미연방기관들의 AI 행정명령(E.O. 14110) 이행 현황과 시사점

(U.S. Federal Implementation of Executive Order 14110: Progress and Implications)

신 원 준(Shin Wonjun) · 양 수 빈(Yang Soobin) ·
이 대 준(Lee Daejun) · 박 상 철(Park Sangchul)
서울고등법원 재판연구원 · 김 · 장 법률사무소 변호사 ·
김 · 장 법률사무소 변호사 · 서울대학교 법학전문대학원 부교수

요 지

미 바이든 행정부의 2023. 10. 30. 자 “인공지능의 안전 · 안보 · 신뢰성 있는 개발 · 이용에 관한 행정명령”(E.O. 14110)은 안전 · 안보, 혁신 · 경쟁, 노동, 시민권, 소비자보호, 프라이버시, 공공부문, 국제협력의 분야별로 각 연방기관에 세부 과업을 부여하였고, 한 해가 지난 현시점에 대부분 이행되었다. 이에 연방기관들의 이행 현황을 개관하고 의의를 다음과 같이 정리한다. 첫째, 미국은 전시동원에 준하는 연방 역량의 결집을 통해 전략적 · 산업적 우위를 도모하며, AI의 공급망 문제와 관련하여 수출통제에 준하는 안보적 접근을 명확히 하고 있다. 둘째, EU식 수평 · 포괄규제를 지양하고 각 연방기관에 관할별 세부 목표를 부여하는 부문 · 맥락특유적 규제를 추진하면서도, 예산편성과 연계된 강력한 국무조정과 경성규범과 연성규범 간 유연한 역할 분담을 통해 정연성을 기한다. 셋째, 연방정부의 AI 활용에 대해서는 기관 간 활용 경쟁을 촉발하고 적정한 규제를 통해 인력과 조달 수요를 열고자 한다. 이로부터 몇 가지 시사점을 도출할 수 있다. 첫째, AI에 있어서는 경제와 안보가 분리될 수 없으며 국제정세에 촉각을 세우고 보조를 맞추며 국익을 도모해야 한다. 둘째, 분야별 과잉 · 중복규제를 초래할 EU식 수평 · 포괄 AI 규제보다는 ① 이를 대신할 유연한 연성규범, ② 각 부처의 개별 용례에 대한 핀셋형 · 맞춤형 규제, ③ 예산편성과 연계된 강력한 국무조정의 조합을 통한 치밀하고 협력적인 거버넌스를 추구해야 한다. 셋째, 국내 개발사들이 빅테크의 공세 앞에 숨 쉴 틈을 확보하고 공공부문의 AI 전환을 앞당기기 위해 AI 공공조달 수요를 진작하는 제도 기반을 확충해야 한다.

주제어: 인공지능, 행정명령, 안전성, 거버넌스, 기반모델

◆**논문접수**: 2024. 10. 18. ◆**심사개시**: 2024. 10. 22. ◆**게재확정**: 2024. 11. 15.

미연방기관들의 AI 행정명령(E.O. 14110) 이행 현황과 시사점*

신 원 준** · 양 수 빈*** · 이 대 준**** · 박 상 철*****

Ⅰ. 들어가며

바이든 행정부의 2023. 10. 30. 자 "AI의 안전 · 안보 · 신뢰성 있는 개발 · 이용에 관한 행정명령(Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)"(E.O. 14110)[1]은 안전 · 안보, 혁신 · 경쟁, 노동, 시민권, 소비자보호, 프라이버시, 공공부문, 국제협력 등 정책 목표를 설정하고 이를 달성하기 위해 각 관할 연방기관에 이행 시점을 정해 과업을 부여하는 범정부적(whole-of-government) 기획으로서, 한 해가 지난 현시점에 대부분 이행되었다.[2] E.O. 14110은 트럼프 행정부 2기 출범 후 개폐 가능성이 거론되며,[3] 특히 혁신 · 경쟁 중 출입국 · 이민 관련 사항, 노동, 시민권 정책은 변화가 예상된다. 그러나 E.O. 14110에는 트럼프 행정부 1기의 2019년 "미국의 AI 리더십의 유지" 행정명령(E.O. 13859)[4]과 2020년 "연방정

* https://doi.org/10.22825/juris.2024.1.70.004

** 서울고등법원 재판연구원, 공동1저자(V, VI, VII장 집필). 이 글의 내용은 저자의 소속기관의 공식 의견이 아닌 저자의 개인적 견해임.

*** 김 · 장 법률사무소 변호사, 공동1저자(VIII, IX장 집필). 위와 같음.

**** 김 · 장 법률사무소 변호사, 공동1저자(Ⅲ, Ⅳ장 집필). 위와 같음.

***** 서울대학교 법학전문대학원 부교수. 교신저자(I, II, X, XI장 집필 및 전체 검토). 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. RS-2024-00509258, 글로벌 AI 프런티어랩)을 받아 수행함.

1) 88 Fed. Reg. 75191 (Nov. 1, 2023).

2) 1년간 이행 현황의 간략한 요약은 White House, "Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order", Oct. 30, 2024.

3) 대선을 앞두고 공화당 강령(party platform)은 E.O. 14110을 "급진좌익적 발상(Radical Leftwing ideas)"이라 칭하며 폐지 및 표현의 자유와 번영에 방점을 둔 정책으로의 전환을 공약하였다(43rd Republican National Convention, "The 2024 Republican Platform - Make American Great Again!", Jul. 15, 2024).

4) Maintaining American Leadership in Artificial Intelligence, 84 Fed. Reg. 3967 (Feb. 14, 2019).

부의 신뢰성 있는 AI 활용의 진흥" 행정명령(E.O. 13960)[5] 이래의 미국의 AI의 안전 · 안보 · 신뢰성 및 연방정부 AI 활용 촉진에 대한 일관된 입장이 투영되어 있다. 특히 E.O. 13859과 E.O. 13960를 승계한 ① 안전 · 안보, ② 혁신 · 경쟁 중 출입국 · 이민 외 부분, ③ 공공부문 활용 촉진 정책은 실질상 이어질 것으로 전망된다. 따라서 E.O. 14110을 2022년 말 ChatGPT 출시의 충격에 대한 대응이나 특정 정권의 입장으로 단순화할 수는 없으며, 연방기관들이 이행을 위해 생산한 방대한 행정규칙, 가이드라인, 정책자료들이 향후 미국 연방 AI 규율체계의 출발점이 될 것이다. 다만 E.O. 14110는 연방기관에만 적용되며 주법들은 종종 엇박자를 내고 있음에 유의해야 한다.[6] 우리가 E.O. 14110를 검토해야 할 필요성은 이하로 정리할 수 있다.

첫째, AI 경제안보 질서의 재편 때문이다. 다자간 무역규범의 붕괴, 미-중 무역전쟁 등을 계기로 AI의 국제공급망 내지 국제가치망 재편이 경제안보 문제의 전면에 등장하면서,[7] 미국 AI 경제안보 정책의 면밀한 파악이 절실해졌다. 바이든 행정부의 2023년 "미국의 우려국에 대한 특정 국가안보기술 · 상품 투자의 검토" 행정명령(E.O. 14105)[8]은 특정 국가안보기술 · 상품은 반도체와 마이크로전자, 양자정보기술, AI 분야의 민감 기술 · 상품으로, 우려국은 중국(홍콩 · 마카오 특별행정구 포함)으로 지정하고, 중국의 국가안보기술 · 상품에 대한 위협을 국가비상사태로 선포하였다. 재무부(Department of the Treasury; USDT)가 2024. 11. 15. 이행을 위한 행정규칙을 완결 · 공포하여 2025. 1. 2.부터 미국인들은 상기 기술 · 상품을 중국과 거래할 때 재무부에 신고해야 하고 급박한 안보위협 시 거래가 금지된다.[9] 같은 맥락에서 E.O. 14110은 제4조를 통해 추가적인 안보 조치들을 규정하고 있는바, 이 글 II.장에서 그 이행 상황을 검토한다.

둘째, AI 규제체계 관련 부문특유적(sector-specific), 맥락특유적(context-specific)

---

5) Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, 85 Fed. Reg. 78939 (Dec. 8, 2020).

6) EU AI법의 영향을 받은 포괄적 AI규제법인 콜로라도 AI법(Colo. Rev. Stat. § 6-1-1701 et seq.)(2024. 5. 17. 통과, 2026. 2. 1. 시행)과 후술하는 캘리포니아 AI법안(SB 1047)이 특히 논란의 대상이다.

7) 이재민, "국제경제협정에서 '비차별 원칙'의 공식적 포기인가?-'공급망 재편' 논의의 법적 함의", 국제거래법연구 32권 2호(2023. 12.), 93~109.

8) Addressing United States Investments in Certain National Security Technologies and Products in Countries of Concern, 88 Fed. Reg. 154 (Aug. 11, 2023).

9) USDT, Provisions Pertaining to U.S. Investments in Certain National Security Technologies and Products in Countries of Concern, 31 C.F.R. Part 850, Nov. 15, 2024.

접근을 참고하기 위함이다. 국내에선 “인공지능 발전과 신뢰 기반 조성 등에 관한 기본법안”(“AI기본법안”)이 2024. 11. 26. 소관 상임위를 통과하였는데, 동 법안 중 규제조항은 EU AI법과 같은 수평적(horizontal) 접근방식에 입각해 있다. 수평적 규제체계는 전문규제기관의 전문성을 살릴 수 있는 장점이 있으나 각 분야의 전통적인 규제들과 중복·저촉되고 과잉규제화될 우려가 있다. 반면 E.O. 14110은 각 분야를 전통적으로 담당하던 부처들에 과업을 할당하는 부문·맥락특유적 접근방식을 취하고 있다. 미국의 접근방식의 장점과 한계를 이해한 후 우리 현실에 맞게 절충적인 대안을 도출할 필요가 있다. 이 글 II. ~ IX.장에서 단초를 모색한다.

셋째, 공공부문 AI 활용의 진흥과 관련하여, 빅테크의 압도적인 공세 앞에 국내 개발자들의 생존을 도모하는 전략이 긴요하다. 그러나 EU와 같이 규제로써 장벽을 쌓아 시간이나 벌어보겠다는 수세적 발상은 답이 될 수 없다. AI 공공조달 시장을 열어 국내 개발자들이 숨 쉴 틈을 만들어 주고 행정의 AI화도 가속화하기 위한 제도적 기틀을 시급히 마련해야 한다. E.O. 14110은 제10조에서 이 주제를 다루고 있어, 이 글 IX.장에서 그 이행 상황을 검토하면서, 시사점을 도출한다.

이하에서는 상기 관점을 중심으로 E.O. 14110의 이행 상황을 개관하고, 핵심적 의미를 정리하며, 우리 입법과 정책 방향에 대한 시사점을 도출한다. 단, 이 글의 초점은 E.O. 14110의 이행 현황에 있기에 조항 자체의 검토는 생략한다.[10]

II. 안전과 안보/보안(제4조)

1. 개요

AI 안전(AI safety), 즉 “AI 관련 위해의 예방 또는 그로부터의 보호”[11]는 AI의 개발·혁신에 주도권을 쥔 미·영의 AI 통제에 대한 입장과 공조를 상징하는 개념이다. AI 개발사인 Anthropic이 2023년 초 제안한 이래, 주요 빅테크들이 지지했고, 영국이 수용하여 “AI 안전성 정상회의”의 핵심 의제로 떠올랐으며, 미국이 수용하

10) 조항들의 요약은 Congressional Research Service (CRS), “Highlights of the 2023 Executive Order on Artificial Intelligence for Congress (R47843)”, last updated Apr. 3, 2024.

11) Yoshua Bengio et al., “International Scientific Report on the Safety of Advanced AI: Interim Report”, DSIT 2024/009 (2024), 89.

여 2023. 11. “블레츨리 선언(Bletchley Declaration)”을 통해 주요국들의 지지를 이끌어냈고, 2024. 5. “서울 선언(Seoul Declaration)”에서 재확인되었다. AI 안전 개념에 대한 공통적 이해는 (AI 전반에 대한 선험적 관념에 입각한 포괄규제보다는) 대규모 또는 다량연산으로 훈련된 범용모델(“프런티어 AI” 또는 “기반모델”)에 집중하여 과학적 모델평가(model evaluation), 즉 AI 시스템에 대한 “견고하고 신뢰할 수 있고 재현가능하고 표준화된 평가”[§ 2(a); 이하 E.O. 14110 조항]와 상호운용되는 국제표준에 기해 AI의 오용이 ① 화생방핵(chemical, biological, radiological, and nuclear; CBRN), ② [주요기반시설(critical infrastructure) 등을 겨냥한] 사이버공격, ③ 통제회피 등[§ 3(k)](여기에 생성AI에 의한 여론조작 등 추가)과 관련하여 초래할 수 있는 실재적 · 파국적인 안전 리스크를 측정 · 경감하자는 것이다. 블레츨리 선언의 이행을 위해 각국이 AI 안전연구소(AI Safety Institute; AISI)를 설치 중인데, 그 핵심 역할도 모델평가와 그 근거가 되는 표준의 확립이다.[12] 때문에 미국은 AISI를 상무부(Department of Commerce; DOC) 산하 국립표준기술원(National Institute of Standards and Technology; NIST)에 설치했으며, 행정명령 제4장이 요하는 표준들도 NIST가 준비했다. 그러나 同床異夢이 계속 중이며, 이 개념을 섣불리 법제화한 캘리포니아 프런티어 AI 법안(SB 1047; Safe and Secure Innovation for Frontier Artificial Intelligence Models Act)의 가결 및 주지사의 2024. 9. 29. 자 거부권 행사 등 혼란도 만만찮다.

AI 안보/보안(AI security)은 안보의 측면에서는 앞서 살펴본 AI 공급망 재편 움직임을 반영하고 있고, 보안의 측면에서는 AI 시스템, 특히 주요기반시설의 운용시스템의 “사이버공격, 모델의 코드나 가중치의 유출 등 기술적 개입으로부터의 보호”[13]를 중심으로 한 개념이다.

2. 생성AI와 이중용도 기반모델 등의 평가 · 관리(§§ 4.1, 4.5)

NIST의 주도로 생성AI(generative AI) 및 이중용도 기반모델(dual-use foundation model)과 관련하여 안전 · 안보 · 보안 관련 표준들과 툴을 이하와 같이 발표하였다.

---

12) U.K. Department of Science, Innovation, and Technology, “Emerging Processes for Frontier AI Safety”, Oct. 2023, 15-18.

13) Bengio et al., supra note 11 at 89.

기반모델은 트랜스포머 기반 언어모델 등 범용모델 중 10억 이상 파라미터로 구성된 대규모 모델[14]을, 이중용도 기반모델은 이 중에서도 대량살상무기, 사이버전, 인적 통제 회피 등 전략기술로 전용될 수 있는 범용모델을 지칭한다[§ 3(k)].

가. "AI 리스크 평가 프레임워크: 생성AI 프로파일"(2024. 7. 25.)[15][§ 4.1(a)(i)(A)]

생성AI 프로파일은 NIST가 국가AI이니셔티브법(National Artificial Intelligence Initiative Act of 2020)에 근거하여 2023. 1. 발한 "AI 리스크 평가 프레임워크"(AI RMF 1.0)를 보충한다. 생성AI에 특유하거나 생성AI가 악화시키는 리스크를 ① 화생방핵 정보 · 역량, ② 허언(confabulation)[환각(hallucination)], ③ 위험 · 폭력 · 증오 콘텐트, ④ 데이터 프라이버시, ⑤ 환경영향, ⑥ 유해한 편향 및 동조화, ⑦ 인간-AI 상호작용, ⑧ 정보 무결성, ⑨ 정보보안, ⑩ 지식재산권, ⑪ 아동성착취물(child sexual abuse material; CSAM)과 비동의사적이미지(non-consensual intimate image; NCII)를 포함한 음란 · 비하 · 착취콘텐트, ⑫ 가치망과 구성요소 통합으로 지목하고, 이들을 통제하기 위한 항목을 AI RMF가 규정하는 리스크 평가 프로세스, 즉 거버넌스(Govern) – 맵핑(Map) – 측정(Measure) – 관리(Manage)에 추가한다. 추가 항목만 해도 방대한데, 우리 법에 특히 시사하는 바 큰 ④, ⑩, ⑪ 항목만 발췌하면 이하와 같다.

[표 1] 생성AI 프로파일 추가 항목의 예시

	데이터 프라이버시	(저작권 등) 지식재산권	음란 · 비하 · 착취 콘텐트
거버넌스	• 프라이버시 법령 준수 • 훈련 · 생성데이터 출처 · 이력 문서화 • 허용가능이용정책(acceptable use policy; AUP) • 출처 정보(출처, 서명, 버전관리, 워터마크)	• 지식재산권 법령 준수 • 훈련 · 생성데이터 출처 · 이력 문서화 • 콘텐트 투명성 • 데이터 출처 정보, 지재권 특유의 권리 등 고려 • 약관 제정	• 리스크 계위 현행화 · 정의 • AUP 관련 이해관계자 의견 수렴 • 위법 콘텐트 생성 방지

14) 범용모델(기반모델)과 생성모델의 상이점에 대해서는 박상철, "인공지능의 법적 규율 I: 범용모델과 생성모델", 서울대학교 법학 65권 1호(2024. 3.), 66~69. 이중용도란 상업용 · 군사용 양자에 쓰일 수 있다는 의미로서 바세나르협정 등 다자간 수출통제체제와 미국 수출통제제도의 핵심 개념이다.

