한국사이버안보학회
070.8019.3590 l kacs@kacs.ne.kr
The Korean Association of Cybersecurity Studies
이슈브리핑
국가전략연구위원회 이슈브리핑 53호 (발간일: 2025. 12. 26.)

인공지능(AI) 시대의 디지털 지정학:

인공지능 안전(AI Safety) 지정학과 중국의 대응 전략

백서인 (한양대학교 ERICA)

1. 글로벌 AI 생태계 경쟁과 AI safety

1) AI 가치사슬 디커플링과 오픈소스 생태계의 역설

미국과 중국 간 AI 패권 경쟁은 단순한 기술 우위 경쟁을 넘어 경제·안보·규범이 융합된 총체적 경쟁으로 전개되고 있다. 미국은 첨단 AI 칩(H100, A100 등)의 중국 수출을 제한하고, 이에 대응하여 중국은 화웨이, 알리바바 등을 중심으로 독자적인 AI 생태계 구축을 추진하고 있다. 흥미로운 지점은 AI 오픈소스 생태계 경쟁의 전개 양상이다. AI 발전 초기에는 오픈소스를 기반으로 한 협력적 혁신이 주를 이루었으나, 생성형 AI의 상업적 가치가 입증되면서 미국 빅테크의 폐쇄형 전략이 강화되고 있다. Open AI, Anthropic 등 주요 AI 기업들은 모델의 세부사항을 공개하지 않으며, 논문화나 특허 출원에도 소극적인 모습을 보이고 있다. 그러나 중국은 오히려 오픈소스 전략을 강화하고 있다. Qwen, DeepSeek 등 중국의 오픈소스 모델들이 글로벌 시장에서 빠르게 확산되고 있으며, 2024년 하반기부터 중국 모델의 글로벌 다운로드 수가 미국을 추월하였다. 이는 중국이 미국의 기술 봉쇄에 대응하여 오픈소스를 통한 생태계 확장 전략을 추구하고 있음을 보여준다.

2) 안보중심의 AI 규제 논의와 AI 거버넌스의 다원화

AI 규제 논의는 초기 “AI Safety(안전성, 공정성, 투명성, 책임성)”에 중점을 두었으나, 최근 “AI Security(국가안보, 기술 우위, 경제 경쟁력)”로 무게중심이 이동하고 있다. 특히 트럼프 2기 행정부 출범 이후 미국은 바이든 행정부의 AI Executive Order를 철회하고, 미국의 글로벌 AI 지배력 확대를 우선시하는 정책으로 전환하였다.

이와함께 2025년 7월 발표된 미국의 AI Action Plan, 2025년 8월 발표된 중국의 WAICO(World AI Co-operation Organization) 구상, 2025년 11월 EU의 AI Act 시행 연기 등 주요 국가 및 권역별로 상이한 AI 거버넌스 체계가 경쟁적으로 형성되고 있다. 미국은 3개 필라(혁신, 인프라, 수출/규범)를 중심으로 미국의 글로벌 AI 지배력 확대에 집중하고 있으며, 중국은 13개 조항으로 구성된 포괄적 AI 협력 체제를 제시하고 있다. EU는 자국의 과도한 규제를 완화함과 동시에 11

개 전략 산업 부문별 플래그십 과제를 추진하고 있다. 이와 함께, UN, OECD, APEC, G7/G20등 다양한 다자체제에서 인공지능 논의가 복잡하게 전개되고 있다.

2. 중국의 인공지능 안전 내실화와 국제화 전략

1) 법률 및 제도적 기반

중국은 인공지능의 활용에 있어 사회 통제, 인권 침해, 중앙집권형 거버넌스 등의 이유로 서방 국가들의 비판을 받아오고 있지만, 역설적으로 자율주행, 디지털 헬스케어, 스마트 제조, 스마트 시티 등 다양한 신산업 영역에서 세계에서 가장 선도적으로 인공지능을 적용하고 있다. 이처럼 광범위한 영역에서 중국은 국가 차원에서 중국식 디지털 혁신의 표준화 작업을 꾸준히 추진해 오고 있는데, 그중 가장 중점적으로 추진하고 있는 분야가 인공지능 기술과 안전에 관련된 표준 작업이라고 할 수 있다.

