ISSUE BRIEF

이슈브리프 480호
(2023.11.10)

바이든 행정부의 첫 인공지능(AI) 행정명령과 시사점

제480호

김경숙 신흥안보연구실
홍건식 신흥안보연구실

국가안보전략연구원
Institute for National Security Strategy

INSS
www.inss.re.kr

국문초록

바이든 행정부의 첫 인공지능(AI) 행정명령은 강력한 규제조치를 담고 있다. 행정명령의 주요 내용은 첫째, AI 안전성 평가 의무화, 둘째, AI 도구의 안전성 표준 마련, 셋째, 콘텐츠인증표준 수립, 넷째, 개인 정보 보호 강화 등이다. 미국이 연방 정부 차원에서 AI 개발과 활용을 안전하고 책임감 있게 촉진하고, 국가 안보, 건강과 안전을 위협하는 AI 기술 개발과 이용을 규제하겠다는 것이 핵심이다. 이 같은 강력한 규제는 AI의 위험에 대한 경각심과 규제의 시급성에 기인한다. AI 개발 기업들은 이번 행정명령에 따라 안전 예방 조치를 해야 하며, 상무부 등 행정부처는 관리 감독을 그리고 미국 클라우드 서비스 제공자는 외국 고객 명단 신고를 의무적으로 해야 한다. 이는 미국 행정부가 전 세계 AI 개발 기업의 정보 수집은 물론 중국을 견제하려는 의도로 보인다. 바이든 행정부의 AI 첫 행정명령은 AI가 가지는 긍정적인 잠재성은 극대화하고 국가 안보, 허위정보 생성, 일자리 등에 미칠 위험성은 최소화하기 위한 규제이다. 따라서, 한국에도 시사하는 바가 크다. 첫째, 개인 정보 보호와 AI 사용의 안전을 강화하기 위한 일련의 신속하고 표적화된 조치를 마련할 필요가 있다. 둘째, 한국의 반도체 산업에 미칠 영향을 점검하고 대비할 필요가 있다. 첨단 반도체 전쟁의 승패가 AI 분야에서 좌우될 것으로 전망되기 때문이다. 셋째, 급물살을 타고 있는 AI 규제와 국제규범 형성 논의에 적극적으로 참여해야 한다. 넷째, 한국의 영향력에 걸맞은 외교력을 발휘해야 한다. 내년 5월 'AI 안전 미니 정상회의'가 한국에서 열릴 예정이므로 우리 정부는 '뉴욕 구상'과 디지털 권리 장전을 기반으로 디지털과 AI 규범 제정 논의를 선도할 필요가 있다. 그 중심은 민주주의 가치, 인간의 생명과 안전 그리고 기본권 보호라는 점을 분명히 할 필요가 있다.

핵심어 : 인공지능(AI), 바이든 행정부, 행정명령, 규제, 안전성, 위험성, 개인 정보 보호, AI 안전 정상회의

바이든 미국 행정부는 인공지능(AI)을 규제하는 첫 행정명령을 발표(2023.10.30)하였다.[1] AI 기술이 급속하게 발전하면서 AI가 주는 편익만큼 부작용에 대한 우려가 커지면서 미국, EU, 중국 등 AI 선도 국가들은 앞다투어 AI 규제에 나서고 있다. 이미 유럽연합(EU)은 전 세계 처음으로 AI 규제법[2]을 채택하였다. 이번 미국의 AI 행정명령은 안전하고 신뢰할 수 있는 AI 개발과 활용을 촉진하기 위해 연방 정부와 기업의 책임을 강화하는 강력한 규제조치를 담고 있다. 이는 위험한 AI 기술이 제도적 통제 없이 개발 및 배포되고 있어 관련 규제가 그만큼 시급하다는 의미이다.

국가 안보와 안전을 위한 강력한 AI 규제조치

바이든 행정부의 첫 AI 행정명령은 국가 안보, 건강, 안전을 위협하는 AI 기술 개발과 이용을 규제하겠다는 것이 핵심이다. 63페이지 분량의 총 8개 부분으로 구성[3]되어 있는 행정명령은 AI 개발 기업의 안전성 평가 의무화, AI 도구의 안전성 표준 마련, 콘텐츠 인증표준 수립과 개인 정보 보호 등에 관한 내용을 담고 있다.