15) NIST, "AI RMF: GenAI Profile" (NIST AI 600-1), Jul. 25, 2024.

	• 부정 영향 정보 공유 및 피드백 • 관련 콘텐트 유형 범주화 • 제3자 리스크 평가, 실사, 사고 대응 계획	• 관련 콘텐트 유형 범주화 • 서비스수준약정 • 제3자 리스크 평가, 변경 기록, 실사, AUP, 사고 문서화	• AUP • 약관 제정
맵핑	• 의도된 목적 식별 • 데이터 · 콘텐트 흐름 시험 · 평가 • 주기적 모니터링 • 새 정책 · 절차 · 프로세스의 적용 • 훈련데이터 큐레이션 정책 • 개인정보 노출 고려한 데이터 수집 · 보유 · 최소품질 정책 • 새 도메인에 적응 시 리스크 재평가 • 출력값의 개인(민감)정보 감지 • 훈련데이터 실사	• 의도된 목적 식별 • 데이터 · 콘텐트 흐름 시험 · 평가 • 데이터 정확성 · 대표성 · 관련성 · 적합성 검토 · 문서화 • 침해청구 대응 프로세스 • 훈련데이터 큐레이션 정책 • 불법 사용 고려한 데이터 수집 · 보유 · 최소품질 정책 • 제3자 지재권 · 훈련데이터 사용 · 보유 · 보관 정책 • 적응 작업 시 리스크 재평가 • 훈련데이터 실사	• 기관 리스크 허용 넘는 예견가능한 불법 이용 식별 및 문서화 • 훈련데이터 큐레이션 정책 문서화 • 출처 위해 식별
측정	• 사람 피드백 담당자와 시스템 개발자의 분리 • 출처 데이터 추적 방식 및 데이터의 프라이버시, 보안과의 상호작용 문서화 • 정보주체 참여, 동의 철회권 • 프라이버시향상기술(PET) • Dioptra 등 시험 환경 활용 • 프라이버시 침해 평가 • 보안조치 적용	• 지재권 침해 평가 • 정책 위배, 차단 요청, 지재권 침해 통계 취합 • 레드티밍 • Dioptra 등 시험 환경 활용	• 출처 리스크 측정 신기법 평가 • 부정 영향 평가 • 훈련 데이터 위해 평가
관리	• 합성데이터 등 PET 사용 • 가치망 리스크 시험 • 사전학습 모델의 특정 생성 태스크에 대한 적응 방식 문서화 • 침해신고	• 합성데이터 등 PET 사용 • 제3자 리소스에 리스크 허용 범위 및 통제 적용 • 훈련데이터 검토 • 훈련데이터와 그 출처, 사전학습 모델의 아키텍처와 훈련 프로세스 문서화	• 출력값의 기관 리스크 허용 범위, 지침, 원칙과 비교 • 출력값 분석 • 콘텐트 필터링

나. "합성콘텐트로 인한 리스크 경감"(2024. 4. 초안)[16](§ 4.5)

본 보고서는 생성AI로 생성 · 변경되는 합성콘텐트(synthetic content)가 사칭, 기관의 신뢰성 약화, CSAM/NCII, 사회 분열 심화, 민주정과 선거의 공정성 위협, 안보 등 사회문제를 악화 · 가속화할 수 있다고 지적한다. 이 문제에 기술적으로 대응하기 위한 디지털 콘텐트 투명성(digital content transparency) 메커니즘을 출

16) NIST, "Reducing Risks Posed by Synthetic Content - Draft for Public Comment" (NIST AI 100-4), Apr. 2024.

처데이터추적(provenance data tracking)(메타데이터기록, 디지털워터마킹)과 합성콘텐트탐지(synthetic content detection)(출처데이터탐지, 자동화콘텐트기반탐지, 인간보조탐지)로 구분하여 평가하면서 개발 현황과 한계를 분석하였는데, 특히 이미지 · 오디오 대비 텍스트에 대한 투명성 기법의 한계를 지적하고 있다. 또한, CSAM/NCII 피해 예방을 위한 기술적 조치로서 학습데이터 · 입력데이터 · 이미지출력 필터링, 해싱, 출처 데이터 추적 기술, 레드티밍을 분석하여 한계점을 정리한다. 이어 NIST AI RMF의 체계에 따라 생애주기에 걸친 프로세스를 제시한다.

대통령실 관리예산실(Office of Management and Budget; OMB)이 이에 입각해서 2025. 6. 21.까지 연방정부의 디지털합성콘텐트 레이블링 · 인증 가이드라인을 낼 것이고[§ 4.5(c)], 연방조달규제위원회(Federal Acquisition Regulatory Council; FARC)는 이를 반영하여 연방조달규정(Federal Acquisition Regulation)을 개정할 예정이다[§ 4.5(d)]. 대통령실 과학기술정책실(Office of Science and Technology Policy; OSTP)과 성정책위원회(Gender Policy Council; GPC)는 2024. 5. 23. “이미지 기반 성착취와의 투쟁을 위한 행동개시”를 통해 성착취물의 수익화 방해, 생성의 저지, 배포의 예방, 피해자에 대한 의미 있는 구제수단의 제공을 촉구했고,[17] 2024. 9. 12. 민간 개발자들로부터 자발적인 확약을 받았다.[18]

다. “생성AI와 이중용도 기반모델을 위한 보안성 있는 소프트웨어 개발 프레임워크: SSDF 커뮤니티 프로파일”(2024. 7. 26.)[19][§ 4.1(a)(i)(B)]

본 프로파일은 NIST가 2021년 “국가 사이버보안의 향상” 행정명령(E.O. 14028)[20]에 따라 2022. 2. 발한 “보안성 있는 소프트웨어 개발 프레임워크”(SSDF) 1.1[21]을 보충한다. 소프트웨어 개발 생애주기별(조직의 준비 – 소프트웨어 보호 – 보안성 있는 소프트웨어의 제작 – 취약성에 대한 대응)로 생성AI와 이중용도 기반

17) White House, “A Call to Action to Combat Image-Based Sexual Abuse”, May 23, 2024.
18) White House, “White House Announces New Private Sector Voluntary Commitments to Combat Image-Based Sexual Abuse”, Sep. 12, 2024.
19) NIST, “Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile” (NIST SP 800-218A), Jul. 26, 2024.
20) Improving the Nation’s Cybersecurity, 86 Fed. Reg. 26633 (May 12, 2021).
21) NIST, “Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities” (NIST SP 800-218), Jul. 2024.

모델을 위한 추가 보안조치를 열거하고 있다. 특히 ① AI모델 특유의 취약성과 위협 유형의 예시로 훈련데이터의 오염, 입출력값 내 악성코드 또는 원치 않는 콘텐트, 적대적 프롬프트에 유래한 서비스 거부 상황, 공급망 공격, 미승인 정보 공개, AI 모델 가중치의 도용, 데이터 파이프라인의 오설정을 열거(PW 1.1, R1)하면서, ② 모델가중치와 설정파라미터에 대한 보안조치의 예시로 훈련 · 시험 · 미세조정 · 정렬 · 데이터로부터의 분리, 기밀성(폐쇄모델만)과 무결성의 감시, 접근권한 최소화, 암호화 · 해시 · 디지털서명 · 다자인증 · 망분리(air gap)를 열거(PS1.3, R1-4)한 것이 주목할 만하다.

라. "이중용도 기반모델 오용 리스크 관리" 초안(2024. 7.)[22][§ 4.1(a)(ii)]

NIST 산하 AISI가 초안을 발표하여 의견 수렴 중이다. 이중용도 기반모델의 오용 유형으로 ① 화생방핵 무기 개발 촉진, ② 사이버공격, ③ 기만과 은폐의 조력, ④ CSAM/NCII의 생성을 예시한다. 대응을 위한 오용 리스크 관리 프로세스를 예견, 관리계획 수립, 관리, 측정, 실행 전 관리, 실행 후 정보 수집 및 대응, 투명성의 단계별로 제시한다. 다만 추상적 프로세스 위주라는 한계도 보인다. AISI는 화학생물AI 모델의 안전 관련 고려사항에 대한 의견제출요청을 공고하였다.[23]

마. AI 모델 시험 · 평가 툴 개발[§ 4.1(a)(i)(C), (b)]

NIST는 AI의 위해 유발 역량 평가 · 감사 이니셔티브[§ 4.1(a)(i)(C)]의 일환으로 오픈소스 AI 시험 플랫폼인 다이압트라(Dioptra)를 개발해서 2024. 7. 26. 공개했고,[24] 산하 AISI는 2024. 8. 29. AI 안전 연구 · 시험 · 평가를 위해 Anthropic 및 OpenAI와 협약을 체결하였다.[25] AI 모델 시험 · 평가도구 개발 및 테스트베드 설치[§ 4.1(b)]는 2차 대전 중 맨해튼 계획(Manhattan Project)을 위해 유수의 국립연구소들을 설치하여 보유 중인 에너지부(DOE)에 맡겨져 있다. 산하 6개 연구소[26]에서 AI

22) AISI, NIST, "Managing Misuse Risk for Dual-Use Foundation Models" (NIST AI 800-1; Initial Public Draft), Jul. 2024.

23) NIST, "Safety Considerations for Chemical and/or Biological AI Models", 89 Fed. Reg. 80886, Oct. 4, 2024.

24) James Glasbrenner et al., "Dioptra Test Platform", NIST, 〈https://doi.org/10.18434/mds2-3398〉.

25) NIST, "U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation With Anthropic and OpenAI", Aug. 29, 2024.

모델 평가도구를 개발하고 레드티밍을 수행하는 테스트베드를 설치하였다.[27] 연방 연구소 외 연구자들에 대한 펀딩을 위해 국립과학재단(National Science Foundation; NSF)도 2024. 7. 23. "AI 준비 테스트베드" 개발 이니셔티브를 출범했다.[28]

3. 이중용도 기반모델, 대규모 컴퓨팅 클러스터, IaaS 외국인 거래 보고의무(§ 4.2)

가. 이중용도 기반모델과 대규모 컴퓨팅 클러스터의 보고의무[§ 4.2(a), (b)]

행정명령은 한국전쟁 중 민간의 전시동원을 위해 제정됐고 COVID-19 팬데믹 중 마스크와 식품 공급망 확보를 위해 발동된 국방물자생산법(Defense Production Act of 1950)을 재소환하여, 잠재적 이중용도 기반모델을 개발 중이거나 개발 의도를 표명한 회사들[§ 4.2(a)(i)] 및 잠재적 대규모 컴퓨팅 클러스터를 취득 · 개발 · 보유한 회사 · 개인 · 조직 · 법인들[§ 4.2(a)(ii)]이 상무부에 일정 정보[29]를 주기적으로 보고하도록 하였다. 이에 따라 상무부는 2024. 9. 11. "향상된 AI모델과 컴퓨팅 클러스터에 대한 보고요건의 설정" 행정규칙 개정안을 입법예고 하였다.[30] 핵심 내용은, 보고의무의 이행 주기를 분기로 정하고, 보고의무 기준을 이중용도 기반모델의 경우 훈련 소요 연산량 10^{26} OP[정수나 부동소수점(FLOP)] 이상으로 유지(생물학적 서열 데이터로 훈련된 모델에 대한 10^{23} OP 기준은 별도 검토)하되, 컴퓨팅 클러스터의 경우 네트워크 300 Gbit/s 초과 및 최대용량 10^{20} OP/s 초과로서 데이터 전송 단위만 상향한 것이다.

---

26) Oak Ridge, Argonne, Lawrence Livermore, CENATE, Sandia, NNSA. 앞의 3개 기관은 맨해튼 계획을 위해 각각 레슬리 그로브스, 엔리코 페르미, 어니스트 로렌스가 설치한 연구소들이다.

27) DOE, "Artificial Intelligence Testbeds at DOE", 〈https://www.energy.gov/cet/artificial-intelligence-testbeds-doe〉.

28) NSF, "NSF Announces New AI Test Beds Initiative to Advance Safety and Security of AI Technologies", Jul. 23, 2024.

29) 이중용도 기반모델의 경우 모델 훈련 · 개발 · 제작 관련 계속 · 예정된 활동(물리적, 사이버보안 보호 포함), 모델 가중치의 소유 · 보유 및 이를 보호하기 위해 취한 물리 · 사이버보안 조치, 개발된 모델의 레드팀 테스팅 성능 결과 및 안전 목표 충족을 위해 회사가 취한 조치들을 보고한다[§ 4.2(i)(A), (B), (C)]. 컴퓨팅 클러스터의 경우 존재 · 위치와 클러스터별 총가용연산능력을 보고한다[§ 4.2(ii)].

30) DOC, "Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters" (15 CFR Part 702), 89 Fed. Reg. 73612, Sep. 11, 2024.

나. 클라우드 제공자의 외국인 거래 및 대규모 모델 훈련 보고의무[§ 4.2(c), (d), (e)]

국제긴급경제권한법(International Emergency Economic Powers Act)과 국가비상사태법(National Emergencies Act)은 적국, 테러단체 등에 대한 제재를 규정하며 9.11 테러에 대한 대응, 대북 제재, 러 · 우전쟁 발발 후 러시아 제재 등에 활용되었다. 트럼프 행정부 1기는 이 법들에 의거 2021년 "중대한 악의적 사이버 활동과 관련한 국가비상사태 대응을 위한 추가 조치들의 이행" 행정명령(E.O. 13984)[31]을 발령하여 상무부로 하여금 미국의 클라우드 인프라 서비스(IaaS) 제공자에 대해 외국인 계정 등록유지 시 신원확인을 시행(고객식별 프로그램)하도록 하고 외국인의 악의적 사이버 활동 수행에 사용되는 계정에 대한 접근을 금지하거나 제약을 설정하도록 하였다. 상무부는 2024. 1. 29. E.O. 13984와 E.O. 14110 양자의 이행을 위해 "중대한 악성 사이버 공격 관련 국가비상사태 대응을 위한 추가 조치"라는 행정규칙의 개정안을 입법예고하였는데, 골자는 IaaS 제공자와 재판매업자가 모든 고객의 신원정보를 취득하여 외국인 여부를 확인하고, 잠재적 외국인 고객에 관한 일정 정보[32]를 수집 · 검증 · 보관하며, 외국인 고객이 "악의적 사이버 활동에 사용될 수 있는 잠재적 역량을 가진 대규모 AI 모델"을 훈련할 경우 보고하고, 필요시 접근을 금지하거나 제약을 설정하도록 하는 것이다.[33]

4. 보안(§ 4.3)

깃헙 코파일럿(Github Copilot)으로 대표되는 언어모델 기반 코드생성모델이 악성코드 생성 툴로 무기화되거나 보안이 취약한 코드 베이스로 훈련되어 취약성을 증폭할 수 있다는 점[34]이 보고되는 등 특히 주요기반시설에 대한 AI의 사이버 공격 오용 내지 취약성 증폭에 대한 우려가 커지고 있다. 국토안보부는 국방부와 AI를 활용해서 미 정부 네트워크의 취약성을 보완하는 시범 프로그램을 추진하여

31) 86 Fed. Reg. 6837 (Jan. 25, 2021).
32) 계정번호, 주소, 이메일, 지불수단, 전화번호, IP주소.
33) DOC, "Taking Additional Steps to Address the National Emergency with Respect to Significant Malicious Cyber-Enabled Activities" (15 CFR Part 7), 89 Fed. Reg. 5698, Jan. 29, 2024.
34) Hammond Pearce, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, and Ramesh Karri, "Asleep at the Keyboard? Assessing the Security of GitHub Copilot's Code Contributions", 2022 IEEE Symposium on Security and Privacy (2022).

결과를 대통령실 안보실장에게 보고[§ 4.3(b)]하였으나 공개되지 않았다. 이하에서는 주요기반시설 보안[§ 4.3(a)]의 이행 현황을 집중적으로 살펴본다.

