국가전략연구위원회 이슈브리핑 33호 (발간일: 2024.9.30.)

AI·사이버 외교 분야 한국의 과제[1)]

김소정 (국가안보전략연구원)

Ⅰ. 서론

한국은 2022년 디지털 심화 시대의 기본 방향을 제시한 뉴욕 구상을 시작으로 작년 유엔 총회 기조연설과 디지털 권리 장전 수립을 통해 새로운 AI 디지털 규범 정립에 앞장서 왔다. 지난 5월에는 주요 국가들과 함께 AI 서울 정상회의를 개최해서 안전, 혁신, 포용의 3대 원칙을 담은 서울 선언을 도출했으며, 「2024 인공지능의 책임있는 군사적 이용에 관한 고위급회의(REAIM : REsponsible AI in the Military domain)」을 개최하고 "REAIM 행동을 위한 청사진"을 발표하는 등 AI 국제 규범을 정립하고 상호간 신뢰가능하고 투명한 AI를 개발 및 사용하고자 노력하고 있다. 또한 국가인공지능위원회를 운영하며, AI의 안전성과 보안성을 담보하는 역할을 할 AI 안전연구소 설립을 추진하고 있다.[2)]

UN, OECD 등 다양한 형태의 국제기구들도 AI 사용에 관한 원칙 등을 공개했으며, 각국은 AI 수용 및 활용, 국가간 활용 등에 대비한 자국의 전략·정책 문서들을 공개하고 있다. 반면, 사이버안보 규범에 대한 논의는 UN의 개방형 워킹그룹(OEWG : Open-Ended Working Group)의 논의만 지속되고 있어, 상대적으로 사이버안보에 관한 규범 논의는 활발하지 않은 것으로 보인다. 사이버분야의 규범논의가 최근 주춤해진 것은, 이미 30여년간 지속되었음에도 불구하고, 국제법 등 경성규범의 형태로 사이버공간에 적용가능한 규범에 합의하지 못했으며, 비가시적인 사이버공간을 가시적 형태로 그 사용을 규범화시키는 데 한계

1) 이 글은 2024년 9월 25일 제11차 사이버 국가전략 포럼에서 발표된 발표문을 수정보완한 것이다.
2) 대한민국 정부, "국가인공지능(AI)위원회 출범식 및 1차회의 모두 발언", 정부24, 2024.09.26.

를 느끼는 때문인 것으로 보인다.

사이버 규범논의가 주춤해진 사이, 오히려 AI, 우주 등 타 과학기술분야 규범논의는 활발해져 왔다. 이들은 모두 국가안보에 치명적 위해를 가할 수 있는 파괴적인 역량을 보유하고 있다는 공통점 외에도, IT에 기반한 사이버안보 및 보안의 요소가 필수적으로 요구된다는 점에서 또 다른 공통점을 갖고 있다.

Ⅱ. 인공지능(AI)의 군사적 이용

AI는 시대를 규정하는 기술이 될 가능성이 있으며, 경제, 사회, 정치적 성공을 결정할 수 있다. 각국이 AI를 얼마나 잘 활용하느냐에 따라 군사 및 경제 균형이 변화할 수 있다. 특히 자율 무기 시스템(LAWS)은 군사적 측면에서 큰 우려를 불러일으키고 있다. 인공지능(AI)의 급속한 기술 발전은 군사적 안정성에 새로운 위험을 초래하며, 이를 완화하기 위해 각국은 규범과 신뢰 구축 조치(CBMs)를 마련하고자 노력하고 있다. 이러한 위험을 이해하고 완화책을 시행함으로써 의도치 않은 군사적 확전을 예방할 수 있기 때문이다.

현재 선진 민주주의 국가들간 긴장관계는 AI 규범형성을 어렵게 만들고 있다. 그럼에도 불구하고 채택된 동 청사진이 효력을 가지기 위해선, 국가간 협력을 모색하는데 있어 정치적 실현 가능성을 높이는 것이 중요하다. 사이버안보 규범 논의가 제자리 걸음을 하는 결정적 이유도, UN의 정부전문가그룹(GGE : Group of Governmental Expert)에서 중국, 러시아 등 국가가 모두 합의한 "주요기반시설 대상 사이버 공격 금지"라는 규범이 깨졌기 때문으로 볼 수 있다. 중, 러가 합의한 원칙임에도 불구하고 정치적 목적에 의해 위성통신망, 에너지망 등에 대한 사이버공격이 버젓이 일어남에 따라 규범의 효과에 대한 신뢰가 상실된 탓이다.