첫째, AI 개발 기업의 안전성 평가 의무화이다. 행정명령은 미국의 안보·건강·안전을 위협할 수 있는 AI 모델에 대해 정부검증 전문가팀(AI 레드팀)의 안전 검사를 받고 AI 개발자는 그 결과를 정부에 제출하도록 의무화했다. 백악관은 '안전성 검증에 대한 정부 보고'를

---

1) The White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, October 30, 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

2) 2021년 4월 AI법안이 발의되었으며, 2023년 6월 EU 의회에서 통과되었다. The European Parliament, EU AI Act: first regulation on artificial intelligence, June 14, 2023, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

3) ▲AI를 위한 새로운 안전 및 보안기준 마련 ▲연방정부의 AI 사용과 조달을 위한 지침 개발 ▲개인정보 보호 ▲평등과 시민권 향상 ▲소비자 보호 ▲노동자 지원 ▲혁신과 경쟁 촉진 ▲국제 파트너와 협력 등이다. The White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, October 30, 2023.

두고 국방물자생산법에 근거한 것으로 첨단 AI 기술이 출시되기 전에 개발·훈련 단계부터 의무적으로 거쳐야 하는 과정이라고 밝혔다. 주목할 것은 마이크로소프트(MS)·구글(Google) 등 미국 기업의 AI 기술을 이용하는 외국인(기업)도 안전성 평가 및 그 결과를 보고해야 하며, AI 훈련도 평가 범위에 두었다. 이는 세계 각국 정부의 AI 규제조치 중 가장 강력한 조치로 볼 수 있다.

둘째, AI 도구의 안전성 표준 마련이다. 행정명령은 국립표준기술연구소(NIST)에 AI 기술의 안전성을 보장하기 위한 최고 수준의 표준 마련을 권고했다. 또한, 행정명령은 에너지부에 핵무기나 생물학무기, 핵심 기반시설, 에너지 안보 분야에서 위협이 될 수 있는 AI 기술을 평가할 수 있는 수단을 개발하도록 했다.

셋째, 콘텐츠 인증 표준 수립이다. 행정명령에 따라 상무부는 AI 기술로 만든 가짜 이미지 등의 콘텐츠 식별을 위해 워터마크 적용을 의무화하는 방안을 추진해야 한다. AI 개발 업체들은 AI를 이용한 거짓 정보 확산을 막기 위해 AI 콘텐츠에 식별 가능한 워터마크를 표시해야 한다. 이는 러시아-우크라이나 전쟁과 최근 하마스의 이스라엘 기습공격에서 나타난 것처럼 AI 기술을 이용한 공격이나 가짜뉴스 확산과 같이 AI를 악용할 위험성을 예방하고 국가안보 위협을 막으려는 조치이다. 바이든 대통령은 행정명령 서명 현장에서 딥페이크 기술의 위험성을 언급하기도 하였다. 이미 구글, 메타, 오픈AI를 비롯한 7개 업체는 AI 기술로 작성한 콘텐츠에 새로운 워터마크를 표시할 시스템 개발에 나서고 있다(월스트리트저널 7.21). 워터마크 표시는 이번 이스라엘-하마스 전쟁에서 하마스의 AI 기술을 악용한 선전전을 막는 데도 활용되었다. 엑스(X), 메타 등 소셜 미디어(SNS) 플랫폼 기업은 AI 생성 가짜뉴스에 식별 표시를 했다.

넷째, 개인 정보 보호이다. ChatGPT[4]는 데이터 수집 출처에 대해 정보를 제공하지 않고, 알고리즘 훈련을 위해 대규모 개인정보를 불법적으로 수집하고 있다. 사용자의 연령 확인에 대한 절차도 없다. 따라서 AI 개발 기업들은 이번 행정명령에 따라 AI 개발 훈련에 개인 정보 불법 사용을 규제하는 지침을 마련해야 한다.