행정명령[§ 4.3(a)(i)]의 이행을 위해 16개 주요기반시설을 관장하는 부문 리스크 관리부처들(Sector Risk Management Agencies; SRMAs)[35]이 부문별 AI 리스크 평가를 실시했고, 국토안보부(Department of Homeland Security; DHS) 산하 사이버보안 · 기반시설보안청(Cybersecurity and Infrastructure Security Agency; CISA)이 분석 결과를 취합하여 150건을 넘는 AI 용례를 식별하고 10개로 유형화(빈도순으로 작동감시, 성능최적화, 작동자동화, 사건감지, 예측, 연구개발, 시스템계획, 고객서비스 자동화, 모델링 · 시뮬레이션, 물리적 보안)하는 한편, 이들 유형별 리스크를 다시 ① AI 사용 공격(사이버공격, 자동화된 물리공격, 물리적 타겟 및 취약성 식별, 사회공학, 공급망 교란, 지식재산권 편취 및 역공정, 무기 개발), ② AI 시스템 겨냥 공격(AI 알고리듬 또는 데이터의 적대적 조작, 회피공격, 서비스중단 공격, 데이터 손실, 모델 전도 및 추출), ③ AI 설계 · 이행상 실패[자율성, 불안정성(brittleness), 불가해성(inscrutability), 의도치 않은 시스템적 · 설계적 결함, 일관되지 못한 시스템 유지보수, AI 시스템과 비AI 시스템 간 상호운용과 설정, AI에 대한 과다 · 과소의존, 통계적 편향, 부문 전문가 부족, 공급망 취약성]로 유형화하여 분석하였다.[36] 국토안보부는 이에 입각하여 2024. 4. 26. 자 "주요기반시설 보유 · 운영자를 위한 안전 · 보안 가이드라인"을 발표하였는데, 위와 같이 식별된 리스크에 대한 감경전략을 열거하는 한편, NIST RMF의 체계에 맞춰 주요기반시설 해당 거버넌스－맵핑－측정－관리 항목을 구체화한 가이드라인을 제시하였다[§ 4.3(a)(iii)].[37] 대통령실 국가안보보좌관(Assistant to the President for National Security Affairs; APNSA)과 관리예산실 주도로 2024. 12. 23.까지 기관들이 이 가이드라인을 의무화하도록 조치를 취하도록 되어 있다[§ 4.3(a)(iv)]. 국토안보부는 2024. 4. 29. AI 안전 · 안보위(AI Safety and Security Board; AISSB)를 설

35) 16개 부문 중 국토안보부는 화학, 상업시설, 통신, 주요제조설비, 댐, 긴급서비스, 정보기술, 원자로 · 핵물질 · 폐기물, 국토안보부와 교통부는 교통시스템, 국방부는 방위산업기반, 에너지부는 에너지, 재무부는 금융서비스, 농무부와 보건복지부는 정부설비, 보건복지부는 보건 · 공중보건, 환경보호청은 상하수도 시스템을 관장한다.

36) DHS, "Safety and Security Guidelines for Critical Infrastructure Owners and Operators", Apr. 26, 2024, 7-9, 16-21.

37) Id., 22-28.

치하였다[§ 4.3(a)(v)].[38]

특히 금융 보안[§ 4.3(a)(ii)]에 관하여 재무부는 2024. 3. "금융서비스 부문에서의 AI 특유의 사이버보안 리스크의 관리"를 발간하였다.[39] 이 보고서는 AI의 발전이 금융기관의 사기탐지 및 사이버보안 강화에 기여했으나, 새로운 보안 위협(데이터 오염, 추론을 통한 데이터 유출, 회피, 모델 추출), 생성AI를 통한 금융사기, 제3자 벤더 관련 리스크(데이터 무결성과 출처)를 야기했음을 지적한다.[40] 대응을 위한 모범사례로 ① AI 리스크 관리의 기존 리스크 관리 프로그램으로의 통합, ② AI 리스크 관리 프레임워크의 개발 · 이행, ③ AI를 위한 리스크 관리 기능의 통합, ④ 최고데이터책임자(chief data officer; CDO) 역할의 진화와 데이터 공급망의 맵핑, ⑤ 벤더의 감독, ⑥ NIST의 사이버보안 프레임워크의 검토, ⑦ 리스크에 기반하여 차등화된 다중인증의 이행, ⑧ 직무와 리스크 수인에 대한 올바른 툴의 선정, ⑨ AI 시스템에 대한 사이버보안 모범규준의 긴밀한 적용을 열거하면서,[41] 특히 금융기관들이 사이버사기에 공동 대응하기 위한 정보 공유 체계의 확립을 강조하고 있다.[42]

5. 화생방핵(CBRN)(§ 4.4)

화생방핵 중 방사능과 핵은 핵연료물질 등 방사성물질에 대한 접근통제라는 진입장벽이 여전히 존재하나, AI 기반 생물 · 화학툴이 대중화되면서 큰 위협이 되고 있다. 아직까진 범용의 대규모 언어모델보다는 알파폴드(AlphaFold) 등 생물학적 설계 툴(biological design tool; BDT)이 단백질과 염기서열 등의 합성에 주로 사용된다.[43] 특히 오픈소스 BDT의 출현으로 진입장벽이 대폭 낮아졌으며, 2022년 그 중 하나인 MegaSyn AI가 6시간도 안 되어 4만 건의 독소를 생성하는 등 우려가 커지고 있다.[44] 미국도 생물화학무기(chemical and biological weapon; CBW) 합성

38) DHS, "Establishment of the Artificial Intelligence Safety and Security Board", 89 Fed. Reg. 33380, Apr. 29, 2024.
39) USDT, "Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector", Mar. 2024.
40) Id., 12-20.
41) Id., 26-32.
42) Id., 3-4, 34-35.
43) National Security Commission on Emerging Biotechnology (NSCEB), "White Paper 1: Introduction to AI and Biotech", Jan. 2024.
44) Fabio Urbina, Filippa Lentzos, Cédric Invernizzi, and Sean Ekins, "Dual Use of Artificial-in-

리스크에 집중해서 대책을 강구하고 있다.

가. 국토안보부의 2024. 4. 26. 자 "AI와 화생방핵 위협의 접점에서의 리스크의 경감에 관한 보고서"[§ 4.4(a)(i)][45]

AI의 화생방핵 오용 관련 쟁점을 정리하고 대응방안을 권고한다.

[표 2] AI의 화생방핵 활용 관련 쟁점과 권고사항

쟁점	권고사항
안보 · 공중보건 · 동물 보건 기관 간 AI 리스크 공통이해 형성	• 정보 공유 · 보고 · 대응 체계, 국가생물방위전략이행계획과 국가안보각서상 리스크 평가에 AI 화생방핵 문제 편입 • AI 화생방핵 용례 맵핑, 기관 간 공유, 교육
모델과 데이터는 주로 민간 · 학술기관이 보유. 오픈소스 모델이 모델과 BDT에 대한 접근의 대중화 초래	• 범용모델 소스코드 · 모델가중치 배포 지침 • AI 툴 개발자의 취약성 신고 경로 마련 • 공개데이터로부터 민감 생물/화학 데이터를 배제/보호하는 기준 • 고위험 생물/화학 툴 · 서비스(BDT, 화학적 역합성 툴, 핵산합성 툴 등) 접근 차등화 및 고객 확인 • 민감 모델가중치의 역공정 · 손실 · 유출 방지 위한 보안 강화
AI로 인해 화생방핵 공격을 개념화 · 실행하는 악의적 행위자의 능력 향상	• 과학계 책임 문화 강화, 훈련 기준 정비 • 위험 물질 필수 스크리닝 모범규준 • AI로 자동화된 실험실에 대한 인적 감독 보장 지침 • 국제사회와 협력하여 AI 안전 및 보안 원칙 마련
개발자들의 자발적 리스크 관리는 한계, 산 · 관 · 학 간, 개발자와 국가안보 · 생물국방계 간 교류 필요	• AI 악용을 막기 위한 정책, 규범, 행동강령 • CBW 관련 훈련 데이터셋의 평가 기준 • 모델 검토자와 레드팀을 위한 AI 화생방핵 위협 인식 훈련 • AI 모델 사전 평가 및 레드티밍과 사후 위해 보고 시스템 • 평가 벤치마크 또는 표준 • CBW 위협 보호 위한 모델 역량 레드티밍 이행 프레임워크 • 이중용도 AI 기반 생명공학 규제를 위한 법적 분석 및 규제 개선
기존 보안 규정의 한계와 AI 도구의 사용 증가에 따라 위험한 연구 결과의 안보 위협 증대	• 생명공학 기술 · 시스템 인식 · 감시 강화 • 연구소 생물보안 지침, 펀딩조건, 규정 검토 · 개정 • 이중용도 연구 관련 정부지원 · 교육 · 내부위협훈련지침 개선 • 고도격리실험실의 이해, 재고안전관리 개선, 리스크 평가 • 화학시설 테러방지 기준 개정

telligence-powered Drug Discovery", 4 Nat. Mach. Intell. 189-191 (2022).

45) DHS, "Department of Homeland Security Report on Reducing the Risks at the Intersection of Artificial Intelligence and Chemical, Biological, Radiological, and Nuclear Threats", Apr. 26, 2024.

	• 생물/화학 안전 인식 확산 • 과학 커뮤니티와 과학, 윤리, 거버넌스 문제 논의 • 안전 · 보안 원칙 · 조치 관련 국제협력
국제적 이해관계자들과 협력하여 원칙 및 프레임워크 개발	• 군비통제, 비확산 등 국제협약 준수를 확인하고 도전과제 · 기회 논의 • 미국의 AI 기술과 데이터셋의 국제 공유 지침 • 동맹과 함께 AI 생물/화학 보안 표준, 프레임워크, 레드티밍 개발 • 국제 사회와의 정보 공유
AI를 화생방핵 예방 · 탐지 · 대응 · 경감에 투입	• 대량살상무기, 테러, 화생방핵 대응 계획에 AI 통합 • CBW 대비 개인보호구 · 의료대책 · 제독제의 설계 · 시험 · 평가에 AI 활용 • AI안전 · 안보위를 통해 정보 공유, 모범규준 및 리스크 경감 확립
AI로써 국가생물방어전략 강화	• 질병 발발, 비정상적 생물/화학 사건 탐지 • 질병 발발 대응 리소스 감시 · 추적 · 점검 · 실행 • AI 화생방핵 리스크 평가, 위협 모델링, 의사결정 지원
AI 툴을 화생방핵 관련 국제협력, 생물/화학 공격 추적, 국제 협정 준수의 감시에 활용	• 국제연합(UN)과 화학무기금지기구(OPCW)의 조사 AI 활용 지원 • CBW 샘플 서명 식별, 테러리스트 네트워크 맵핑 · 분석, CBW 주문 스크리닝 AI 툴 개발 • CBW 이정표 · 표지자 식별, 조달망 분석, 취약성 식별 AI 앱 연구 · 조달 • 미지 · 신규 생물/화학 구조 포렌식 역량 확충, 포렌식 국제표준 지원

나. 국가과학기술위원회(National Science and Technology Council; NSTC) 산하 합성핵산 구매 스크리닝 패스트트랙 조치 위원회(Fast Track Action Committee on Synthetic Nucleic Acid Procurement Screening)의 2024. 4.(2024. 9. 개정) "핵산 합성 스크리닝 프레임워크"[§ 4.4(b)][46]

이 프레임워크는 보건복지부(Department of Health & Human Services; HHS)의 2023. 10. 자 "합성핵산 제공 · 이용자를 위한 스크리닝 프레임워크 가이드라인"[47]을 보충하며, ① 동 프레임워크의 실행을 증명하는 방침을 웹사이트에 게시하거나 연방기금 기관 · 수혜자 요청 시 제공, ② 합성핵산의 구매주문(P/O)을 스크리닝하여 우려서열(sequence of concern; SOC)을 식별, ③ SOC 포함 합성핵산

46) NSTC, "Framework for Nucleic Acid Synthesis Screening", Apr. 29, 2024.

47) HHS, "Screening Framework Guidance for Providers and Users of Synthetic Nucleic Acids", Oct. 2023.

과 벤치탑(benchtop)[48] 핵산합성장비의 구매자를 스크리닝하여 합법성을 확인, ④ SOC 관련 합성핵산, 벤치탑 핵산합성장비 P/O가 위법 소지가 있으면 보고, ⑤ 합성핵산, 벤치탑 핵산합성장비 P/O 기록 보관, ⑥ 사이버・정보보안 보장 조치 등 합성핵산염기서열 제공자의 합성핵산 구매 스크리닝 절차를 규정하고 있다.

다. 전미과학공학의학원(National Academies of Sciences, Engineering, and Medicine; NASEM)의 생명공학 AI 활용의 안보 우려와 편익 연구[§ 4.4(a)(ii)]

국방부(Department of Defense; DOD)로부터 연구를 위탁받아 진행 중이다.[49]

6. 공개모델(§ 4.6)

행정명령은 "널리 가용한 모델 가중치(widely available model weights)를 가진 이중용도 기반모델", 즉 Llama 등 오픈소스로 배포된 대규모 범용모델의 편익과 리스크에 관한 의견 수렴 절차를 거쳐 상무부가 보고서를 작성하게 하였다.[50] 오픈소스 모델이 경쟁을 촉진하고 AI의 활용도를 높이나, 수출통제에 준하는 안보적 조치를 무력화하고, 무기 전용에 대한 진입장벽을 낮추는 리스크를 검토하자는 취지이다. 그러나 최종결과물인 상무부의 2024. 7. 30. 자 "널리 가용한 모델 가중치를 가진 이중용도 기반모델" 보고서는 공개모델의 제한이 당장 필요하다거나 미래에도 부적절하다고 단언할 증거는 현재로서 불충분하다고 결론 내리면서, ① 표준과 감사・공개・투명성 요구, 연구 지원・수행, 외부 연구 지원, 리스크 포트폴리오, 지표, 하한 개발을 통해 증거를 수집하고, ② 전유모델과 공개모델 간 시차 평가, 벤치마크와 정의 개발, 연방정부 전문가 역량 유지를 통해 증거를 평가하며, ③ 모델 접근 제한 등 리스크 경감조치를 통해 평가를 시행하고, ④ 추가적 조치 가능성을 열어놓자고 제안하는 등 일단 지켜보자는 입장인데,[51] 오픈소스 대규모 언어모델에 의존 중인 우리 스타트업들은 한 시름 놓게 되었다.

48) 실험실에서 간편하게 사용될 수 있는 장비를 의미.
49) NASEM, "Assessing and Navigating Biosecurity Concerns and Benefits of Artificial Intelligence Use in the Life Sciences - August Information Gathering Meeting", Aug. 13-14, 2024.
50) DOC, "Dual-Use Foundation Models with Widely Available Model Weights", Jul. 2024.
51) Id.

7. 연방데이터의 안전한 공개 및 악의적 AI 학습 활용 방지(§ 4.7)

미국은 증거기반정책수립기본법(Foundation for Evidence-Based Policymaking Act of 2018)의 Title II인 공개 · 공공 · 전자 · 필수정부데이터법(Open, Public, Electronic, and Necessary Government Data Act)에 따라 연방정부 데이터를 공개한다. 각 부처의 최고데이터책임자(CDO)의 협의체인 연방총무청(GSA) 산하 최고데이터책임자위원회(CDO Council)가 이 공개가 화생방핵 무기나 자율사이버공격 개발 오용에 초래할 리스크를 검토하여 지침을 작성할 예정이다[§ 4.7(a)]. 이에 따라 각 기관들이 보안 검토를 수행하고 최우선순위 잠재 보안 리스크에 대응할 것이다[§ 4.7(b)].

8. AI 국가안보각서(National Security Memorandum; NSM)(§ 4.8)[52]

대통령실이 2024. 10. 24. 발표한 AI 국가안보각서는 정보기관들과 AI를 국가안보시스템의 구성요소로 활용하는 기관들[대통령실, 의회 정부책임처(Government Accountability Office; GAO), 연방선거위원회(Federal Election Commission; FEC) 제외]에 적용되며, 간략히만 요약하면 다음과 같다.

- AISI의 프런티어 AI 안전성 시험 및 선제적인 AI 시험 인프라: AISI가 사이버공격, CBW, 자율적인 악성 행동의 실행, 이러한 능력을 가진 다른 모델의 개발 · 실행의 자동화 등 리스크를 중심으로 프런티어 AI의 안전성을 시험하고, 리스크 및 경감조치에 대해 국가안보보좌관에 보고하도록 한다. 단, 원자력은 에너지부가 담당한다.
- 각 기관의 사이버, 화생방핵 리스크 관련 부문특유적 AI 시험: 사이버 · 핵 · 방사능 리스크와 관련하여 국가안보국(National Security Agency; NSA)의 AI보안센터(AI Security Center; AISC)는 AISI와 협업해서 AI 모델의 사이버공격의 적발 · 생성 · 가중 역량에 대해 신속하고 시스템적이며 기밀분류된 평가를 수행하고, 에너지부는 AISI 및 국가안보국과 협업해서 핵 · 방사능에 유사한 평가를 실시한다. 생물 · 화학적 리스크에 대해서도 기밀분류된 평가를 시행하

52) White House, "Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence", Oct. 24, 2024.

며, 에너지부, 국토안보부, AISI, 국방부 등이 로드맵을 개발하고, 에너지부가 시범사업을 시작한다. AISI가 생물 · 화학 안전 지침을 발표하면 실질적으로 생물 · 화학 데이터로 훈련된 공개 기반모델을 개발하는 모든 기관이 준수해야 한다. 나아가 국방부 · 보건복지부 · 에너지부 · 국토안보부 · 국립과학재단은 실질적으로 생물 · 화학 데이터로 기반모델 훈련 시 생물안전 · 보안 우선 조치들(가상연구 평가 툴 개발, 합성 핵산 감시 · 스크리닝 알고리듬 개발, 보안 · 신뢰성 있는 소프트웨어 프레임워크 개발, 클라우드 랩과 바이오 제조설비의 데이터 스트림과 주문의 스크리닝, 의료적 대응 등 리스크 감경 전략 개발)을 시행해야 한다. 국립과학재단은 연구기관들과 과학저널들을 소집해서 연산적 생물 · 화학 모델 · 데이터셋 · 접근법의 간행에 대한 자율모범관행 · 표준을 개발시킬 예정이다.