그럼에도 불구하고, 경쟁하는 국가들간의 공통의 우려는 규범이나 CBMs 준수를 장려할 수 있으며, 이는 핵의 전략적 안정성이나 AI 안전 및 보안을 위한 국제 평가 체계와 같은 협력 가능한 분야를 포함하게 된다. 특히 정치·군사 위기 상황에서는 AI 시스템이 의사결정을 지원하는 역할을 하되, 인간의 통제 아래에 두는 것이 매우 중요하다. 인간의 상황 인식을 보장할 수 있도록 투명한 추론과 논리 체계의 검증을 지원하는 AI 시스템을 구축하는 것이 필수적이다.

이러한 측면에서, 지난 9월 개최된 REAIM은 청사진에서 책임있는 군사분야 AI 이용에 대해 △AI 적용은 윤리적·인간중심적일 것 △군사분야 AI 역량은 국내법·국제법에 합치하도록 적용할 것 △인간이 AI 적용에 책임을 져야하고 어떠한 경우에도 기계에 전가하지 말 것 △오작동과 데이터 편향 등으로 발생하는 위험을 줄이기 위한 보호장치를 마련할 것 △군사분야 AI의 개발·배치·이용시 인간의 적절한 개입이 유지될 것 등에 합의했다[3]는 점에서는 고

3) 박민희, ""핵 사용에 AI 오용 안돼, 인간의 통제·개입 유지되야", 한국 주최 'REAIM 고위급회의'서 '행

무적이라 볼 수 있다. 다만, 그 실현가능성을 높이기 위해 합의문 채택 동의 국가에 더 많은 국가를 포함시켜야 하며, 그 실효성을 검증할 수 있는 체계를 구축하는 것이 필수적이다.

정치적·군사적 역학에 대한 이해 없이 AI 기능을 통합하면 예상치 못한 결과를 초래할 수 있다. 군사 분야에서 AI 시스템의 운영을 평가하고, 가능한 위험을 예측하며, 완화책을 제안하는 전담 연구 그룹을 구성해야 한다. 특히, 의사결정자들은 AI 기능과 그 잠재적 영향에 대한 충분한 이해가 부족한 채로, 정확하고 필요한 정보에 근거한 정책 결정 및 전략 계획을 세우지 못할 수 있다. 이러한 정보 격차는 AI 위험, 모범 사례, 완화 방안, 적시 정보 교환 및 사건 대응을 위한 정부 간 접점을 마련함으로써 해결할 수 있다.

Ⅲ. AI와 외교

지정학 측면에서, AI가 발전한 국가는 군사적, 경제적, 사회적 힘을 더 많이 발휘할 수 있다. AI는 다양한 협정과 조약이 논의되는 외교 의제의 중요한 주제이기도 하다. AI는 국제 의제에서 매우 중요한 주제로 떠오르며, 정치, 사회, 경제 전반에 걸쳐 많은 영향을 미치고 있다. AI는 앞으로도 국제적 논의에서 중요한 역할을 할 것이며, 사회와 국제 관계를 계속해서 변화시킬 것이다.

최근 몇 년 동안 인공지능(AI) 분야에서는 상당한 진전이 있었으며, 이는 지능형 디지털 개인 비서, 스마트 홈 기기, 자율주행차, 스마트 빌딩, 의료 로봇과 같은 형태로 우리의 일상생활에 점점 더 통합되고 있다. 이는 더 이상 공상 과학 소설의 이야기에 머물지 않는다. 이러한 발전은 경제, 사회, 교육 등 여러 정책 분야뿐만 아니라 외교, 인프라, 사회 전반에 걸쳐 영향을 미칠 것으로 예상되며, 전 세계의 정부, 기술 커뮤니티, 민간 부문은 이를 점점 더 중요한 이슈로 고려하고 있다.

특히 AI 시스템은 방대한 양의 데이터를 처리하므로, 이는 개인정보 보호와 데이터 보호와 관련된 우려를 불러일으킨다. 데이터 처리 과정에서 정부는 AI를 설계, 개발, 배포하는 과정에서 인권을 존중하도록 민간 부문을 독려해야 할 의무가 있다. AI 시스템 개발자들은 사용되는 데이터의 무결성을 보장하고, AI 애플리케이션에 개인정보 보호 및 데이터 보호 기능을 내재화해야 한다.