중국은 2017년 발표한 ‘차세대 인공지능 발전 계획’에서 이미 인공지능 발전의 불확실성이 가져올 새로운 도전에 대해 언급하면서, 인공지능을 발전시키는 동시에 잠재적인 리스크와 도전을 중시하고 사전 예방을 통해 안전하고 신뢰할 수 있으며 통제할 수 있는 발전을 보장해야 한다고 명시하였다. 또한 인공지능 법률, 규범, 정책 체계 구축을 통해 인공지능 안전 평가와 관리 통제 능력을 구축해야 한다고 강조한 바 있다.

이후 2017년 6월 1일 시행된 ‘중화인민공화국 네트워크안전법’, 2021년 시행된 ‘중화인민공화국 데이터안전법’과 ‘중화인민공화국 개인정보보호법’은 생성형 인공지능 거버넌스 법률법규 발전의 기반이 되었다. 이들 규정에는 추천 알고리즘 관리 조치와 합성 생성 이미지 및 생성형 인공지능 유사 서비스에 대한 관리 세칙도 포함되었다.

중국의 현행 법률의 책임 조항은 기본적으로 생성형 인공지능의 안전 위험을 포괄하고 있으며, 데이터, 개인정보, 네트워크 안전, 알고리즘 추천, 딥페이크, 콘텐츠 정보 및 생성형 인공지능 서비스와 사용에 관한 규정이 모두 포함되어 있다. 예를 들어 ‘생성형 인공지능 서비스 관리 임시방법’ 제4조는 생성형 인공지능 서비스를 제공하고 사용할 때 법률과 행정법규를 준수하고 사회 도덕과 윤리를 존중해야 한다고 규정하고 있다.

[그림 2] 중국의 생성형 인공지능 관리규정에서 강조되는 핵심 가치

첫째, 사회주의 핵심가치관을 견지하고 국가 안전과 이익을 위협하거나 국가 이미지를 훼손하는 내용을 생성해서는 안 된다는 것이다. 둘째, 알고리즘 설계, 학습 데이터 선택, 모델 생성 및 최적화, 서비스 제공 등의 과정에서 민족, 신앙, 국적, 지역, 성별, 연령 등에 대한 차별을 방지하기 위한 효과적인 조치를 취해야 한다. 셋째, 지식재산권과 상업도덕을 존중하고 알고리즘, 데이터, 플랫폼 등의 우위를 이용해 독점과 부당 경쟁 행위를 해서는 안 된다. 넷째, 타인의 합법적 권익을 존중하고 초상권, 명예권, 사생활권, 개인정보 권익을 침해해서는 안 된다. 다섯째, 서비스 유형의 특징을 바탕으로 효과적인 조치를 취해 생성형 인공지능 서비스의 투명성을 높이고 생성 콘텐츠의 정확성과 신뢰성을 제고해야 한다.

중국의 인공지능 관련 법률이 작동하는 거버넌스를 살펴보면 크게 4개 층위로 구성되어 있음을 알 수 있다. 첫 번째 층위는 국가인터넷정보판공실이 총괄하고 국무원 공업정보화부, 공안부

등 관련 부서가 협력하는 형태로 구성되고, 두 번째 층위는 지방 네트워크 판공실이 총괄하고 지방 공안부 등 관련 부서와 업계 협회, 관련 기업이 협력하여 구성되며, 세 번째 층위는 기술이나 서비스를 제공하는 기업으로 구성되고, 마지막인 네 번째 층위는 개인, 조직, 기관 등을 포함한 사용자 집단으로 구성된다. 이렇게 구조화된 거버넌스 체계는 각 거버넌스 주체의 직책과 권한 구분을 명확히 함으로써 생성형 인공지능의 안전 거버넌스의 효율성 제고를 목표로 한다.

2) 안전 평가와 비안(备案) 제도

안전 평가는 사이버 안전과 데이터 거버넌스로 구성되어 있다. 네트워크 안전은 네트워크 핵심 설비 및 네트워크 안전 전용 제품 안전 인증 및 감독과 클라우드 컴퓨팅에 대한 안전 검사로 구성되어 있으며, 데이터 거버넌스는 데이터 안전의 평가와 앱 및 개인정보 수집 정황에 대한 테스트로 구성된다. 네트워크 안전 체계가 비교적 체계적으로 구축된 데 비해 중국의 데이터 안전 거버넌스는 아직 시작 단계에 있는 것이 사실이다. 중국은 국가 차원의 '데이터안전법'을 2021년에 공포하였고, 2022년 사이버공간관리국과 국가시장감독관리총국이 공동으로 제정한 '데이터 안전 관리 인증 실시 규칙'을 시행하면서 네트워크 운영자의 네트워크 데이터 수집, 저장, 사용, 가공, 전송, 제공, 공개 등 처리 활동을 규제하고 있다.