- 국가안보 AI 리스크 관리 프레임워크: 대통령실은 같은 날 이 프레임워크를 별도 문서로 발표했는데,[53] 후술하는 관리예산실의 M-24-10을 보충하며, 4개 필러(pillar)로 구성된다. 첫째, 금지 대상 AI 용례를 ① 개인의 권리행사에만 근거한 활동들(표현 · 집회 · 결사 등)의 프로파일링 · 타겟팅 · 추적, ② 표현의 자유, 변호인의 조력을 받을 권리의 불법적 억압 · 제약, ③ 불법적 차별, ④ 데이터로부터 개인의 감정 상태를 감지 · 측정 · 추론(공무원 건강 지원 목적으로 동의하에 진행시 예외), ⑤ 생체인식정보에만 의존하여 민감정보를 추론 · 결정, ⑥ 무력작전 전 비전투원 존재 식별을 포함하여 치명적 피해와 사상자 예측(단, 견고한 시험과 보증, 합당한 판단과 주의, 예측에 책임을 지는 훈련된 인원의 감시의 경우 예외), ⑦ 이민자 분류 최종 결정, ⑧ 충분한 주의 없이 AI로만 생성된 보고서나 정보분석을 생성 · 유포, ⑨ 대통령의 핵무기 사용 개시 · 중단 결정의 고지 · 실행 조치에서 인적 관여 제거로 정한다. 고영향 AI 용례는 ① 군사 또는 형사사법 목적으로 생체인식정보에만 기하여 실시간으로 개인을 트래킹 · 식별, ② 개인을 테러리스트, 내부위협자, 안보위협자로 분류, ③ 이민자 분류 결정, ④ 이민 · 출입국 관련 권리 · 안전영향 활동, ⑤ 무기화 리스크가 있는 민감한 화생방핵 관련 활동, ⑥ 오로지 완결된 정보분석을

53) White House, “Framework to Advance AI Governance and Risk Management in National Security”, Oct. 24, 2024.

생성 · 유포할 목적의 AI 활용으로 정한다. 둘째, 고영향 AI에 대해 최소 리스크관리 실무를 정하며, ① AI 리스크영향평가, ② 시험, ③ 평가 및 AI 거버넌스 위원회 보고, ④ 차별 · 편향 기여요소의 식별 · 경감, ⑤ AI 과의존 리스크 경감, ⑥ AI 운용자 훈련 · 평가, ⑦ AI 기반 결정의 인적 고려 및 감독, ⑧ 보고체계, ⑨ 정기 모니터링 및 시험, ⑩ 주기적 인적 검토, ⑪ 신규 리스크 경감, ⑫ 내부보고체계를 포함한다. 셋째, 고영향 AI 사용을 목록화 · 모니터링한다. 넷째, 효과적인 훈련 및 책임성 메커니즘을 제시한다.

- 국가안보 목적의 AI의 획득 · 조달: AI의 효율성 · 책임성 있는 활용을 촉진하고 기관 간 AI 기능의 공유와 상호운용성을 확보하도록 하며, 각 기관이 국가안보시스템에 대한 거버넌스 및 리스크 관리 지침을 발하도록 한다.
- 외국의 정보 위협으로부터의 보호: 외국이 “미국의 AI 혁신에 대한 과실을 자신의 안보 목표를 위해 취득하고 전용”하고자 (전쟁도 평화도 아닌) 회색지대(gray-zone) 방식을 통해 미국의 AI 지식재산(“핵심 기술 유물(critical technical artifacts)”)을 빼내고 있다고 지적하면서, 국가정보장실(Office of the Director of National Intelligence; ODNI)이 AI 공급망에서 유출이 이뤄지는 핵심 지점과 리스크를 식별하고, 외국인투자위원회(Committee on Foreign Investment in the U.S.; CFIUS)가 외국인투자 심사 시 AI 훈련 기법 등 전유 기술 접근 여부를 검토하도록 한다.
- 기타: 그 밖에 안보기관 내 AI 인재 임용을 늘리고, 국립과학재단의 국가AI연구자원(National AI Research Resource; NAIRR) 시범사업을 중심으로 AI 반도체 및 연산 인프라를 촉진하며, AI를 신속하게 개발하여 안보 목적에 활용하고, 동맹 및 파트너들과 공동 개발 · 실행할 것을 규정하고 있다.

Ⅲ. 혁신과 경쟁(제5조)

1. 개요

행정명령은 자유시장 원칙을 바탕으로 하면서도 혁신을 촉진하고 중국 등 주요국들과의 AI 패권 경쟁에서 우위를 점하기 위한 연방정부의 적극적 역할, 소규모 개발자 등에 대한 제도적 지원, 해외 AI 인재 유치를 강조한다. 그 실현을 위해 인재

유치, 연구개발강화, 지식재산권 보호를 중심으로 한 다각적 전략을 추진하였다.

2. 해외 AI 인재의 미국 유치(§ 5.1)

해외 AI 인재 유치를 위해 각 부처가 비자 및 취업허가 완화 조치를 시행했다.[54] 다만 이 정책은 트럼프 행정부 2기에서 상당 부분 후퇴할 것으로 예상된다.

3. 연구개발 진흥[§ 5.2(a)~(b), (e)~(h)]

부처별로 다각적인 국가연구개발 정책이 전개되고 있다.[55]

---

54) 각 부처가 취한 조치들은 이하와 같다.
- 비자 등록·갱신 절차의 간소화 등 개선[§ 5.1(a), (b)]: 국토안보부(DHS) 연방이민국(USCIS)은 2024. 2. 2. 비이민 취업비자인 H-1B의 등록절차를 개선(건별 추첨 대신 인별 추첨, H-1B 상한 적용 시 시작일자의 유연화 등)했고[8 CFR Part 214 (2024)], 동시에 입법예고된 다른 개선사항들도 추가 시행할 예정이다. 국토안보부도 이공계(STEM) 분야 F-1 비자의 졸업 후 실무수습(OPT) 기간 연장 등의 추가 개선 계획을 밝히고 있다(DHS, "Artificial Intelligence Roadmap", Mar. 17, 2024).
- 비자 국내 갱신 재개[§ 5.1(c)]: 국무부(DOS)는 H-1B의 국내 갱신 재개를 위한 시범사업을 2024. 1. 29.부터 4. 1.까지 시행하였다. 9·11 테러 후속조치로 2004년부터 비자 발급 시 지문을 날인받았는데, 국내에 지문 날인 시설이 없어 국내 비자 갱신을 중단했었던 것을, 20여 년 만에 재개한 것이다.
- 비자 자격요건 정비[§ 5.1(d)]: USCIS는 행정명령 발령 전인 2023. 9. 12. 이미 EB-1의 자격요건에 AI 관련 항목을 추가했다.
- 취업허가 개선[§ 5.1(e)]: 노동부도 정규직 취업허가 절차(Schedule A of 20 C.F.R. § 656.5)의 현대화를 위한 입법예고(99 Fed. Reg. 11788)를 하였다.
- 비자 정보 제공[§ 5.1(g)]: 국토안보부는 외국 AI 인재를 위한 비자 절차를 웹사이트(ai.gov)에서 안내하고, 2024. 5. 9. EB-2와 O-1A 관련 통계를 제공하고 있다.

55) 각 부처별 국가연구개발 프로젝트들은 이하와 같다.
- 국립과학재단과 에너지부는 국가AI연구자원(NAIRR) 시범사업으로 신뢰 가능 AI, 건강, 인프라 관련 AI 연구를 지원한다[§ 5.2(a)(i)](NSF, "National Artificial Intelligence Research Resource Pilot", Jan. 24, 2024).
- 국립과학재단의 지역혁신엔진(NSF Engines) 프로그램은 지역 기반 기술주도적, 포괄적 혁신 자금 지원 프로그램으로서, 반도체과학법(CHIPS and Science Act of 2022)상 제한적이었던 지원 대상도 확대하였다[§ 5.2(a)(ii)](NSF, "NSF Engines Eligibility Update", Sep. 18, 2024).
- 국립과학재단은 천문학, 재료발견 연구, AI 강화의 3가지 주제로 5개 국가AI연구소를 선정하여 총 1억 달러의 펀딩을 할 예정이며[§ 5.2(a)(iii)][NSF, "National Artificial Intelligence (AI) Research Institutes" (NSF 23-610), Aug. 1, 2023], 이 중 2개 천문 AI 연구소를 출범했다.
- 에너지부는 NAIRR 프로그램과 관련하여 딥페이크 탐지, 차세대 의료 진단 촉진 등 주요 AI 과제들을 위한 고성능 컴퓨팅 자원을 제공하는 35개 프로젝트를 지원함과 동시에 고성능 컴퓨팅 시스템, 클라우드 컴퓨팅 플랫폼, 기초 모델, 소프트웨어 및 프라이버시 향상 기술 툴, 모델 교육을 위한 교육 플랫폼을 제공한다[§ 5.2(b)](NSF, "NSF-led National AI Research Resource Pilot Awards First Round Access to 35 Projects in Partnership with DOE", May 6, 2024).
- 보건복지부는 책임 있는 AI를 위한 데이터 품질 등에 초점을 둔 "2024 LEAP in Health IT" 지원자

4. 특허 등 지식재산권

가. 특허상표청(U.S. Patent and Trademark Office; USPTO)의 AI 지원 발명(AI-assisted invention) 관련 AI의 발명자 적격(inventorship) 검토[§ 5.2(c)(i), (ii)]

특허상표청의 2024. 2. 13. 자 "AI 지원 발명에 대한 발명자 적격 지침"[§ 5.2(c)(i)][56]은 ① DABUS 판결[57]에 따라 발명자란 용어는 법적으로 자연인을 의미할 뿐 AI 시스템은 발명자 적격이 없음, ② AI 지원 발명의 경우 AI가 발명에 기여했더라도 자연인만이 발명자로 기재될 수 있음, ③ AI 시스템은 공동발명자로도 지정될 수 없음을 명확히 하고 있다. 한편 AI 지원 발명에 있어서 특정 자연인이 발명자가 될 만한 기여를 했는지 판단 기준으로 Pannu 요소[58]를 적용하여 ① 특정 문제 해결을 위한 프롬프트 구성 방식을 직접 고안하였는지, ② AI 시스템의 산출물을 실험적 과정에 적용하여 활용하였는지, ③ AI 시스템의 설계 · 구축 · 훈련에 직접 참여하

를 공모 · 선정했다[§ 5.2(e)](HHS, "HHS Announces 2024 LEAP in Health IT Awardees Focused on Data Quality for Responsible AI and Accelerating Adoption of Behavioral Health IT", Sep. 17, 2024).

- 보훈부(Department of Veterans Affairs; VA)는 재향군인의 건강관리 AI 툴을 개발하는 "AI 테크 스프린트" 프로그램을 실시하였다[§ 5.2(f)](VA, "AI Tech Sprint", May 24, 2024).
- 에너지부는 대규모 언어모델을 이용한 연방 인허가 절차 지원, AI를 활용한 재생에너지 생산 예측, 지능형 전력망을 통한 회복탄력성 강화 등 청정에너지 목표를 달성을 지원할 수 있는 AI 용례를 탐색하면서 전력망 애플리케이션에 적용되는 AI 모델의 보안, 데이터 편향 리스크를 해결하고 견고성을 촉진할 방안을 제시하였고[§ 5.2(g)(i)](DOE, "AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy", Apr. 2024), 산하 태평양북서국립연구소(Pacific Northwest National Laboratory)에 국가환경정책법(National Environmental Policy Act; NEPA)상 환경평가 관련 자료들(유형제 외, 환경평가, 환경영향평가서)로 훈련하여 환경실무 평가, 지식영역 식별, 지식격차 결정에 활용되는 PolicyAI라는 툴을 개발시켰으며[§ 5.2(g)(ii)], 기후변화 대응을 위한 민간 · 학계의 AI 툴 개발을 지원하고 테스트베드를 개방하는 한편[§ 5.2(g)(iii), (iv)], 산하 국립연구소들을 통할하는 중대신흥기술청(Office of Critical and Emerging Technology)을 설치하였다[§ 5.2(g)(iii), (v)].
- 대통령실 과학기술자문위원회는 단백질 구조 예측 및 치료용 단백질 설계 등 생명과학 분야, 기후 모델링 등 AI를 통한 글로벌 문제의 해결, AI의 윤리적 사용, 자원의 공정접근 연구를 제안하였다[§ 5.2(h)](PCAST, "Supercharging Research: Harnessing Artificial Intelligence to Meet Global Challenges", Apr. 2024).

56) USPTO, "Inventorship Guidance for AI-Assisted Inventions", Feb. 13, 2024.

57) Thaler v. Hirshfeld, 558 F.Supp.3d 238 (E.D. Va. 2021). Thaler v. Vidal, 43 F.4th 1207, 1213 (Fed. Cir. 2022), cert denied, 143 S. Ct. 1783 (2023). Thaler는 우리나라에서도 소를 제기했으며, 서울고등법원이 청구기각(서울고법 2024. 5. 16. 선고 2023누52088 판결)하여 대법원에 계속 중이다.

58) Pannu v. Iolab Corp, 155 F.3d 1344 (1998).

였는지를 통해 기여도를 판단하도록 한다.

특허상표청은 이를 반영하여 특허심사기준(Manual of Patent Examining Procedure; MPEP) § 2103-2106.07(c)[59]을 개정했다[§ 5.2(c)(ii)].[60] 우선 특허적격성(subject matter eligibility of patents) 판단[61]은 청구항의 AI 지원 발명 여부와 무관하며 동일 잣대를 적용할 것임을 명확히 한다. 이에 따라 AI 지원 발명의 특허적격성 판단에 있어서도 연방대법원의 기존 Alice/Mayo 2단계 테스트[62]에 따라 청구항이 추상적 발상(abstract idea) 기타 특허 부적격 사항에 해당하는지, (이들에 해당하더라도) 부적격 사항을 그 실용적 응용(practical application)으로 통합하는지에 관한 판단 기준과 사례를 제시한다. 추상적 발상의 판단 기준으로는 ① 수학적 개념, ② 인간 활동을 조직하는 방법, ③ 정신적 과정이 제시된다. 실용적 응용 통합의 판단 기준으로는 ① 추가 요소가 컴퓨터 등 기술의 기능을 향상시키는지, ② 청구항이 특허 부적격 사항을 특정 기술 환경이나 사용 분야에 일반적으로 연결하는지, ③ 청구항에서 특정 질병이나 의학적 상태에 대한 치료나 예방에 적용하는 단계가 포함되는지를 기준으로 청구항이 추상적 발상 등 특허 부적격 사항을 "직접 지칭한(directed to)" 것이 아니라면 특허적격성을 인정한다.

나. 특허상표청의 AI 저작권 관련 검토[§ 5.2(c)(iii)]

의회도서관 산하 저작권청(U.S. Copyright Office; USCO)은 AI 저작권 관련 검토를 수행 중이며 디지털 복제본(digital replica) 관련 1차 보고서를 발간한 상황인바,[63] 전체적인 검토 완료 후 180일 내에 특허상표청이 추가 조치사항을 권고할 예정이다.

다. AI 관련 지식재산권 리스크 대응 프로그램[§ 5.2(d)]

2024년 AI 로드맵에 따르면, 국토안보부는 AI 관련 지식재산권을 보호하기 위해

59) USPTO, "Manual of Patent Examining Procedure (MPEP)" (9th Edition, Rev 01.2024), Jan. 2024.
60) USPTO, "2024 Guidance Update on Patent Subject Matter Eligibility, Including on Artificial Intelligence", 89 Fed. Reg. 58128, Jul. 17, 2024.
61) 35 U.S.C. § 101.
62) Alice Corp. Pty. Ltd. v. CLS Bank Int'l, 573 U.S. 208 (2014). Mayo Collaborative Servs. v. Prometheus Labs., Inc., 566 U.S. 66 (2012).
63) USCO, "Copyright and Artificial Intelligence – Part 1: Digital Replicas", Jul. 2024.

NIST, 법무부, 국방부와 분야별 AI 위험 식별 지원, 기관에 대한 내용 표시 및 인증 방법 안내, AI 모델 보안 검토, AI 안전 · 보안 지침을 마련하고 정보분석국, 연방재난관리국 등과도 정보를 공유할 예정이다[§ 5.2(d)(i), (ii)].[64] 국토안보부는 국가지식재산권조정센터(National Intellectual Property Rights Coordination Center)를 통해 피해자 및 제3자로부터 AI 관련 지식재산 도용 신고를 접수한다[§ 5.2(d)(v)].[65]

Ⅳ. 노동자 지원(제6조)

1. 개요

오바마 행정부가 2016년 "AI, 자동화, 경제 보고서"를 공표[66]한 이후 AI의 노동시장 영향의 분석을 요한 첫 행정명령이다. AI로 인해 예상되는 노동시장의 변화를 예측하고, 선제적으로 대응할 수 있는 정책적, 법적 틀을 마련하는 것을 목표로 하며, 특히 AI를 통한 감시 · 차별 · 해고 등의 위험을 감소시키고, 노동자의 이익의 증진을 위한 모범사례의 개발에 중점을 둔다. 다만 트럼프 행정부 2기 중 개폐가 예상된다.