평가 : 양면적인 AI, 선의의 잠재성은 극대화하고 위험성은 규제

바이든 행정부의 AI 첫 행정명령은 AI의 긍정적인 잠재성은 극대화하고 국가 안보, 허위정보 생성, 일자리 등에 미칠 위험성은 최소화하려는 규제로 보인다. 트럼프 행정부가 발표한 행정명령이 미국의 우위를 유지하기 위한 AI 촉진에 초점을 두었다면,[5] 바이든 행정부의 행정명령은 이에 더해 AI의 위험성을 규제하는 데 방점을 두고 있다.

AI 기술이 급속하게 발전하면서 의료, 교육, 교통 등 다양한 분야에서 AI는 효율성과 삶의 질을 높이는 데 활용되고 있다. AI는 신약 개발이나 수술용 로봇, 고령층 보살핌, 방대한 데이터 처리능력, 자율주행 자동차 등 우리 사회에 긍정적 영향을 준다. 무엇보다도 AI 기술은 기술패권 경쟁에서 핵심이다. 최첨단 반도체 분야에서 미국이 격차를 유지하기 위해서는 AI 기술력에서 우위를 점해야 한다. 미국이 신산업에 대한 시장주도와 민간의 역할을 강조한 이유이기도 하다.

그러나, AI가 선의의 잠재성만 있는 것은 아니다. AI를 기반으로 한

4) OpenAI가 개발한 대화형 AI로 딥러닝 기반 모델인 GPT(Generative Pre-trained Transformer) 모델의 변형이다. https://chat.openai.com/c/673b9f41-1ce9-450f-9687-7a2c9bd1a79b

5) The Executive Office of the President, Maintaining American Leadership in Artificial Intelligence, Executive Order 13859, February 11, 2019; The Executive Office of the President, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, Executive Order 13960, December 3, 2020.

개인 정보 유출, 지식재산권 침해, 허위정보 및 가짜뉴스와 사이버 위협 등 AI 악용은 개인과 사회는 물론 국가 안보를 위협하고 있다. AI 전문가들은 악의적인 행위자들이 AI를 악용해 인류에 해악을 끼칠 수 있다고 경고한다.[6] 극단적으로는 AI가 핵전쟁이나 전염병과 유사한 '멸종 위험'을 초래할 수 있다고 우려한다. 'AI의 대부'로 알려진 제프리 힌튼(Geoffrey Hinton) 캐나다 토론토대학 교수를 비롯해 AI 리더들마저 AI가 머지않은 미래에 인류를 대신할 수 있다며 AI의 위험성을 경고하고 있다.[7] 힌튼 교수는 지난 10월 CBS와의 인터뷰에서 AI가 "5년 안에" 인간의 통제 능력을 넘어 진화할 수 있다고 경고했다.[8] 힌튼 교수는 AI의 위험성을 경고하기 위해 10년 동안 겸임한 구글 석학 연구원직을 사임했다. 결국, AI의 미래를 우려하는 전문가들은 안전하고 신뢰할 수 있는 AI를 보장하고 군사 로봇이나 드론 무기와 같은 '이중 용도 기반 AI 모델'의 악용을 규제하기 위한 보호장치가 필요하다는데 뜻을 모으고 있다.

이번 행정명령에 따라 기업들은 안전 예방 조치를 하고, 상무부 등 행정부처는 이를 관리 감독해야 한다. 국방물자생산법에 근거한 AI 모델의 안전성 평가 의무화는 미국 기업의 AI 기술을 이용하는 외국인(기업)과 AI 훈련까지 적용 대상이라는 점에서 AI 관련 가장 강력한 조치이다. 또한, 미국 클라우드 서비스 제공자의 외국 고객 명단 신고 의무화는 미국 행정부가 전 세계 AI 개발 기업의 정보 수집을 쉽게 하고 중국을 견제하려는 의도로 보인다. AI 기술발전을 선도하는 미국이 격차를 유지하기 위해 세계 AI 규제 표준을 만들려는 포석이라는 관측도 나오고 있다.