비안(备案) 제도는 기업 자율 규제와 정부의 사후 규제 및 관리 감독이 결합된 중국식 모델 관리 시스템으로 규정할 수 있다. 기업이 정부가 규정한 범위에 해당하는 서비스, 예컨대 인터넷 정보 서비스 알고리즘, 딥페이크 서비스 알고리즘, 생성형 인공지능 서비스 등을 개발하여 출시하기 전에 자체적으로 국가의 법령을 준수하는지에 대한 분석을 진행하여 승인을 신청하고, 국가가 심사 후에 이를 허가하는 형태로 추진된다.

2023년 12월, 중국은 '인터넷 정보 서비스 알고리즘 추천 관리 규정'을 정식으로 공포했는데, 이 규정은 알고리즘 비안의 법률 의무를 확립하고 알고리즘 안전 평가, 알고리즘 검사와 함께 알고리즘 감독관리 체계의 3축 체계를 구성했다는 점에서 큰 의미를 지니고 있다. 2023년 1월부터 시행된 '인터넷 정보 서비스 딥페이크 관리 규정'에서도 딥페이크 서비스 제공자에게 비안 및 변경, 취소를 이행할 것을 명확히 요구하였으며, 이러한 비안 규정은 '생성형 인공지능 서비스 관리 임시방법'에서 더욱 세분화되었다.

3) 표준화 및 국제화 전략

중국 정부는 2020년 국가 표준, 지역 표준, 기업 표준 등의 형태로 분산되어 있던 자국의 인공지능 기술 표준을 통합 조정하여 세계 최초의 통합 인공지능 표준 초안을 공개하였다. 최초로 공개된 인공지능 초안을 기반으로 지속적으로 개정되고 있는 '국가 인공지능 산업 종합 표준화 체계 가이드라인'의 2024년 최신 버전에는 2026년까지 중국이 표준과 산업 과학기술 혁신의 연계 수준을 꾸준히 향상시켜 50개 이상의 새로운 국가 표준과 산업 표준을 제정하고, 20개 이상의 국제 표준 제정에 참여하며, 1,000개 이상의 기업을 표준 홍보와 실행 보급에 참여시킴으로써 인공지능 산업의 글로벌화 발전을 촉진할 것을 천명하였다.

이 가이드라인은 스마트 도시, 과학 계산, 스마트 농업, 스마트 에너지, 스마트 환경 보호, 스마트 금융, 스마트 물류, 스마트 교육, 스마트 의료, 스마트 모빌리티 10개 분야의 인공지능 응용 표준화에 집중하고 있다. 특히 주목할 만한 점은 이러한 기반 기술과 산업 응용의 표준 제정과 함께 안전과 거버넌스 분야의 표준 활동 제정에도 힘을 쓰고 있다는 점이다. 안전 표준은 인공지능 기술, 제품, 시스템, 응용, 서비스 등 전체 생명주기의 안전 요구사항을 규범화하며, 기초 안전, 데이터·알고리즘·모델 안전, 네트워크·기술·시스템 안전, 안전 관리 및 서비스,

안전 테스트 평가, 안전 라벨링, 콘텐츠 식별, 제품 및 응용 안전 등의 표준을 포함한다.

적극적인 국내 표준화 작업을 기반으로 중국 정부는 자국 디지털 규범의 국제화에도 박차를 가하고 있다. 2023년 8월 세계 최초로 공개한 '생성형 인공지능 관리 방안'과 2023년 10월 20일 발표한 '글로벌 인공지능 거버넌스 이니셔티브', 그리고 2025년 7월 발표한 '글로벌 인공지능 거버넌스 액션플랜' 등이 그 대표적인 사례이다.

중국의 인공지능 관련 글로벌 이니셔티브와 거버넌스는 포용적 활용 가치와 인간 중심의 사용 원칙을 강조함과 동시에 타국 주권 존중, 국가 체제와 기술 활용의 연계 및 배타적 그룹 형성에 대한 반대를 천명하면서 중국식 인공지능 안전 규범의 세계화 의지를 보여주고 있다. 이는 중국이 단순히 국내 시장을 관리하는 것을 넘어 글로벌 AI 거버넌스에서 규범 제정자로서의 역할을 추구하고 있음을 의미한다.