2. 노동시장 영향 연구[§ 6(a)]

대통령실 경제자문위원회(Council of Economic Advisers; CEA)는 2024. 7. "AI의 잠재적 노동시장 영향: 실증 분석" 보고서를 발표하였다.[67] 본 보고서는 높은 AI 노출도와 낮은 AI 관련 직무 수행 요건을 가진 노동자들을 "잠재적으로 AI에 취약한" 집단으로 분류하고 있으나, AI가 전체 고용에 부정적인 영향을 미친다는 점은 입증되지 않았다고 결론짓고 있다.[68] 한편 노동부(Department of Labor; DOL)

64) DHS, supra note 54.

65) DHS, "Intellectual Property Rights Center (IPR Center)", 〈https://www.iprcenter.gov/referral/report-ip-theft-form〉.

66) White House, "Artificial Intelligence, Automation, and the Economy", Dec. 12, 2016.

67) CEA, "Potential Labor Market Impacts of Artificial Intelligence: An Empirical Analysis", Jul. 2024.

68) CEA는 이러한 결과를 도출하기 위해 AI 노출도를 ① 소득별, ② 성별 및 인종별, ③ 연령별, ④ 지리적 패턴, ⑤ 해당 직종의 노동조합 가입 여부, ⑥ 산업별 등 기준에 따라 분석하고 AI 관련 직무 수행 요건이 높은 직종과 낮은 직종 간 고용 변화를 검토하였다. 이에 따라, AI 노출의 민감도는 선택한 기준에 따라 결과가 가변적이고, 본 연구가 단기적 데이터를 기반으로 이루어진 점을 고려한다면, 향후 지속적인 모니터링 및 데이터 보완을 통해 AI의 노동시장 영향을 장기적 관점에서 분석할 필요가 있으며, 단순한 AI 노출도가 아닌 각 직무에서의 AI 용례 등 다양한 기준을 고려해야 한다고 결론짓고 있다.

는 연방기관의 AI 대체 노동자 지원 역량에 관한 보고서를 제출하도록 되어 있으나, 공개자료에서 검색되지 않는다.

3. 노동자 피해의 최소화[§ 6(b)(i), (ii)]

노동부는 2024. 10. 19. “인공지능과 노동자 안온” 보고서를 통해 ① 노동자 참여를 중심에 둠, ② 윤리적 AI 개발, ③ AI 거버넌스와 인적 감시 확립, ④ 투명성, ⑤ 노동고용권 보호, ⑥ 노동자들의 능력 향상을 위한 AI 활용, ⑦ AI의 영향을 받는 노동자들을 지원, ⑧ 노동자 정보의 책임 있는 활용을 원리와 모범규준으로 제시하였다.[69]

4. 근로기준[§ 6(b)(iii)]

노동부는 2024. 4. 9. “공정노동기준법 및 기타 연방 노동기준의 적용 관련 직장 내 AI 및 자동화 시스템 이용 지침을 위한 현장지원안내서”를 공지하였다.[70] 본 지침은 직장 내 “AI 및 자동화 시스템”(이하 AI로 축약) 사용 시 이하와 같은 공정노동기준법(Fair Labor Standards Act of 1938)상 문제가 발생하므로 인적 감독이 필요함을 강조한다. 예컨대, (1) 임금 지급과 관련하여 ① 사용자가 노동자 위치추적 등 AI를 이용하여 근로시간, 휴게시간, 대기시간을 측정 시 근로시간이 과소 계산되어 공정노동기준법상 초과근무수당[71]이 제대로 지급되지 않을 수 있는 점, ② 초과근무수당의 기준이 되는 통상임금을 AI로 산정 시 연방임금기준을 위반하지 않도록 적절한 인적 감독이 이뤄져야 함을 지적하고, (2) 가족・의료 휴가법(Family and Medical Leave Act of 1993)이 보호하는 노동자의 휴가 사용과 관련하여, ① AI 기반 휴가 사용 요건 판단 시 인적 감독이 필요한 점, ② 중대한 건강상 문제 등 특별 휴가를 위한 승인 절차를 AI로 진행시 과도한 민감정보 수집 등 추가적 위법이 우려되는 점, ③ 긴급 모유수유 보호법(PUMP for Nursing Mothers Act)에 따른 모유수유 휴게시간을 AI 기반 일정설정・시간기록 시스템에 미산입하는 등 법정 휴게시

69) DOL, “Artificial Intelligence and Worker Well-being: Principles and Best Practices for Developers and Employers”, Oct. 19, 2024.

70) DOL, “Artificial Intelligence and Automated Systems in the Workplace under the Fair Labor Standards Act and Other Federal Labor Standards”, Apr. 29, 2024.

71) 29 U.S.C. §§ 206(a) and 207(a).

간을 위반할 수 있음을 지적한다. 이외에도 노동자의 노동부 신고나 근로감독관 협조 여부의 감시 등 자동화된 감시 시스템이 공정노동기준법상 보복금지[72] 위반으로 간주될 수 있음을 지적한다.

V. AI 편향과 시민권(제7조)

1. 개요

행정명령은 평등과 시민권 보장이 특별히 강조되는 분야를 관장하는 부처들에게 AI 이용에 따른 차별 및 시민권 침해 위험을 연구하고 정책적 조치를 취하도록 하였으나, 트럼프 행정부 2기에 상당 부분 개폐가 예상된다.

2. 형사사법(§ 7.1)

법무부(Department of Justice; DOJ)는 법집행 과정에서 AI가 야기할 수 있는 차별 등 시민권 침해 등을 해결하기 위해 유관기관과 협력하고 지원할 과업을 부여받았다[§ 7.1(a)]. 이에 따라 민권 담당 법무차관보는 2024. 10. 11.까지 네 차례에 걸쳐 기관 간 회의를 소집하였다. 특히 2024. 4. 3. 자 회의에서는 "자동화 시스템에서의 시민권, 공정경쟁, 소비자보호, 평등기회법 집행에 관한 공동 성명"이 발표되었고,[73] 10. 11. 자 회의에서는 연방거래위원회(Federal Trade Commission; FTC)의 소셜미디어 및 비디오 스트리밍 회사의 사생활 통제 없는 방대한 사용자 정보 수집 관련 대책,[74] 고용평등기회위원회(Equal Employment Opportunity Commission; EEOC)의 첨단기술 분야의 기회의 평등에 대한 장애물과 해소 방안,[75] 후술하는 교육부(Department of Education; ED)의 교육 AI 제품 개발 지침 등이 발표되었다. 법무부는 형사사법체계에서의 AI의 사용에 관한 보고서를 작성

72) 29 U.S.C. § 215(a)(3).

73) CFPB/DOJ/EEOC/FTC/DOE/HHS/HUD/DOL, "Joint Statement on Enforcement of Civil Rights, Fair Competition, Consumer Protection, and Equal Opportunity Laws in Automated Systems", Apr. 4, 2024.

74) FTC, "A Look Behind the Screens: Examining the Data Practices of Social Media and Video Streaming Services", Sep. 2024.

75) EEOC, "High Tech, Low Inclusion: Diversity in the High Tech Workforce and Sector 2014 - 2022", Sep. 2024.

중이며[§ 7.1(b)], 이를 위해 산하 사법연구소(National Institute of Justice; NIJ)가 의견제출요청을 공고하였다.[76] E.O. 14110은 2022년 "대중의 신뢰와 공공안전을 높이기 위한 효율적이고 책임 있는 경찰 및 형사사법 관행에 관한 행정명령"(E.O. 14074)[77]의 형사사법 인력 임용 조항에 보충하여 법무부가 기계학습, 인프라, 데이터 프라이버시, 데이터과학, 사용자 경험 등 AI 기술 전문성도 갖춘 형사사법 전문가들을 임용할 최선의 방안을 밝히고 공유하도록 하였고[§ 7.1(c)(i), (ii)], 형사사법 절차 중 AI 사용으로 인한 인종 차별에 대한 조사 역량을 재평가하도록 하나[§ 7.1(c)(iii)], 공개자료에서 검색되지는 않고 있다.

3. 공공부조(§ 7.2)

보건복지부의 "공공부조 행정에 있어 주 · 부족 · 지방 · 준주정부의 자동화 알고리즘 시스템에서 AI의 책임 있는 사용을 촉진하기 위한 계획"(2024. 4. 29.)[78] [§ 7.2(b)(i)]은 VIII.장에서 살펴볼 관리예산실의 M-24-10의 체계에 따라 공공부조 AI 활용 정책 목표로 ① 책임감 있는 AI 혁신의 촉진, ② 리스크 관리, ③ NIST AI RMF 등을 참고한 안전 · 보안 확보, ④ 직원들과의 협의 및 판단 존중, ⑤ 다양 · 형평 · 표용(DEI), 접근성, 시민권 보장, ⑥ AI를 사용하지 않을 선택지 유지, ⑦ 인적 감독, ⑧ 대중 보호, ⑨ 프라이버시, 시민권, 시민 자유 보호, ⑩ AI 거버넌스 강화를 제시하였다. "푸드 스탬프(food stamp)"라 불렸던 현물(식품) 공공부조를 담당하는 농무부(Department of Agriculture; USDA)도 같은 날 "공공부조 행정을 위한 주 · 부족 · 지방 · 준주의 AI 사용에 대한 프레임워크"[79][§ 7.2(b)(ii)]를 발표하여 공공부조 영역에서 AI의 사용이 권리 · 안전에 미치는 영향과 대책을 M-24-10의 체계에 맞춰 제시하였다.

---

76) NIJ, DOJ, "Request for Input From the Public on Section 7.1(b) of Executive Order 14110, 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence'", 89 Fed. Reg. 31771, Apr. 25, 2024.

77) Exec. Order No. 14074, 87 Fed. Reg. 32,945 (May 25, 2022).

78) HHS, "HHS Shares its Plan for Promoting Responsible Use of Artificial Intelligence in Automated and Algorithmic Systems by State, Local, Tribal, and Territorial Governments in the Administration of Public Benefits", Apr. 29, 2024.

79) USDA, "Framework for State, Local, Tribal, and Territorial Use of Artificial Intelligence for Public Benefit Administration", Apr. 29, 2024.

4. 정부계약자의 채용AI 활용[§ 7.3(a)]

린든 존슨 대통령의 1965년 행정명령(E.O. 11246)[80]에 따라, 정부계약자들의 직원 채용 · 고용 시에도 차별이 금지된다. 노동부 산하 연방계약준수프로그램(Office of Federal Contract Compliance Programs; OFCCP)은 정부계약자들이 채용을 위해 AI 활용 시 개발 · 사용 시 편향 완화, AI 활용 관련 고지, 노동자 또는 그 대표와 협의, 시스템 표준화, 정기 모니터링 및 분석, 인적 감시, 훈련, 데이터 문서화, 내부관리계획, 시스템 감시를 위한 거버넌스 수립 등을 이행하도록 하였다.[81]

5. 금융[§ 7.3(b)]

소비자금융보호국(Consumer Financial Protection Bureau; CFPB)과 연방주거금융청(Federal Housing Finance Agency; FHFA)은 금융감독기관들과 함께 Dodd-Frank법에 따라 주택담보에 적용되는 자동가치평가모델의 품질관리기준에 관한 행정규칙(2025. 10. 1. 시행)을 확정했으며, 특히 연방차별금지법령의 준수를 요하고 있다.[82]

6. 주거[§ 7.3(c)]

주택도시개발부(Department of Housing and Urban Development; HUD)는 2024. 4. 29. "AI 적용에 관한 공정주택법 가이드라인"을 발표하였다.[83] 이 중 첫 번째인 "공정주택법의 임차인 심사 적용에 관한 가이드라인"[84][§ 7.3(c)(i)]은 임대인의 명확한 공개기준에 따른 임차인 심사, 임차 지원자에 대한 정보제공, 입주 거부 시 근거 설명 및 이의 제기 허용, 신용 · 퇴거 · 전과기록 취급 시 특별한 주의를 강조한다. 두 번째인 "디지털 플랫폼을 통한 주택, 신용, 기타 부동산 관련 거래의

80) Fed. Reg. 65-10340 (Sep. 28, 1965).
81) DOL, "Artificial Intelligence and Equal Employment Opportunity for Federal Contractors", Apr. 29, 2024.
82) 89 Fed. Reg. 64538 (Aug. 7, 2024).
83) HUD, "HUD Issues Fair Housing Act Guidance on Applications of Artificial Intelligence", May 2, 2024.
84) HUD, "Guidance on Application of the Fair Housing Act to the Screening of Applicants for Rental Housing", Apr. 29, 2024.

광고에 공정주택법을 적용하는 것에 관한 가이드라인"[85][§ 7.3(c)(i)]은 특히 목표 광고 수신자 선별 및 맞춤형 추천 시 차별이 초래되지 않도록 하고 있다.

VI. 소비자 보호(제8조)

1. 개요

특히 보건복지 · 교육 · 교통 분야에서 AI 관련 소비자 피해를 경감하고자 한다.

2. 보건복지[§ 8(b)]

보건복지부는 2023. 12. 19. 의료제공, 복지, 연구개발을 비롯한 8가지 분야를 아우르는 범부처 태스크포스를 설치하였음을 밝혔고[86] 태스크포스는 설치 후 365일 이내에 보건복지 AI에 전반에 관한 전략계획을 수립하여 보고서를 발표해야 한다[§ 8(b)(i)]. 보건복지부는 AI 품질보증 정책 및 인프라 수요 개발 등 품질평가 전략도 개발해야 하는데[§ 8(b)(ii)], 개발 완료 사실을 밝혔으나[87] 상세가 확인되지는 않는다. 또한 보건복지부는 의료 AI에 대한 차별금지 법령들의 적용을 명확하게 하기 위해[§ 8(b)(iii)], 산하 민권국(Office for Civil Rights; OCR)이 2024. 5. 6. 의료비용합리화법(Affordable Care Act; "오바마케어") § 1557[88]에 대한 행정규칙 개정안을 입법예고하였는데,[89] 특히 § 92.210를 정비하여 환자치료결정지원도구(patient care decision support tool)에 있어 AI 활용이나 자동화 여부를 불문하고 차별이 금지됨을 명확히 했다. 나아가 보건복지부는 의료AI로 인한 임상적 오류를 중앙에서 취합하여 축적된 데이터와 증거를 가지고 모범규준을 만들어 전파하는

85) HUD, "Guidance on Application of the Fair Housing Act to the Advertising of Housing, Credit, and Other Real Estate-Related Transactions through Digital Platforms", Apr. 29, 2024.
86) Micky Tripathi, "Leveraging Agency Expertise to Foster American AI Leadership and Innovation", Dec. 19, 2023.
87) White House, "Biden-Harris Administration Announces Key AI Actions 180 Days Following President Biden's Landmark Executive Order", Apr. 29, 2024.
88) 의료 분야의 차별금지사유를 민권법(Civil Rights Act of 1964)상 사유들(인종, 피부색, 종교) 이외의 사유들(출생국, 성별, 연령, 장애)까지 확장한 조항이다.
89) HHS, "Nondiscrimination in Health Programs and Activities", 45 C.F.R. Part 92, 89 Fed. Reg. 37522, May 6, 2024.