또한 AI에 대한 외교와 동시에 외교에 AI를 활용하는 방식도 고민이 필요하다. 이는 전세계적으로 외교분야에서 외교행정 처리에 있어, 데이터의 적절한 처리와 의사결정 지원에 AI를 활용하고자 하는 노력과 일맥상통하다고 볼 수 있다.

사이버외교의 노력이 2012년 런던프로세스를 추진하면서 지속된 바와 동일하게 AI분야 외고의 역할도 결정적일 수 있다. 사이버외교 분야는 규범 형성을 위한 다양한 노력의 경주, 신뢰구축조치의 개발 및 이행, 역량강화 지원 등 다양한 형태로 진행되었다. AI분야의

동을 위한 청사진' 채택", 한겨레신문, 2024.09.10.

외교도 유사한 형태로 발전해나갈 것으로 예상된다. 다만, 사이버와 AI 등 과학기술분야에 있어 외교의 스펙트럼이 넓기 때문에 기술적 이해에 기반한 전문지식 확보가 가장 중요한 선결문제가 될 것이다.

Ⅳ. AI의 안전성과 보안성, 사이버 보안과의 연계

생성형 AI 기술은 사이버 공격 및 사이버 스파이 행위의 주요 표적이 될 수 있다. 이들의 훈련 데이터는 조작될 수 있으며, 사이버 취약성이 있는 기술 기반에서 실행된다. 내부 데이터를 분할하고 접근을 제어할 수 있는 효과적인 보안 구조는 존재하지 않는다.

사이버 공격역량을 지원하는 AI 수준은 현재 사이버 방어 지원보다 앞서가고 있다. 그러나 충분한 자원이 투입된다면, 방어 능력은 극적으로 향상될 수 있을 것이다. 향상된 사이버 방어 역량은 AI의 안전성과 보안성에 신뢰를 구축하고 국제 관계를 안정시킬 수 있을 것이다. AI의 안전성과 보안성을 향상시키기 위해 1) 설계 단계에서의 안전성 강화를 위한 보안 설계, 2) 데이터 익명화 및 검증, 암호화, 3) AI 모델 편향 제거, 성능 검증 및 보안테스트 실시 등이 요구된다. 이들은 기관별 보안정책으로 반영되어야 하며, 지속적으로 감시 및 모니터링 해야 된다.

역사적으로 기술 산업은 메모리 안전성 등 컴퓨터 보안 과학을 구현하는 데 실패했으며, 이는 지난 35년간 지속된 사이버 불안정성 위기에 크게 기여했다. 민간주도의 기술개발에 AI의 발전을 모두 일임하게 되면, 경쟁 압박과 이익에 비해 보안 및 안전에 대한 유인책이 약한 AI가 광범위하게 배치될 것이다. 이는 정보기술의 급속한 사용확대시 보안을 고려하지 못했던 것과 유사한 실패로 이어질 수 있다. 정부는 AI와 관련된 실패와 재난에 대비하여 예방, 완화 및 사고 대응을 위한 정책을 수립해야 한다.

주요 자산을 보유하고 있는 기술 회사들은 그들의 전문성과 규제 이행 능력 때문에 AI 거버넌스에서 중요한 역할을 해야 한다. AI 사이버 보안 및 AI 거버넌스에 대한 공공과 민간의 협력이 더 나은 규제 우선순위 설정과 효과적인 규제를 이끌어낼 수 있다.

Ⅴ. 결론

사이버안보 규범논의와 유사하게 AI의 규범형성과 외교적 활동에 대해서 한국은 아직 많은 내부적 논의가 진행되지 않고 있다. 정책지향성 부재, 전문인력의 부족, 학제간 소통 및 문제의식 공유 부재, 정책결정 지원 가능한 생산의 한계, 국가 R&D와의 정합성 부족, 지속적 실행력 부재 등이 지속되고 있으며, 결국 발달하는 과학기술을 외교와 정책, 전략적 측면의 접점을 이해하고 풀어가는 데 아직 충분한 사유를 하지 못한 결과일 것이다. REAIM

및 AI 정상회의 등을 기반으로 한국내 전사회적인 접근이 가능할 수 있는 체계를 마련하고 이를 지속하면서, 논의 주제도 군사적 사용 외 민간 사용, 기술개발의 정책적 함의, 표준, 윤리 및 인권적 측면 등까지 포괄할 수 있도록 다양한 분야가 고려되어야 한다.

<참고자료>

※ 이슈브리핑에 수록된 내용은 필자 개인의 의견이며, 한국사이버안보학회의 공식적인 입장이 아님을 알려드립니다.