3. 시사점

1) 중국 AI 거버넌스의 강점과 한계

중국의 인공지능 안전 규범은 국가 주도의 선제적 대응, 다층적 거버넌스 모델 도입, 영역별 특화된 안전성 평가 체계 구축을 통해 규제의 신속성과 실효성, 국가 경쟁력 확보 측면에서 성과를 거두고 있다. 특히 생성형 인공지능의 출현 이후 빠른 속도로 법제도를 정비하고 관리 체계를 구축한 것은 중국 정부의 강력한 추진력과 조정 능력을 보여주는 사례라고 할 수 있다. 또한 비안(备案) 제도를 통한 승인 방식은 기업들로 하여금 출시 전에 법령 준수를 철저히 점검하도록 함으로써 사후 규제의 부담을 줄이는 효과도 있다. 전세계적으로 인공지능 안전 관리 체계가 기업 및 활용 친화적으로 변화하는 시점에서 이러한 중국식 모델의 확산·채택 가능성이 높아진 것이 사실이다.

그러나 중국의 접근 방식에는 몇 가지 해결과제도 존재한다. 먼저, 정부 주도 시스템의 효율성 약화 위험이다. 현재 운영하고 있는 정부 주도의 AI 서비스 안전 평가, 비안 승인 절차에 시간 소요가 길어지거나 불확실성이 높아진다면, 기업의 서비스 출시와 글로벌 경쟁력 확보에 차질을 빚을 수 있다. 둘째, 윤리적 규제 장치의 부족이다. 중국의 AI 법안이 기술적 안전성과 사회적 통제에 초점을 맞추고 있는 반면, 개인의 자율성 존중, 알고리즘의 공정성, 의사결정 과정의 투명성 등 윤리적 측면에 대한 고려는 상대적으로 부족하다. 셋째, 복잡한 규제 체계에 따른 부작용이다. 9개 이상의 정부 부처가 관여하는 다층적 거버넌스 구조는 부처 간 역할과 책임이 불명확할 경우 규제의 중복이나 공백을 초래할 수 있다. 또한 각 부처가 제정하는 세부 규칙이나 지침이 상충될 경우 기업들은 혼란을 겪을 수 있다.

2) 글로벌 AI 거버넌스에 대한 영향

중국의 선제적 AI 규범 제정은 글로벌 AI 거버넌스 경쟁에서 중요한 의미를 갖는다. 세계 최초의 생성형 AI 관리 법안과 체계적인 안전 평가 제도는 개도국을 중심으로 중국 모델의 확산 가능성을 높이고 있다. 많은 개도국들이 AI 거버넌스 체계를 구축하는 과정에서 참고할 수 있는 구체적인 사례를 찾고 있는 상황에서, 중국의 모델은 실질적인 대안으로 인식될 수 있다. 특히 일대일로 국가들과의 디지털 협력을 통해 중국식 AI 거버넌스가 사실상의 국제 표준으로 자리잡을 가능성이 높다.

중국은 '글로벌 인공지능 거버넌스 이니셔티브'를 통해 인간 중심의 AI 발전, 각국의 주권 존중, 포용적 발전이라는 보편적 가치를 강조하면서도, 동시에 서방 국가들의 배타적 그룹 형성

과 기술 봉쇄에 반대하는 입장을 명확히 하고 있다. 이는 미국과 유럽이 주도하는 AI 거버넌스 체계에 대한 대안을 제시하려는 전략적 의도를 담고 있다. 중국의 이러한 움직임은 글로벌 AI 거버넌스가 단일한 모델로 수렴되지 않고 복수의 경쟁적 모델이 공존하는 상황으로 이어질 가능성을 높이고 있다.

또한 중국의 적극적인 국제 표준화 참여는 국제기구에서의 영향력 확대로 이어지고 있다. 2026년까지 20개 이상의 국제 표준 제정에 참여하겠다는 목표는 ITU, ISO 등 국제 표준화 기구에서 중국의 목소리를 강화하려는 의도를 반영한다. 이는 장기적으로 기술 표준의 지정학적 경쟁이 더욱 심화될 것임을 시사한다.