절차에 관한 공청회를 진행했다[§ 8(b)(iv)].[90]

3. 교통[§ 8(c)]

교통부(Department of Transportation; DOT) 산하 비전통 · 신흥교통기술위원회(Nontraditional and Emerging Transportation Technology Council; NETT Council)는 2018년 새 교통기술이 기존 교통부 산하 기관 분류체계에 맞지 않거나 새로운 접근이 필요한 경우에 대응하고자 설치되었다. NETT 위원회는 교통부 산하 기관별 AI 용례와 영향을 분석하고 있다[§ 8(c)(i)].[91] NETT 위원회는 교통부의 AI 거버넌스 위원회로서 ① AI 거버넌스 절차와 정책 검토 · 승인, ② 최고AI책임자(Chief AI Officer; CAIO)의 AI 용례 비공개 기준 승인, ③ 안전 · 권익 영향 AI 용례 지정 기준 승인, ④ 외부 의견 수렴 등을 수행한다.[92] 교통부는 또한 2023. 12. 29. 교통혁신의 계획과 접근 관련 조언을 제공받고자 교통혁신자문위원회(Transforming Transportation Advisory Committee; TTAC)를 설치하였으며[§ 8(c)(ii)],[93] 기반시설 고등연구계획국(ARPA-I)[94]에 교통 분야에서 AI의 기회 및 과제를 연구하도록 하고 2024. 5. 3. 정보요청을 공고했다[§ 8(c)(iii)].[95]

4. 교육[§ 8(d)]

교육부 산하 교육기술실(Office of Educational Technology; OET)은 2024. 7. "AI를 활용한 교육의 설계: 개발자들을 위한 필수 지침"을 발표하였다.[96] 지침서는 교육 분야 AI 제품 개발자들의 제품 신뢰 확보를 위한 혁신 · 책임의 통합을 강조

90) HHS, "Meeting on the Artificial Intelligence in Healthcare Safety Program", Oct. 31, 2024.
91) DOT, "Opportunities and Challenges of Artificial Intelligence (AI) in Transportation; Request for Information, 2024-09645", 89 Fed. Reg. 36848, May 3, 2024.
92) DOT, "US Department of Transportation (DOT) Compliance Plan for OMB Memorandum M-24-10 (September 2024)", Sep. 24, 2024.
93) DOT, "DOT Launches New Advisory Committee to Help Shape Federal Approach to Transportation Innovation", Dec. 29, 2023.
94) 초당적 인프라법(Bipartisan Infrastructure Law)의 수권하에 국방부 산하 국방고등연구계획국(DARPA)과 에너지부 산하의 ARPA-E를 모델로 하여, 상업적 투자가 어려운 고위험 · 고보상의 차세대 교통기술의 국가적 연구를 목적으로 교통부 산하에 설치된 기관이다.
95) DOT, supra note 91.
96) OET, "Designing for Education with Artificial Intelligence: An Essential Guide for Developers", Jul. 2024.

하면서, ① 인간중심적 설계를 위해 교육자와 학생의 피드백을 반영, ② 투명성·정의·공정성 등 윤리 및 교육자와 학생의 시민권을 보호, ③ 제품의 교육성과의 긍정적 영향에 대해 분명한 증거를 제시, ④ 프라이버시, 보안 법령을 준수하고 위험을 경감·해소, ⑤ AI 시스템의 투명성 확보를 위해 교육 분야 당사자들의 AI 문해력 증진을 제안한다.

Ⅶ. 프라이버시(제9조)

1. 개요

이 장은 AI 도입에 따른 데이터 수집·이용·보관·파기 정책에 관한 검토 및 프라이버시 향상 기술(privacy-enhancing technology; PET)의 연구를 요하고 있다.

2. 이행 현황

대통령실 관리예산실은 연방기관들이 수집하는 상업적 이용 가능 정보(commercially available information; CAI)가 적정한 보유·보고 절차를 거치는지 평가해야 하는데[§ 9(a)(i)], 이를 위해 정보요청을 공고하였고,[97] 이에 입각하여 기관들의 CAI 처리에 관한 표준 및 절차를 제시할 예정이다[§ 9(a)(ii)]. 나아가 관리예산실은 전자정부법(E-Government Act of 2002)에 따른 프라이버시 보호 영향평가의 AI 관련 개선 방안에 대한 정보요청을 공고하였다[§ 9(a)(ii)].[98] NIST는 2023. 12. 11. 차분 프라이버시 보장 평가 지침서의 초안을 발표하였다[§ 9(b)].[99] 국립과학재단과 에너지부는 2024. 2. 26. PET의 개발·도입·확장을 위한 연구 협력 네트워크를 발족시켰다[§ 9(c)].[100] 국립과학재단은 2024. 6. 26. PET 중 현실성 있는 기술을 발전시켜 기관들에 도입하는 프로그램을 발표하였다.[101]

---

97) OMB, "Request for Information: Executive Branch Agency Handling of Commercially Available Information Containing Personally Identifiable Information", 89 Fed. Reg. 83517, Oct. 16, 2024.

98) OMB, "Request for Information: Privacy Impact Assessments", 89 Fed. Reg. 5945, Jan. 30, 2024.

99) NIST, "Guidelines for Evaluating Differential Privacy Guarantees (NIST SP 800-226 ipd), Dec. 11, 2023.

100) NSF, "NSF and DOE Establish a Research Coordination Network Dedicated to Enhancing Privacy Research", Feb. 26, 2024.

101) NSF, "NSF Launches New Investment to Accelerate the Transition of Privacy-enhancing Technologies

Ⅷ. 연방의 AI 활용 촉진(제10조)

1. 개요

행정명령에 따라 50여 개가 넘는 연방기관들에 100여 개의 과업이 부여되고 막대한 예산이 투여되었기 때문에,[102] 미국 특유의 강력한 국무조정이 없었다면 제대로 이행될 수 없었을 것이다. 대통령실 관리예산실, 과학기술정책실, 과학기술자문위원회(President's Council of Advisors on Science and Technology; PCAST), AI위원회(AI Council)가 컨트롤타워 역할을 하고 있으나,[103] 특히 대통령실 최대 조직인 관리예산실의 활약이 두드러졌다. 관리예산실은 우리 시스템에 비유컨대 기획재정부 예산실과 국무조정실이 합일되어 있을 뿐 아니라 대통령실로 격상되어 더 강력해진 형태로 이해할 수 있다. 관리예산실은 예산과 연계된 강력한 관리 기능으로 연방기관들을 총동원하여 2차 대전 중 전시동원 체제를 지휘하였을 뿐 아니라, 현재는 AI에 대해 각 부처의 경쟁적인 AI 활용을 독려하고 범부처적 리스크 관리 체계를 구축하고 있다. 이하에서 살펴본다.

2. 관리예산실, "기관의 AI 활용에 대한 거버넌스, 혁신, 리스크 관리의 진전" 각서(M-24-10)(2024. 3. 28.)[§ 10.1(a), (b), (c), (e)][104]

M-24-10은 연방기관들에 지시와 이행지침을 하달하는 관리예산실 각서의 형태로 발령되었는데, 모든 연방기관이 개발 · 사용 · 조달하는 신규 · 기존 AI(단, 국가안보체계 활용 시 제외)에 대해 적용된다. 상세는 이하와 같다.

가. AI 거버넌스의 강화[105]

모든 연방기관들은 M-24-10에 따라 최고AI책임자(CAIO)를 지정하고 관리예

---

to Practice", Jun. 27, 2024.

102) CRS, supra note 10, 3.

103) GAO, "Artificial Intelligence: Agencies Are Implementing Management and Personnel Requirements", Sep. 2024, 6-12.

104) OMB, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence" (M-24-10), Mar. 28, 2024.

105) Id., 4-9.

산실에 보고하였다[§ 10.1(b)(i)]. CAIO는 M-24-10의 이행, 타 기관과의 협조, AI 활용 조율, 혁신 진흥, 리스크 관리에 대한 책임을 진다. CAIO는 전문성과 학력을 갖춰야 하고 기관장과 주기적으로 의견을 교환할 정도의 직급을 보유해야 한다. 최고재무책임자법(Chief Financial Officers Act of 1990; CFO Act) 적용 대상 기관들은 AI 거버넌스 위원회(AI Governance Board; AIGB)도 조직하여야 하는데[§ 10.1(b)(iii)], 의장은 차관, 부의장은 CAIO가 맡고, IT, 사이버보안, 데이터, 개인정보 보호 담당자들로 구성되며, 외부 전문가와의 협의가 권고된다. 각 부처는 M-24-10의 이행계획을 관리예산실에 제출하고 각자의 웹사이트에 공개했으며, 2036년까지 격년마다 제출 · 공개할 예정이다.

나아가, 각 기관은 매년 1회 이상 AI 용례를 정리하고 용례집을 관리예산실에 제출하며 "연방 AI 활용 인벤토리(Federal AI Use Inventory)" 웹사이트에서 공개한다[§ 10.1(e)]. 국방부와 정보기관은 예외이나, 이들도 용례의 집합적 통계, 권리 · 안전영향AI의 건수, M-24-10 준수 여부를 취합하여 매년 관리예산실에 보고해야 한다.

나. AI의 책임 있는 혁신의 진전[106)]

M-24-10은 연방정부에서의 AI 활용 촉진을 위해 다음 조치들을 요한다.

- AI전략: CFO Act 적용 대상 기관들은 AI 활용 장애요소의 확인 · 제거 및 AI 성숙도 증진 전략을 개발하여 웹사이트에 공개했다.
- AI 활용 장애요소 제거: 각 기관은 AI 프로젝트의 IT 기반시설 접근, AI 개발자의 소프트웨어 툴 및 오픈소스 라이브러리에 대한 접근을 보장하여야 한다. 또한 AI 훈련 · 검사 · 운영에 사용할 수 있도록 기관의 데이터의 공유 · 큐레이션 · 관리 인프라를 개방하며, 데이터 품질 · 대표성 · 편향성 평가를 실시한다. AI 앱의 운영을 위해 정보시스템 승인, 지속적 모니터링 프로세스를 갱신하고, 생성AI의 유익성 평가, 안전장치, 감독 체계를 확립하여야 한다.
- AI 인재 임용: AI 및 기술 인력 채용지침을 준수하고, AI 인력 채용을 기관장에 보고하며, 해당 기관의 AI 인력 채용을 추적 · 관리하고, AI 인재 책임자를 지

106) Id., 5-14.

정하고, AI 인재 개발 및 관내 교육을 지원한다.

- AI 자원 공유 · 협력: 각 기관은 AI 앱을 위해 개발하여 사용 중인 맞춤형 코드를 적극 공유하고, 적절할 경우 오픈소스로 배포 · 유지 · 관리한다. 또한 AI 훈련 · 시험 데이터가 공개 · 공공 · 전자 · 필수정부데이터법상 데이터자산에 해당하면 공개해야 한다. 모델이 비공개대상이면 평가데이터를 공개하거나 연방 내부적으로만 모델을 공유하는 등 조치를 취한다. AI 맞춤형 코드 개발, 훈련 · 시험 데이터, 기존 데이터 강화 시 외부 공개 가능 방식으로 진행하는 것을 권장한다. AI 모델 및 모델 가중치 공개 여부 결정시 데이터 내 세부 민감정보 추론 리스크를 평가한다.
- AI 요구조건의 통합: 리소스 및 용례 공유를 위해 연방 내 기관들이 일관된 방식(예컨대 동일 서식 사용, 모범 사례 등 공유, 기술자원 공유, 용례 홍보)을 사용하도록 한다.

다. AI 활용으로부터의 리스크 관리[107)]

정보기관을 제외한 모든 정부기관은, AI 용례에 대해 이하 권리 · 안전영향AI 해당 여부를 검토하고, 이하 최소준수기준을 준수하며, 부합되지 않는 방식의 AI 활용을 중단하여야 한다. 또한 권리 · 안전영향AI 해당 여부에 관한 결정 및 최소준수기준 적용 면제의 정당성을 설명하는 문서를 공개해야 한다.

[표 3] M-24-10상 권리 · 안전영향 AI의 개념과 최소준수기준

쟁점	권리영향 AI (rights-impacting AI)	안전영향 AI (safety-impacting AI)
해당 여부	이하 항목 중 하나에 해당하는 AI[108)]: • 시민권 · 시민자유 · 프라이버시: 표현의 자유, 투표, 사람의 자율성, 차별 · 과잉처벌 · 불법감시로부터의 보호 • 기회균등: 교육 · 주택 · 보험 · 신용 · 고용에 대한 동등접근 • 중대 정부 자원 · 서비스에 대한 접근 · 신청권에 법적 · 실질적 · 구속적 또는 유사하게 중대한 영향이 있는 특정 개인 · 법인에 대한 결정 · 행동의	이하 항목 중 하나에 해당하는 AI[109)]: • 사람의 생명 · 건강 관련 • 기후변화 및 환경 관련 • 주요기반시설 관련 • 전략자산 · 자원 관련: 댐, 응급 서비스, 전력망, 에너지 생성 · 이동, 소방 안전 시스템, 식품 안전 메커니즘, 교통 제어 및 물리적 운송 제어 시스템, 상하수도, 원자로 · 자재 · 폐기물 제어 시스템 내의 중요 안전 제어 기능 등

107) Id., 14-26.
108) Id., 29-33.
109) Id., 29-33.

	주요 기초가 되는 AI: 보건 · 금융 · 공공주택 · 사회보장 · 교통 · 필수재화/용역	
공통 최소 준수 기준	• AI 영향 평가의 주기적 실시: AI의 의도된 목적과 기대되는 이익, AI 사용에 따른 잠재적 리스크, 관련 데이터의 품질 및 적절성 포함 • AI 성능 테스트: 현실 상황에서 의도된 대로 작동할 수 있는지, 예상 편익 달성 및 리스크 통제가 가능한지 모범규준과 각계의 피드백을 고려하여 확인 • CAIO, AIGB의 검토: 현실 상황에서의 성능 시험 결과 및 AI 영향평가 결과 포함 • AI 기능 저하의 모니터링, 권리 · 안전영향의 지속적 변화 감지 절차 마련, 부정적 성능 · 결과 적시 확인 기능 마련 • AI 리스크 정기 평가: 인적 검토가 요구되며, 본 최소준수기준의 리스크 경감 여부 및 개선 필요성에 대한 판단 포함 • 권리 · 안전에 대한 신규 · 변경 리스크 감지 시 현행화 또는 절차적 · 수동적 완화 장치 마련 등 리스크 경감 조치 • 교육 · 평가 · 감독: AI 운영자가 AI의 출력을 해석하여 필요한 조치를 취하고, 자동화편향 등 인간-기계 협업 문제를 해결하며, 사람 기반 요소로 리스크를 관리 • 사람의 추가적인 감독 · 개입 · 책임 부담 없이는 AI가 실행할 수 없는 결정 · 행동 구분 • AI 용례의 누적을 통해 AI의 기능성을 대중에게 쉽게 설득	
항목별 최소 준수 기준	• 형평성 · 공정성 영향 식별 · 평가 및 차별 경감 조치 • 영향을 받는 지역사회 및 대중으로부터 피드백을 받아 반영 • 차별에 대한 지속적 모니터링 · 경감 • 부정적으로 영향 받는 개인에 통지 • 인적 검토와 구제 절차 유지 • AI 기반 의사결정 거부권	• 준법 • 투명성 및 성능 개선 • AI 조달에서의 경쟁 촉진 • AI를 위한 데이터의 품질 최대화 • 알려진 시험데이터 과의존 해결 • 생체인식정보에 대한 책임 있는 조달 • 생성AI의 책임 있는 조달 • 환경 효율성 및 지속가능성 평가

3. 연방정부 내 생성AI 활용[§ 10.1(f)]

가. FedRAMP, "신흥기술 우선 프레임워크"(2024. 1. 26.)[110][§ 10.1(f)(ii)]

FedRAMP는 연방기관이 조달하는 클라우드 제품 · 서비스의 보안 평가 · 인증 · 모니터링에 대한 표준 접근 방식을 규정하는 연방 클라우드 위험 · 인증 관리 프로그램으로서, 우리의 클라우드 보안인증제(CSAP)(클라우드컴퓨팅 발전 및 이용자 보호에 관한 법률 제23조의2)도 FedRAMP를 참고해서 만든 것이다. 본 프레임워크는 최고정보관리자(CIO) 위원회 또는 최고정보안전관리자(CISO) 위원회가 신흥기술 후보를 추천하면, FedRAMP 프로그램 관리국이 최대 3개의 신흥기술을 선정하며, FedRAMP 이사회는 이를 공공클라우드 제공 가능 서비스 목록(cloud

110) FedRAMP, "Emerging Technology Prioritization Framework", Jan. 26, 2024.

service offering; CSO)으로 승인할 수 있다. 관리국은 승인된 신흥기술을 기초로 프로세스, 웹사이트 및 시스템을 현행화해야 한다. 관리국은, CSO가 신흥기술 기준에 부합하는지 판단하고, 기준이 충족되면 그 사업적 수요를 타 CSO와 비교하며, 최종적으로 해당 CSO를 FedRAMP 승인 대기 목록에서 우선 배치한다. 현재 CSO의 우선순위는 챗 인터페이스, 코드 생성 및 디버깅 툴, 프롬프트 기반 이미지 생성기에 맞춰져 있다.

나. 인사관리처, "연방 인력의 책임 있는 생성AI 활용"[111][§ 10.1(f)(iii)]

연방정부의 공무원이 공직 수행 과정에서 생성AI 사용 시 준수사항을 ① 관련 정책과 절차 준수, ② 사적 용도 사용 금지, ③ 공무 사용 가능 생성AI 기술 목록 확인 및 사용 지침 준수, ④ 공개된 생성AI 인터페이스에 비공개 정보 입력 금지, ⑤ 사용 감독, ⑥ 생성결과물의 (출처를 포함한) 유효 · 정확 · 완전 · 안전성 검토, ⑦ 생성결과물의 편향 · 유해성 · 고정관념성 등 검토, ⑧ 생성결과물에 대한 위법행위(조작, 변조, 표절 등) 방지, ⑨ 공개 방식에 대한 정책(레이블, 워터마킹, 면책조항 등) 준수, ⑩ 악의적 · 부적절 · 불법적 자료 생성을 위한 활용 금지로 규정하고 있다.

