통일된 국제 시스템을 구축한다는 점에서 시급한 문제이다. 일론 머스크와 워런 버핏은 AI와 관련된 위험의 심각성과 그 크기가 핵무기와 관련된 위험보다 더욱 크다고 강력하게 강조한 바 있다.

IAIA는 AI에 대한 국제적 규제와 국가 AI 시스템 및 어플리케이션에 대한 독립적이고 객관적인 검증을 수행하여 국가 및 국제적 의무가 준수되고 있음을 보장할 책임이 있다. 이러한 검증 과정은 실재하는 금지된 AI 개발 및 응용을 식별하면서 안전하고 평화로운 목적만을 위해 AI가 개발되고 사용되고 있음을 보장할 것이다.



IAEA 안전조치: IAIA를 위한 모델

IAEA의 경험은 매우 높은 관련성이 있다. IAIA 규제의 이유와 역할은 본질적으로 IAEA의 국제 안전조치 시스템이 만들어진 이유와 일치한다. IAEA 안전조치는 국가 안전조치 시스템의 독립적인 검증을 포함하여, 핵물질, 핵물품 및 핵시설이 평화적인 목적으로만 사용되고 있는 것에 대한 신뢰성 있는 보증을 제공한다.

마찬가지로, IAIA도 인공지능이 안전하고 평화적인 목적으로 사용되는 것을 촉진하면서, 동시에 인공지능이 불법적이거나 불안정한 목적으로 사용되지 않도록 검증하고 통제하는 두 개의 임무를 가질 것이다. IAIA 참여 국가들이 지켜야 할 의무와 받을 혜택을 정리하면 아래와 같다.

- 국제적인 AI 규정을 준수하고 자국의 AI 활동에 대해 국제적 검증을 수용.
- 국가의 법적 및 기타 의무를 준수하며 인공지능의 평화적이고 이로운 사용을 위해 모든 관련 기관의 AI 연구 및 응용 프로그램, 기타 인공지능 활동을 신고.
- 국가 AI 선언의 완전성과 정확성을 검증.
- 국제 검증관 자격을 부여하여 인공지능 활동 선언을 검증, 확인하여 신고되지 않은 AI 활동을 밝힘.(이는 핵물질 및 활동에 관한 IAEA 안전조치 검증관의 역할과 유사하다.)
- AI 기술에 대한 접근 및 평화적 목적을 위한 자국의 AI 시스템 개발을 위한 기술 협력.

- AI가 불법적이거나 반사회적, 또는 범죄적 목적으로 사용되거나 그럴 것이 의심되는 경우에 작동할 조기경보 시스템 구축.
- AI 시장을 통제하기 위한 불법적인 개발, 응용 프로그램 및 경쟁을 식별하고 방지.
- 국제적인 AI 데이터베이스에 대한 인공지능 검사 및 기타 관련하여 합의된 데이터를 제공.
- IAIA 규정의 심각한 위반사항은 IAEA 안전 조치의 심각한 위반사항과 마찬가지로 UN 안보리에 제출될 수 있음.

IAIA: AI 규제에서의 포괄적 접근 vs 배타적 접근

IAIA는 생성적 AI의 평화적이고 생산적인 사용을 촉진하고 보장하기 위해 국제적인 프로세스를 통해 생성될 것이다. IAIA가 효과적으로 기능하려면 AI 활동에 관심을 가진 모든 국가들과 관련 조직, 전문가들이 참여하는 국제적인 협력이 필요하다. 국제적인 접근만이 범지구적 차원에서 이뤄지는 AI 기술의 도전, 급속한 발전에 대처하는 데 성공할 수 있다. 따라서 G7과 EU뿐만 아니라 브라질, 중국, 인도, 대한민국, 러시아 등 다른 주요 국가들도 처음부터 참여해야 할 것이다.

소셜 미디어, 바이러스, 가짜뉴스들의 범지구적인 가용성과 영향력을 고려할 때, AI 규제를 위해 가능한 가장 넓은 망을 짜는 것은 필수적이다. 어떤 국가나 조직도 빠져서는 안 된다. 새로이 구성된 UN AI 고위급 자문기구가 빠른 시일 내에 IAIA 등 국제 IA 규제를 위한 포괄적 절차에 착수하기를 기대한다.

Reference

¹ Dr. Frances Mautner Markhof는 오스트리아 국제연구센터(Austrian Center for International Studies)의 사무총장으로, IAEA에서 중사한 경험이 있다. 주요 연구 관심사는 핵비확산과 동북아 안보이다.