---

6) Tom Huddleston Jr. 'Godfather of AI,' ex-Google researcher: AI might 'escape control' by rewriting its own code to modify itself, October 11, 2023, https://www.cbsnews.com/news/ai-risk-of-extinction-warning/

7) https://www.cbsnews.com/news/ai-risk-of-extinction-warning/

8) https://www.cnbc.com/2023/10/11/tech-godfather-geoffrey-hinton-ai-could-rewrite-code-escape-control.html

핵심은 AI 기술개발 기업들이 얼마나 행정명령을 이행하는가에 있다. 행정명령은 행정부 내에서 효력이 발생하는 조치라는 점에서 한계가 있을 수밖에 없다. AI 기술개발 기업에 일정 수준의 책임을 부여하고 있지만, 법적 구속력이 없어 벌칙 부과 등을 강제할 방법이 없다. 기술 유출을 우려하는 기업들이 신고를 꺼릴 가능성도 있다.

물론 거대 AI 기업들 역시 AI가 초래할 수 있는 기술적 위험에 대한 규제의 필요성에 대해서는 동의하고 있다. '챗GPT'로 생성형 AI 열풍을 주도하고 있는 오픈 AI의 최고기술책임자(CTO)는 미국 시사 주간지 <타임> 인터뷰(2.5)에서 'AI 기술이 가져올 영향을 고려할 때 모두가 관여하는 것이 중요하다'라는 입장을 피력했다.[9] 이미 일부 AI 기업을 중심으로 안전과 책임성을 확보하기 위한 자율적인 조치도 취하고 있다. OpenAI는 2023년 4월 말, 정보 주체의 권리 보호, 사용자 연령 확인용 기술개발 등 개인 정보 보호를 위한 대책을 마련하였다.[10] 지난 7월에는 구글과 마이크로소프트, 오픈AI, 앤스로픽 등 4개 회사가 AI 안전 표준 개발을 위한 '프런티어 모델 포럼(Frontier Model Forum)'을 결성했다(파이낸셜타임스 7.26).[11] AI 개발 경쟁을 주도하고 있는 기업들의 이러한 선제 행보는 바이든 행정부의 AI 규제 움직임과 무관하지 않다. 동 포럼이 AI 행정명령 조치를 얼마나 이행할지는 지켜볼 필요가 있다.

한국에 주는 시사점

바이든 행정부의 첫 AI 행정명령은 한국에도 시사하는 바가 크다. 행정명령 내용 중 일부는 한국의 기업에 즉각적인 영향을 미칠

9) John Simons, The Creator of ChatGPT Thinks AI Should Be Regulated, February 5, 2023 https://time.com/6252404/mira-murati-chatgpt-openai-interview/

10) 이탈리아 개인정보보호청(GPDP: Garante per la protezione dei dati personali)은 2023년 3월 정부 당국 차원에서는 최초로 ChatGPT의 사용을 제한하는 조치를 실시하였으며, OpenAI의 조치 마련 이후 서비스를 재개하였다. https://now.k2base.re.kr/portal/issue/ovseaIssued/view.do?polilsueId=ISUE_0000000000001045&menuNo=200046

11) https://zdnet.co.kr/view/?no=20230727111331

가능성이 크다. 우리 기업에 미칠 영향을 점검하고 대책을 마련해야 한다. 첫째, 개인 정보 보호와 AI 사용의 안전을 강화하기 위한 일련의 신속하고 표적화된 조치를 마련할 필요가 있다. 미국의 AI 기업 기술을 이용하는 한국 기업은 AI 훈련을 비롯해 미국의 안정성 검증을 받아야 한다. 한국을 포함한 외국 기업이 미국에서 이중 용도 이상의 '범용 AI'를 서비스하려면 90일 이내에 상무부에 보고해야 한다. 따라서, 한미 간 긴밀한 대화 채널을 통해 새롭게 마련될 표준에 대한 정보를 입수할 필요가 있다.