4. AI 조달[§ 10.1(d)(ii), (h)]

관리예산실은 2024. 9. 24. AI 시스템 및 서비스의 조달 절차에 M-24-10을 반영하기 위해 "정부 내 책임 있는 AI 조달 촉진"(M-24-18)을 발표했다[§ 10.1(d)(ii)].[112] 동 각서는 ① 기능 · 기관 간 협력 구축과 관련하여, 각 기관이 AI 조달 정책 · 절차 · 실무를 신규 작성 및 갱신하고 타 기관과 정보를 공유하도록 하고, ② AI 리스크 및 성과 관리와 관련하여, AI 시스템 및 서비스가 개발 · 훈련 · 배포되는 방식을 고려하여 AI 조달 정책 · 실무(특히 권리 · 안전AI 관련)를 조정하는 한편, 생성AI 및 AI 생체 인식 시스템 조달 관련 추가 조치를 하며, ③ 혁신을 통한 AI 시장 경쟁력 증진과 관련하여, 정부기관이 항시 최첨단 AI를 확보할 수 있도록 AI 채택 의사

111) OPM, "Responsible Use of Generative Artificial Intelligence for the Federal Workforce", 〈https://www.opm.gov/data/resources/ai-guidance〉.

112) OMB, "Advancing the Responsible Acquisition of Artificial Intelligence in Government" (M-24-18), Sep. 24, 2024.

결정에 있어 우선순위를 정하고 특정 공급업체에 대한 고착화를 방지하도록 하면서, 연방정부의 AI 조달 시장이 다양성·경쟁력·회복탄력성을 유지할 수 있도록 지원을 다할 것을 권고하고 있다.

나아가 총무청(General Services Administration; GSA) 산하 IT벤더관리국(Information Technology Vendor Management Office; ITVMO)은 공공조달 과정에서 생성AI 개발자들에 대한 접근성을 높이기 위해 "생성AI 및 특수 컴퓨팅 인프라 확보 자원 가이드"를 발표하였다[§ 10.1(h)].[113] 동 가이드는 ① 통합사업관리팀이 조달사업의 목표, 제한사항, 연구 방향, 검증 방식을 설정할 것, ② 각 기관의 목표 및 수요를 최고정보책임자(CIO)나 CAIO 등을 통해 확인할 것, ③ 조달에 앞서 클라우드 플랫폼상의 샌드박스에 데이터를 업로드하고 다양한 생성AI 모듈을 테스트할 것, ④ 데이터의 내역을 구분해서 파악할 것, ⑤ 다양한 구매 전략을 기획하고, 신규 구매 시 경쟁을 유도할 것, ⑥ 기술의 변동, FedRAMP 인증 획득 여부, 공급업체에 대한 표준실사 결과에 따라 시장을 평가하고, 과금 방식별로 가격을 평가할 것, ⑦ 조달 후 성능 모니터링 및 평가를 수행할 것을 요한다.

5. 연방정부 내 AI 인재 확대(§ 10.2)

대통령실 과학·정책·사회프로그램국(Office of Science, Policy and Society Programs; OSPSP), 과학기술정책실, 관리예산실의 주도로 "AI와 기술인력 태스크포스"가 2023. 11. 결성되어 매월 소집되고 있고[§ 10.2(b)],[114] 태스크포스는 연방정부 내 AI 전문가의 채용·유지를 목표로 하는 "국가 AI 인재 증진(National AI Talent Surge)" 프로그램을 출범시켰으며, 목표를 ① 정부 AI 활용의 확장, ② AI 규제 및 정책 역량 구축, ③ 연구개발 생태계의 강화로 설정했다[§ 10.2(a)].[115] 인사관리처 등 인사부처들은 실행을 위해 다각적 조치들을 수립하여 이행 중이다.[116] 관리예산처는 2023. 12. 각 부처의 인사 담당자들로 구성된 부처 간 워킹그

113) GSA, "Generative AI and Specialized Computing Infrastructure Acquisition Resource Guide", 〈https://itvmo.gsa.gov/genai〉.

114) GAO, "Artificial Intelligence: Agencies Are Implementing Management and Personnel Requirements", GAO-24-107332, Sep. 2024, 17.

115) AI and Tech Talent Task Force, "Report to the President: Increasing AI Capacity Across the Federal Government: AI Talent Surge Progress and Recommendations", Apr. 2024.

116) OPM의 조치는 이하와 같다.

룹을 구성해서 소집했다[§ 10.2(d)(vi)].[117]

Ⅸ. 국제적 리더십(제11조)

1. 배경

미국은 이하 조치들로써 AI 관련 국제표준을 주도하고, 국제개발과 국제연구를 지원하며, 국경을 넘는 AI 리스크에 대응함으로써 국제적 리더십을 강화하고자 한다.

2. 이행 상황

NIST의 2024. 6. 24. 자 "AI 표준에 대한 글로벌 관여 계획"은 NIST AI RMF와 "핵심신흥기술에 대한 미국 정부의 국가표준전략"에 입각하여 AI 표준 관련 목표와 우선조치를 제시한다[§ 11(b)(i)].[118] 국가표준전략은 민간 주도의 표준 생태계

- AI 인재의 신속한 임용을 위해 직접임용권(Direct Hire Authority)과 특별임용(excepted service appointments)의 근거를 명확히 했다[§ 10.2(c), (d)(ii)][OPM, "Government-wide Hiring Authorities for Advancing Federal Government Use of Artificial Intelligence (AI)", Dec. 29, 2023].
- AI 인재 임용 유연성 증진을 위해 신규 · 지역근무 · 이직고려인력 기본급 가산, 학자금 상환 지원, 민간 근무경력 산입, 고위공직자 연차 적립율 확장, 유연 · 재택근로제 확장 등 인센티브를 제시했다[§ 10.2(d)(i), (iv)][OPM, "Pay Flexibility, Incentive Pay, and Leave and Workforce Flexibility Programs for Artificial Intelligence (AI), AI-enabling, and Other Key Technical Employees", Feb. 27, 2024].
- 미국판 "나라일터"인 USAJOBS의 신규 인재 풀 기능을 확충했다[§ 10.2(d)(iii)][OPM, "New Talent Pools (Shared Certs) Feature for USAJOBS' Agency Talent Portal", Dec. 1, 2023].
- 인사와 AI 전문가로 구성된 범부처 작업반인 Tech to Gov를 운영 중이다[§ 10.2(d)(v)(i)].
- 토목사업 AI 활용에 대한 역량 명세(데이터 분석력, 설계능력, 수학 · 통계 전문성)를 정했다[§ 10.2(d)(viii)][OPM, "Artificial Intelligence (AI) Competency Model for Civil Engineering, 0810", Apr. 29, 2024].
- 정부AI법(AI in Government Act of 2020)에 따라 AI 인재의 직무 분류를 명확히 하며 보수 · 직렬 · 직책 · 업무난이도 등을 고려하도록 했다[§ 10.2(f)][OPM, "Artificial Intelligence (AI) Classification Policy and Talent Acquisition Guidance", Apr. 2024].
- 관리예산실과 함께 기술직 임용 전문훈련개발 프로그램(Tech Hiring Professional Training and Development Program)을 준비하였다[§ 10.2(g)].

그 밖에, 최고인적자원책임자협의회(Chief Human Capital Officers Council; CHCOC)는 비전통전공자의 임용을 촉진하기 위해 기술 기반 임용 절차를 정비하였다[§ 10.2(d)(v)](CHCOC, "Skills-Based Hiring Guidance and Competency Model for Artificial Intelligence Positions", Apr. 29, 2024).

117) GAO, supra note 114, 19.

118) NIST, "A Plan for Global Engagement on AI Standards" (NIST AI 100-5), Jul. 2024.

의 참여자로서의 연방정부의 역할을 강조하면서, AI 표준에 대해 ① 접근·수정을 허용하는 과학성·합리성, ② 세계의 다양한 이해관계자의 요구·필요 반영, ③ 개방성, 투명성, 합의 기반 절차에 따른 개발, ④ 국제 관계의 강화를 추구한다. 그 과정에서 AI 표준 개발·채택에 있어 이해관계자의 참여를 촉진하고, 국제적 정렬을 촉진하며, AI 표준이 AI 생태계에서 수행할 역할에 대한 공통 이해를 확인하고자 한다. 상무부는 이에 입각해서 우선조치 보고서를 대통령에게 제출할 예정이다[§ 11(b)(ii)].

국제개발처(U.S. Agency for International Development; USAID)와 국무부(Department of State; DOS)는 "국제개발에서의 AI 플레이북"을 발표하여 AI 관련 국제개발과 관련하여 ① 역량 증진, AI 기법 향상, 인력 보호, ② 신뢰받고 지속가능한 디지털 기반 구축, ③ 데이터 저장 및 컴퓨팅 리소스에 대한 접근 확대, ④ 지역을 대표하는 데이터셋의 구축 및 문화유산의 보전, ⑤ AI 유망성 환기 전략 개발, ⑥ 안전하고 권리를 존중하는 AI 시스템의 개발·이용에 대한 거버넌스 프레임워크 촉진, ⑦ 개방성, 투명성, 설명가능성을 통해 AI 신뢰 증진, ⑧ 기후대응을 위한 활용을 권고하였다[§ 11(c)(i)].[119] 국무부 등은 2024. 9. "글로벌 AI 연구 어젠더"를 통해 핵심 의제로 포용과 형평, 책임 있는 연구 실행, 동반관계와 협력을 들면서, 국제연구 우선순위를 사회기술적 연구, 포용적 연구 인프라, 국제적 도전과제를 위한 AI를 지원하는 연구, AI 안전·보안·신뢰성을 포함한 AI 관련 근본적 연구, AI의 국제적 노동시장에 대한 함의의 연구로 제시한다[§ 11(c)(ii)].[120]

국토안보부는 2024. 7. 26. 캐나다, 멕시코, EU, 파이브아이즈(Five Eyes)와 함께 사이버보안과 국경을 넘는 주요기반시설 문제와 협력 방안을 다루는 AI 국제관여 계획을 작성했다고 발표하였으나[§ 11(d)(i)][121] 상세는 공개되어 있지 않다. 국토안보부는 이에 입각해서 주요기반시설에 대한 초국경적 리스크의 경감을 위한 우선조치 보고서를 제출할 예정이다[§ 11(d)(ii)].

119) DOS/USAID, "AI in Global Development Playbook", Sep. 2024.
120) DOS/USAID/DOL/DOE/NSF, "The Global AI Research Agenda", Sep. 2024.
121) DHS, "FACT SHEET: DHS Deploys Pilot Programs, Expands Efforts to Leverage AI to Secure Nation's Critical Infrastructure", Jul. 26, 2024.

Ⅹ. 의미와 시사점

안보, 민간, 공공 영역을 나누어 미국의 E.O. 14110 이행의 핵심적 의미를 결산하고 우리 입법과 정책에 대한 시사점을 도출하면 다음과 같다.

1. 안보 영역

E.O. 14110의 이행은 50여 개 넘는 연방기관들에 100여 개의 조치를 요구하여 막대한 예산을 들여 방대한 조사 · 검토 · 보고를 수행한 범연방적 과업이었다. 미국은 전시동원법인 국방물자생산법(DPA)을 재소환하여 이중용도 기반모델과 대규모 컴퓨터 클러스터의 현황을 주기적으로 보고받아 파악, 유사 시 동원하는 체계를 준비하고, 16개 주요기반시설의 150건이 넘는 AI 용례를 식별하여 안보 · 보안 대책을 수립하는 한편, 2차 대전 중 연방 역량의 결집을 이끌었던 대통령실 관리예산실을 중심으로 연방기관들을 AI 전환 경쟁에 밀어붙이는 등, 전시에 준하는 총력적 동원을 통해 전략적 · 산업적 우위를 도모하고 있다.

또한 AI의 공급망 문제에 대해 이중용도 기반모델을 통제 대상으로 삼고, 국제긴급경제권한법(IEEPA)과 국가비상사태법(NEA)의 적용 범위를 확대하여 적국과 테러집단 등의 클라우드 인프라 서비스(IaaS)와 대규모 모델 훈련을 통제하며, 생물화학무기(CBW)에 대해 선제적 대응을 갖추는 등, 군사기술 내지 대량살상기술로 전용될 수 있는 AI에 대해 수출통제에 준하는 안보적 접근을 하겠다는 취지 또한 명확히 하고 있다.

이처럼 미국은 AI에 있어 역량을 총동원하여 우위를 확보하는 한편, 대외적으로 경제와 안보를 분리하지 않겠다는 분명한 메시지를 전달하고 있다. 이는 향후 우리의 AI에 대한 안보적 입장이, 방위산업 등 제한된 범위가 아니라, AI가 사용될 전 산업 영역에 근본적 영향을 미칠 수 있음을 시사한다. 미국의 입장을 포함한 국제정세에 촉각을 세우고 보조를 맞추어 미국이 보유한 압도적인 핵심 기술과 모델에 대한 접근에 있어 우선적 지위를 도모할 수밖에 없다. 이번엔 다행히 보류했으나 공개모델의 폐쇄모델 전환 문제까지 의제에 올렸음은 분명하므로, 향후 갑작스러운 입장 전환이 전산업적 충격으로 이어지지 않도록 독자적인 AI 기술 개발과 투자에도 박차를 가해야 한다. 미국이 비상한 관심을 보이고 있는 합성생물학과 생

물안보 분야에 대한 전략적 투자도 늘려야 함은 물론이다.

2. 민간 영역

미국은 EU식 수평·포괄규제를 지양하고 각 연방기관에 관할별 세부 목표를 부여하는 부문·맥락특유적 규제를 추진하는 한편, 대통령실 차원에서의 예산편성과 연계된 강력한 국무조정 및 경성규범과 연성규범 간 유연한 역할 분담을 통해 정연성(coherency)까지 챙기고 있음을 확인할 수 있다. 우리는 EU식 수평·포괄규제로 다소 치우치는 중인데, 방송통신위원회가 AI이용자법안을 준비 중인 것을 고려하면, AI의 용례마다 해당 분야의 전통적 규제법, AI기본법, AI이용자법의 AI 3중 규제 체제가 자리 잡을 우려가 있다. 그러나 E.O. 14110의 이행 현황을 면밀히 주시하여 현재의 방향성을 재고하고 과잉·중복규제를 피하며 혁신 역량을 보존하기 위해 부문·맥락특유적 접근방식을 바탕으로 규제를 설계해야 한다.[122]

다만 우리 환경의 미국과의 차이점도 고려해야 한다. 우리는 미국 수준의 부처의 통할이 어려우며, 부문·맥락특유적 규제체계를 섣불리 도입 시 도리어 각 부처 간 규제·관할경쟁을 촉발할 수 있다. M-24-10이 요구하는 대로 각 부처들의 신속한 AI 인력 임용과 전문성 확충이 이루어져야 하나, 단기간에 이 목표를 이루기는 어렵다. 따라서 규제체계의 설계에 있어 정연성을 확보하기 위해 AI에 대해 탁월한 전문성이 있는 전문 부처들의 역할도 필요하다. 절충점으로 참고할 수 있는 것은 일본이 국제동향을 면밀히 주시하다가 최근에 정한 방향이다. 일본은 경제산업성/총무성의 AI 전체에 대한 연성규범과 각 부처의 분야별 경성규범을 조합하는 방식으로 규제체계를 설계하고자 한다.[123] 이를 참고하여 ① 전문 부처들은 전문성을 살려 (AI 진흥을 주도함과 아울러) AI 안전과학 연구의 지원에 박차를 가하고 이를 근거로 AI의 유형별 연성규범을 생산해서 각 부처에 공급하며, ② 각 부처는 이를 참고하여 AI의 개별 용례에 대해 기존 규제체계를 적응(adapt)하고 핀셋형·맞춤형 규제를 마련, 집행하거나 민간 표준과 역할을 분담하고, ③ 대통령실은 부처별 규제체계가 정연성을 확보할 수 있도록 예산편성과 연계된 강력한 국

122) EU식 포괄·수평규제체계의 문제점에 대해서는 Sangchul Park, "Bridging the Global Divide in AI Regulation: A Proposal for a Contextual, Coherent, and Commensurable Framework", 33(2) Wash. Int'l L. J. 216, 234-238 (2024).

123) 経済産業省/総務省, AI事業者ガイドライン(第1.0版), 2024年 4月 19日.

무조정을 시행하여 국가적 역량을 결집하는 등 치밀하고 협력적인 거버넌스를 추구해야 한다.

3. 공공 영역

미국은 연방정부의 AI 활용에 대해서는 관리예산실의 강력한 예산과 국무조정권한을 통해 기관 간 활용 경쟁을 촉발하여 연방의 디지털 전환을 가속화시키는 한편, 연방기관들의 인력과 조달 수요를 열어 민간의 AI에 대한 인적, 물적 투자까지 견인하고자 한다. 미국이 자유방임국이라는 인식과 달리 방위, 우주항공, 제약, 인터넷 등 핵심 산업마다 공공 영역의 주도로 시장을 조성 · 발전시켰음을 상기시킨다.

