㊾ 선도형 글로벌 AI 안전 거버넌스 구축

AI시대 국민의 안전을 보장하는 것은 대한민국의 존립과 번영을 위한 최우선 과제이다. 대한민국은 강력한 AI 기술을 개발하는 것을 넘어 그 기술의 잠재적 위험을 통제하고 관리하는 'AI 안전' 분야에서 세계를 선도하는 국가가 되어야 한다. AI 안전 기술 경쟁에서의 승리는 곧 글로벌 신뢰 경쟁에서의 승리이며, 이는 우리 국민과 기업을 위한 새로운 경제 및 안전 질서를 여는 초석이 될 것이다. 첨단 AI 기술의 발전은 가짜뉴스, 사이버 공격, 자율 무기 시스템 등 예측 불가능한 위협을 동반하고 있다. 정부는 AI의 잠재적 위험에 선제적으로 대응하기 위해 영국 안전성 정상회의에 참여하고 AI 안전·과학에 대한 국제협력을 위한 의향서를 도출하고, AI 안전연구소를 설립하는 등 국제사회와의 공조를 위한 기반을 마련해왔다. 본 과제에서 제안하는 AI 안전 거버넌스 구축은 이러한 기존 노력의 자연스러운 연장선이다. 이는 파편화된 대응을 하나의 통합된 국가전략으로 발전시켜 국제 논의의 단순 참여자를 넘어 글로벌 AI 안전 규범을 주도하는 리더로 거듭나기 위한 필수적인 단계이다.

○ 정책 권고사항

- 과기정통부는 AI안전연구소와 함께 첨단 AI 안전성 확보를 위한 기술 개발과 전주기 상시 평가체계 마련 등 종합 방안을 '26년 4분기까지 수립한다. AI안전연구소는 미국·영국·일본 등 주요국 AI 안전 연구기관들과 '국제 AI 안전 연구 공동 이니셔티브'를 추진하여 공동 연구와 인력 교류를 활성화하고, 이를 기반으로 '26년 4분기까지 글로벌 규제와 상호운용 가능한 'AI 안전성 프레임워크' 개발을 주도한다.

- 외교부는 유사 입장국과의 AI 안전 거버넌스 논의를 추진하고, 이를 기반으로 OECD, G7, Global Partnership on AI(GPAI) 등 국제 협력체와 연계하여 '26년 4분기까지 AI 안전 글로벌 규범 형성을 주도한다.