둘째, 한국의 반도체 산업에 미칠 영향을 점검할 필요가 있다. 행정명령은 미국 반도체 산업의 경쟁과 혁신을 촉진하기 위해 AI 기술 강화와 반도체가 AI 경쟁에 중요하다는 점을 언급하고 있다. 첨단 반도체 전쟁의 승패가 AI 분야에서 좌우될 것이라는 전망이 나오는 만큼 이로 인한 여파에도 대비할 필요가 있다.

셋째, AI 규제와 국제규범 형성 논의에 적극적으로 참여해야 한다. 초거대 AI는 국가 경쟁력을 좌우하는 전략기술이다. 초거대 AI는 언어모델 중심 '범용 AI'를 거쳐 알고리즘의 비약적인 연산처리 속도에 기반을 둔 '멀티 모달(Multi Modal) AI'로 진화하고 있다.[12] AI 논의가 윤리 규범 차원을 넘어 규제 움직임으로 급물살을 타고 있는 이유이기도 하다. 생성형 AI의 등장과 급속한 성장은 AI가 가져다주는 효용성과 함께 국가, 사회 그리고 사람에게 미치는 영향력이 고려되면서 규제가 선택이 아니라 시급하게 필요한 시점에 이른 것이다.[13]

넷째, 한국의 영향력에 걸맞은 외교력을 발휘해야 한다. 미중 간

12) 정보통신기획평가원, 2023년 ICT 10대 이슈 (2023.1.30.), p.13.
13) 생성형 AI는 방대한 데이터를 통해 학습하지만, 데이터의 유한성과 데이터 이해가 아니라 패턴을 기반으로 분석하고 결과를 도출하기 때문에 부정확하거나 편향된 응답을 생성하기도 한다. https://now.k2base.re.kr/portal/issue/ovsealssued/view.do?polilsueId=ISUE_000000000001045&menuNo=200046

AI ‘패권경쟁’이 심화되고 있는 가운데 초거대 AI 동맹이 형성될 수 있다는 예측도 나오고 있다. 바이든 행정부의 행정명령 직후 영국에서 첫 ‘AI 안전 정상회의(AI Safety Summit)’가 개최(11. 1~2)되었다. 지난 5월 히로시마 주요 7개국(G7) 정상회의를 계기로 ‘히로시마 AI 프로세스’가 출범하였고, 미국을 비롯한 G7은 AI 안전과 책임을 위한 ‘국제 지침’과 ‘행동 규범’을 발표(10.30)했다. 그 연장선상에서 개최된 첫 AI 안전 정상회의[14]에 한국을 비롯해 미국, EU와 중국 등 28개국이 참석하였다. 참석 국가들은 ‘블레츨리 선언(Bletchley Declaration)’을 발표하고 AI 기술 안전에 관한 협력을 다짐했다.[15] 윤석열 대통령은 이틀째 회의(11. 2)에 화상으로 참여해 글로벌 디지털 규범 정립을 위한 연대와 관련 국제기구 설립 추진을 강조하였다. 내년 5월, 1차 정상회의 후속 논의를 위한 ‘AI 안전 미니 정상회의’가 한국에서 열리는 것은 글로벌 거버넌스 구축을 위한 한국의 노력 성과로 볼 수 있다.

AI 피해와 위험은 국경이 없다는 점에서 AI 위험을 관리하고 통제하는 것은 시급하다. AI 규제와 규범 형성을 위한 국제 논의가 급물살을 타면서 IT 선도국인 한국의 역할은 더 주목받을 것이다. 한국은 윤 대통령이 지난해 9월 선언한 ‘뉴욕 구상’과 올해 9월 발표한 디지털 권리 장전의 5가지 기본 원칙을 기반으로 디지털과 AI 규범 제정 논의를 선도할 필요가 있다. 그 중심은 민주주의 가치, 인간의 생명과 안전 그리고 기본권 보호라는 점을 분명히 할 필요가 있다.

//끝//

본 내용은 집필자 개인의 견해이며,
국가안보전략연구원의 공식입장과는 다를 수 있습니다.

14) https://www.gov.uk/government/topical-events/ai-safety-summit-2023/about
15) The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023, Policy paper published 1 November 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