이 점은 빅테크의 압도적인 공세 앞에 국내 개발사들의 생존 기반을 마련해 줘야 하는 과제를 안고 있는 우리에게 시사하는 바 크다. EU식 보호주의적 규제로 시간이라도 벌어보겠다는 수세적 대응은 오히려 규제의 집행이 용이한 국내 개발자들만 옥죄는 역설을 초래할 수 있다. 대신 보안요건 등으로 인해 빅테크의 접근이 상대적으로 어려운 공공 영역의 AI 조달을 통해 이 목표를 달성해야 한다. 또한 공공 영역은 적절한 규제(조달 절차에서의 인증 요구로 구체화)를 가함으로써 공무원의 책임 부담을 줄여주고 AI 활용을 가능하게 하면서 시장을 열 수 있는 측면이 있다. 이 목표를 달성하기 위해 재량행위 · 기속행위라는 전혀 다른 목적을 가지고 형성된 법적 요건으로 AI의 금지 · 허용 여부를 좌우하는 독소조항인 행정기본법 제20조를 폐지[124]하고 대신 관리예산실의 M-24-10을 적극 참고하여 행정절차법(또는 전자정부법)을 개정함으로써, ① 각 행정기관은 관련 법률에 따라 금지되거나 제한되지 아니하면 행정작용에 AI를 활용할 수 있고, AI의 활용을 통해 효율성 · 정확성 · 일관성 · 예견가능성 등을 향상시킬 수 있는 행정작용의 용례를 발굴하여 AI를 적용하도록 적극 노력하고, ② 적절한 부처가 범용모델과 클라우드를 통합 구축하고 행정각부에 제공하며, ③ 각 행정기관은 AI의 용례를 제출 · 공개하고, ④ 각 행정기관은 AI 기반 행정작용 시 권익영향AI의 차별 위험에 대한 영향평가와 설명의 제공, 안전영향AI의 안전 위험에 대한 영향평가, 보안관리조치를 해

124) 행정기본법 제20조에 대한 비판으로는 박상철, “인공지능의 법적 규율 II: 판별모델”, 서울대학교 법학 65권 2호(2024. 6.), 338~344.

야 함을 명기[125]할 필요가 있다. 행정기관의 AI 활용을 진작함으로써 행정의 AI 전환과 그로 인한 효율화, 국민의 편의도 기할 수 있음은 물론이다.

XI. 결어

이상 AI에 관한 압도적 개발 · 사업화 역량을 갖춘 미국이 중국 등 경쟁국에 맞서 경제 · 안보적 패권을 유지하기 위해 지난 1년간 총력을 기울여 어떻게 법제와 정책을 정비했는지 조망하였다. 북핵, 불안한 동아시아 정세, 성장 동력의 상실, 저출생 등 실존적 위기나 다름 없는 냉엄한 현실을 맞이하여 AI 등 신기술을 기민하게 수용 · 전개하여 살 길을 도모해야 하는 우리가 이 과정에서 생산된 방대한 자료들에 대해 그간 충분한 연구 · 검토를 하지 못하였음은 아쉬운 일이다. 상기 시사점을 되새기고 우리 현실에 맞춰 소화하여 입법과 정책을 합리화함으로써 AI의 개발, 활용, 신뢰성 확보에 민관의 총력을 동원하며 미래를 도모할 기틀을 마련하기를 바란다.

125) 상세 및 이를 반영한 행정절차법 개정안은 박상철(주 124), 363~366.

<참고문헌>

박상철, “인공지능의 법적 규율 I: 범용모델과 생성모델”, 서울대학교 법학 65권 1호 (2024. 3.)

_____, “인공지능의 법적 규율 II: 판별모델”, 서울대학교 법학 65권 2호(2024. 6.).

이재민, “국제경제협정에서 ‘비차별 원칙’의 공식적 포기인가?–‘공급망 재편’ 논의의 법적 함의”, 국제거래법연구 32권 2호(2023. 12.).

Park, Sangchul, “Bridging the Global Divide in AI Regulation: A Proposal for a Contextual, Coherent, and Commensurable Framework”, 33(2) Wash. Int’l L. J. 216, 234–238 (2024).

Pearce, Hammond, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, and Ramesh Karri, “Asleep at the Keyboard? Assessing the Security of GitHub Copilot’s Code Contributions”, 2022 IEEE Symposium on Security and Privacy (2022).

Urbina, Fabio, Filippa Lentzos, Cédric Invernizzi, and Sean Ekins, “Dual Use of Artificial-intelligence-powered Drug Discovery”, 4 Nat. Mach. Intell. 189 (2022).

AI and Tech Talent Task Force, “Report to the President: Increasing AI Capacity Across the Federal Government: AI Talent Surge Progress and Recommendations”, Apr. 2024.

AISI, NIST, “Managing Misuse Risk for Dual-Use Foundation Models” (NIST AI 800-1; Initial Public Draft), Jul. 2024.

Bengio, Yoshua et al., “International Scientific Report on the Safety of Advanced AI: Interim Report”, DSIT 2024/009 (2024).

CEA, “Potential Labor Market Impacts of Artificial Intelligence: An Empirical Analysis”, Jul. 2024.

CFPB/DOJ/EEOC/FTC/DOE/HHS/HUD/DOL, "Joint Statement on Enforcement of Civil Rights, Fair Competition, Consumer Protection, and Equal Opportunity Laws in Automated Systems", Apr. 4, 2024.

CHCOC, "Skills-Based Hiring Guidance and Competency Model for Artificial Intelligence Positions", Apr. 29, 2024.

CRS, "Highlights of the 2023 Executive Order on Artificial Intelligence for Congress", last updated Apr. 3, 2024.

DHS, "Artificial Intelligence Roadmap", Mar. 17, 2024.

____, "Department of Homeland Security Report on Reducing the Risks at the Intersection of Artificial Intelligence and Chemical, Biological, Radiological, and Nuclear Threats", Apr. 26, 2024.

____, "Establishment of the Artificial Intelligence Safety and Security Board", 89 Fed. Reg. 33380, Apr. 29, 2024.

____, "FACT SHEET: DHS Deploys Pilot Programs, Expands Efforts to Leverage AI to Secure Nation's Critical Infrastructure", Jul. 26, 2024.

____, "Intellectual Property Rights Center (IPR Center)", 〈https://www.iprcenter.gov/referral/report-ip-theft-form〉.

____, "Safety and Security Guidelines for Critical Infrastructure Owners and Operators", Apr. 26, 2024.

DOC, "Dual-Use Foundation Models with Widely Available Model Weights", Jul. 2024.

____, "Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters" (15 CFR Part 702), 89 Fed. Reg. 73612, Sep. 11, 2024.

____, "Taking Additional Steps to Address the National Emergency with Respect to Significant Malicious Cyber-Enabled Activities" (15 CFR Part 7), 89 Fed. Reg. 5698, Jan. 29, 2024.

DOE, "AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy", Apr. 2024.

DOE, “Artificial Intelligence Testbeds at DOE”, 〈https://www.energy.gov/cet/artificial-intelligence-testbeds-doe〉.

DOL, “Artificial Intelligence and Automated Systems in the Workplace under the Fair Labor Standards Act and Other Federal Labor Standards”, Apr. 29, 2024.

____, “Artificial Intelligence and Equal Employment Opportunity for Federal Contractors”, Apr. 29, 2024.

____, “Artificial Intelligence and Worker Well-being: Principles and Best Practices for Developers and Employers”, Oct. 19, 2024.

DOS/USAID, “AI in Global Development Playbook”, Sep. 2024.

DOS/USAID/DOL/DOE/NSF, “The Global AI Research Agenda”, Sep. 2024.

DOT, “DOT Launches New Advisory Committee to Help Shape Federal Approach to Transportation Innovation”, Dec. 29, 2023.

____, “Opportunities and Challenges of Artificial Intelligence (AI) in Transportation; Request for Information, 2024-09645” (89 Fed. Reg. 36848), May 3, 2024.

____, “US Department of Transportation (DOT) Compliance Plan for OMB Memorandum M-24-10 (September 2024)”, Sep. 24, 2024.

EEOC, “High Tech, Low Inclusion: Diversity in the High Tech Workforce and Sector 2014 – 2022”, Sep. 2024.

FedRAMP, “Emerging Technology Prioritization Framework”, Jan. 26, 2024.

FTC, “A Look Behind the Screens: Examining the Data Practices of Social Media and Video Streaming Services”, Sep. 2024.

GAO, “Artificial Intelligence: Agencies Are Implementing Management and Personnel Requirements”, Sep. 2024.

Glasbrenner, James et al., “Dioptra Test Platform”, NIST, 〈https://doi.org/10.18434/mds2-3398〉.

GSA, “Generative AI and Specialized Computing Infrastructure Acquisition Resource Guide”, 〈https://itvmo.gsa.gov/genai〉.

HHS, “HHS Announces 2024 LEAP in Health IT Awardees Focused on Data Quality for Responsible AI and Accelerating Adoption of Behavioral Health IT”, Sep. 17, 2024.

____, “HHS Shares its Plan for Promoting Responsible Use of Artificial Intelligence in Automated and Algorithmic Systems by State, Local, Tribal, and Territorial Governments in the Administration of Public Benefits”, Apr. 29, 2024.

____, “Meeting on the Artificial Intelligence in Healthcare Safety Program”, Oct. 31, 2024.

____, “Nondiscrimination in Health Programs and Activities” (45 CFR. Part 92), 89 Fed. Reg. 37522, May 6, 2024.

HHS, “Screening Framework Guidance for Providers and Users of Synthetic Nucleic Acids”, Oct. 2023.

HUD, “Guidance on Application of the Fair Housing Act to the Screening of Applicants for Rental Housing”, Apr. 29, 2024.

_____, “Guidance on Application of the Fair Housing Act to the Advertising of Housing, Credit, and Other Real Estate-Related Transactions through Digital Platforms”, Apr. 29, 2024.

_____, “HUD Issues Fair Housing Act Guidance on Applications of Artificial Intelligence”, May 2, 2024.

NASEM, “Assessing and Navigating Biosecurity Concerns and Benefits of Artificial Intelligence Use in the Life Sciences-August Information Gathering Meeting”, Aug. 13-14, 2024.

NIJ, DOJ, “Request for Input From the Public on Section 7.1(b) of Executive Order 14110, ‘Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence’”, 89 Fed. Reg. 31771, Apr. 25, 2024.

NIST, “A Plan for Global Engagement on AI Standards” (NIST AI 100-5), Jul. 2024.

_____, “AI RMF: GenAI Profile” (NIST AI 600-1), Jul. 25, 2024.

NIST, “Guidelines for Evaluating Differential Privacy Guarantees (NIST SP 800-226 ipd), Dec. 11, 2023.

_____, “Reducing Risks Posed by Synthetic Content – Draft for Public Comment” (NIST AI 100-4), Apr. 2024.

_____, “Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities” (NIST SP 800-218), Jul. 2024.

_____, “Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile” (NIST SP 800-218A), Jul. 26, 2024.

_____, “U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation With Anthropic and OpenAI”, Aug. 29, 2024.

NSCEB, “White Paper 1: Introduction to AI and Biotech”, Jan. 2024.

NSF, “National Artificial Intelligence (AI) Research Institutes” (NSF 23-610), Aug. 1, 2023.

___, “National Artificial Intelligence Research Resource Pilot”, Jan. 24, 2024.

___, “NSF and DOE Establish a Research Coordination Network Dedicated to Enhancing Privacy Research”, Feb. 26, 2024.

___, “NSF Announces New AI Test Beds Initiative to Advance Safety and Security of AI Technologies”, Jul. 23, 2024.

___, “NSF Engines Eligibility Update”, Sep. 18, 2024.

___, “NSF Launches New Investment to Accelerate the Transition of Privacy-enhancing Technologies to Practice”, Jun. 27, 2024.

___, “NSF-led National AI Research Resource Pilot Awards First Round Access to 35 Projects in Partnership with DOE”, May 6, 2024.

NSTC, “Framework for Nucleic Acid Synthesis Screening”, Apr. 29, 2024.

OET, “Designing for Education with Artificial Intelligence: An Essential Guide for Developers”, Jul. 2024.

OMB, “Advancing Governance, Innovation, and Risk Management for Agency

Use of Artificial Intelligence" (M-24-10), Mar. 28, 2024.

OMB, "Advancing the Responsible Acquisition of Artificial Intelligence in Government" (M-24-18), Sep. 24, 2024.

_____, "Request for Information: Executive Branch Agency Handling of Commercially Available Information Containing Personally Identifiable Information", 89 Fed. Reg. 83517, Oct. 16, 2024.

_____, "Request for Information: Privacy Impact Assessments", 89 Fed. Reg. 5945, Jan. 30, 2024.

OPM, "Artificial Intelligence (AI) Competency Model for Civil Engineering, 0810", Apr. 29, 2024.

_____, "New Talent Pools (Shared Certs) Feature for USAJOBS' Agency Talent Portal", Dec. 1, 2023.

_____, "Pay Flexibility, Incentive Pay, and Leave and Workforce Flexibility Programs for Artificial Intelligence (AI), AI-enabling, and Other Key Technical Employees", Feb. 27, 2024.

_____, "Responsible Use of Generative Artificial Intelligence for the Federal Workforce", 〈https://www.opm.gov/data/resources/ai-guidance〉.

PCAST, "Supercharging Research: Harnessing Artificial Intelligence to Meet Global Challenges", Apr. 2024.

Tripathi, Micky, "Leveraging Agency Expertise to Foster American AI Leadership and Innovation", Dec. 19, 2023.

U.K. Department of Science, Innovation, and Technology, "Emerging Processes for Frontier AI Safety", Oct. 2023.

USCO, "Copyright and Artificial Intelligence – Part 1: Digital Replicas", Jul. 2024.

USDA, "Framework for State, Local, Tribal, and Territorial Use of Artificial Intelligence for Public Benefit Administration", Apr. 29, 2024.

USDT, "Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector", Mar. 2024.

USDT, Provisions Pertaining to U.S. Investments in Certain National Security

Technologies and Products in Countries of Concern, 31 C.F.R. Part 850, Nov. 15, 2024.

USPTO, "2024 Guidance Update on Patent Subject Matter Eligibility, Including on Artificial Intelligence", 89 Fed. Reg. 58128, Jul. 17, 2024.

_____, "Inventorship Guidance for AI-Assisted Inventions", Feb. 13, 2024.

_____, "Manual of Patent Examining Procedure (MPEP) – 9th Edition", Jan. 2024.

White House, "A Call to Action to Combat Image-Based Sexual Abuse", May 23, 2024.

_____, "Artificial Intelligence, Automation, and the Economy", Dec. 12, 2016.

_____, "Biden-Harris Administration Announces Key AI Actions 180 Days Following President Biden's Landmark Executive Order", Apr. 29, 2024.

_____, "Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order", Oct. 30, 2024.

_____, "Framework to Advance AI Governance and Risk Management in National Security", Oct. 24, 2024.

_____, "Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence", Oct. 24, 2024.

_____, "White House Announces New Private Sector Voluntary Commitments to Combat Image-Based Sexual Abuse", Sep. 12, 2024.

VA, "AI Tech Sprint", May 24, 2024.

43rd Republican National Convention, "The 2024 Republican Platform-Make American Great Again!", Jul. 15, 2024.

経済産業省/総務省, AI事業者ガイドライン(第1.0版), 2024年 4月 19日.

〈Abstract〉

U.S. Federal Implementation of Executive Order 14110: Progress and Implications

Shin Wonjun* · Yang Soobin** · Lee Daejun*** · Park Sangchul****

The Biden administration's "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" assigned detailed tasks to federal agencies across various sectors, including safety and security, innovation and competition, worker protection, civil rights, consumer protection, privacy, federal use, and international cooperation. Now, a year later, most of these tasks have been implemented, with the following key takeaway. First, the U.S. is pursuing strategic and industrial superiority through the mobilization of federal capabilities comparable to wartime efforts, as well as clarifying its national security-focused approach towards AI supply chain. Second, while avoiding the EU-style horizontal regulatory framework, the U.S. is pursuing sector- and context-specific regulations by assigning detailed goals to each federal agency according to their jurisdictions. At once, the U.S. is ensuring coherency through strong coordination linking regulations to budget planning and a flexible division of roles between hard regulations and soft standards. Third, with respect to the federal use of AI, it seeks to encourage competition among agencies in AI adoption and open up demands for workforce and procurement through appropriate regulation. Korea can draw several lessons from these developments. First, in the realm of AI, economics and security cannot be separated, and

* Seoul High Court. The authors contributed equally.
** Kim & Chang. Same as above.
*** Kim & Chang. Same as above.
**** Corresponding author. Seoul National University School of Law, parks@snu.ac.kr.

national interests must be pursued in tandem with global developments including the U.S. national security policy. Second, instead of adopting the EU-style horizontal and comprehensive regulations, Korea should seek a rigorous regulatory framework that combines: (i) flexible soft standards, (ii) targeted, tailored regulations for specific use cases governed by each agency, and (iii) coherent coordination linked to budget planning. Third, to secure breathing room for domestic developers and to accelerate the public sector's AI transition, Korean should establish institutional foundations for boosting AI public procurement demand.

Keywords: AI, Executive Order, Safety, Security, Governance, Foundation Model