



배포: 일반
2025년 6월 16일

원본: 영어

번역: 정보인권연구소(이 문서는 기계번역 후 사람이 검수하고 교정하였습니다. 각주 제외)

인권 이사회

제59차 회의

2025년 6월 16일~7월 11일

의제 항목 2 및 3

유엔 인권최고대표의 연례 보고서,

최고대표사무소 및 사무총장의 보고서

모든 인권, 시민적·정치적·경제적·사회적·문화적

권리(발전권 포함)의 증진 및 보호

기술 기업의 활동(AI 관련 활동 포함)에 대한 기업과 인권 이행원칙의 실제적 적용

유엔 인권 최고대표사무소 보고서*

요약

본 보고서는 인권이사회 결의 53/29에 따라 제출된 것으로, 이사회는 유엔 인권최고대표사무소에 <기업과 인권 이행원칙>의 실제 적용에 관한 보고서, 특히 AI 관련 활동을 포함한 기술 기업의 활동에 이행원칙을 적용하는 데 있어 직면하는 과제, 모범 사례 및 교훈에 관한 보고서를 작성하여 제59차 이사회에 제출할 것을 요청하였습니다. 본 보고서는 결의안 53/29에 따라 1.5일에 걸쳐 개최된 전문가 자문 회의에서 논의된 내용을 바탕으로 작성되었습니다.



I. 서론

1. 인권이사회는 디지털 신기술과 인권에 관한 결의 53/29에서 디지털 기술이 인권에 미치는 영향을 다루었습니다. 특히 해당 기술이 인권과 기본적 자유를 존중하는 방식으로 개발 및 사용되도록 보장하기 위해서는 국제적·다자간 협력이 필요하다는 점을 강조하고, 각국에는 <기업과 인권 이행원칙>을 이행할 것을, 기업들에게는 이를 기업 운영에 통합할 것을 촉구하였습니다.
2. 이 결의에 따라 유엔 인권최고대표사무소(OHCHR)은 2024년 11월 28일과 29일에 전문가 자문회의를 소집하여 AI(AI)을 설계, 개발 및 배치하는 국가와 기업이 <기업과 인권 이행원칙>을 적용하여 인권 보호와 존중을 강화할 수 있는 방안을 모색했습니다. 본 보고서는 전문가 자문과 OHCHR의 작업을 바탕으로 작성되었으며, 디지털 기술, 특히 AI가 인권에 미치는 부정적 영향을 예방하고 해결하는 데 있어 이행원칙이 지닌 가치와 실제 적용을 제시합니다. 본 보고서는 또한 기술 분야에 이행원칙을 적용하는 일과 관련된 기존의 인권이사회 제출 보고서를 참고하여 작성되었습니다.

II. AI 관련 활동을 포함한 기술 분야에 대한 <기업과 인권 이행원칙>의 적용

A. 배경

3. AI 혁신의 급속한 가속화는 21세기의 특징이 되어 노동 세계를 비롯한 다양한 영역을 변화시키고 있습니다. 이러한 기술적 도약은 전례 없는 발전을 약속하는 동시에 인권 보호에 중대한 과제와 위험을 가져옵니다. '미래를 위한 협약'은 AI 혁신이 2030 지속가능발전목표 달성에 핵심적이라고 강조했습니다. 그러나 인권 보호 장치가 내재되지 않은 AI는 추구하는 결과를 달성하지 못할 뿐만 아니라, 오히려 발전을 저해하고 평화와 안보를 훼손할 수 있습니다. AI 개발 및 배치 과정에서 국가의 역할과 의무, 기업의 책임을 명확히 하는 것은 책임 있는 AI 혁신 생태계를 보장하고 인류를 이롭게 하는 데 핵심적입니다.
4. 본 보고서에서는 경제협력개발기구(OECD)가 정의한 AI, 즉 명시적 또는 암묵적 목표를 위해 입력된 데이터로부터 예측, 콘텐츠, 권고 또는 물리적 다양한 공론의 장에서 현재부터 중기적 시점까지 AI으로 인해 인간에게 발생하는 부정적 영향은 물론 사회적 또는 존재적인 위험을 야기할 수 있는 장기적이고 복잡한 문제들이 주목받아 왔습니다. AI의 존재적 함의를 이해하고 완화하는 일이 중요하긴 하지만, 생성형 AI와 같은 다양한 하위 AI 기술의 현재적 영향이 이미 존재해 왔습니다. 생성형 AI는 입력 프롬프트를 통해 콘텐츠(텍스트, 이미지, 음악, 영상)를 생성할 수 있는 증력을 가지고 있으며, 광범위한 인권이 부정적 영향을 받을 수 있다는 사실이 드러났습니다.

이러한 영향에는 편향과 차별, 사생활 침해, 노동 시장 교란 등이 포함됩니다. 자율 무기 체계에서 AI를 활용하는 것도 막대한 위험을 야기할 가능성이 있습니다. 이 모든 영역에서 AI 시스템의 불투명성은 해당 문제들을 모니터링하고 완전히 이해할 수 있는 길을 복잡하게 만듭니다.

5. 민간 부문은 기업과 정부에 제품을 제공하고 글로벌 AI 거버넌스 형성에 상당한 영향력을 행사하면서 AI 혁신을 주도해왔습니다. <기업과 인권 이행원칙>은 AI 및 신기술 환경에서 인권을 옹호하기 위한 국가와 기업의 역할을 정의하는 데 있어 핵심적인 체계를 구성합니다. 국가와 기업에 공평한 환경을 조성함으로써, 이행원칙은 세계적으로 AI의 위험을 완화하고 혜택을 증진할 수 있습니다. 그러나 AI에 이행원칙을 적용하는 일은 디지털 인프라 거버넌스와 그 기반이 되는 가치 체계가 국제 인권법(조약상 의무 및 관습법 포함)과 어떻게 조화를 이룰 수 있을지 더 넓은 맥락과 논의 속에 이루어져야 합니다.
6. AI 기술 발전이 가져온 새로운 문제들 외에도, 기술 분야 및 기타 영역에서 데이터가 수집, 처리 및 분석되는 방식과 관련된 근본적인 문제들이 지속되고 있습니다. AI모델은 개발, 학습 및 배치를 위해 대규모 데이터 세트에 의존합니다. 이러한 데이터, 특히 개인정보의 불투명한 출처는 개인정보 보호 및 사생활 침해 우려를 불러왔습니다. 학습 데이터가 사용자 동의 없이 수집되었거나, 향후 모델 출력에서 재노출될 수 있는 민감정보를 포함하고 있을 수 있습니다. 또한 많은 국가에서 AI 가치사슬 내 데이터 정제 노동자들의 노동조건은 저임금과 정신 건강 보호 부족 등 착취적 관행으로 인해 심각한 비판을 받고 있습니다.
7. 글로벌 거버넌스 환경은 AI가 가져온 사회적 문제와 기회를 다루는 데 진전을 보여 왔으며, 규범적 기반, 규율성, 깊이, 부문별 초점 및 지리적 범위 측면에서 다양한 접근법을 취하고 있습니다. 이러한 노력은 칭찬할 만하지만, AI의 급속한 확장에 비하여 범위와 정교함 모두에서 여전히 파편적이고 상대적으로 초기인 단계에 머물러 있습니다. 많은 개발도상국들은 AI 전략이 부재하며, 대다수 국가들이 거버넌스 이니셔티브 측면에 공백이 있습니다. 유엔 총회는 민간 부문이 <기업과 인권 이행원칙>을 준수할 것을 권장해 왔으며, '글로벌 디지털 협약'은 AI 거버넌스 지침으로서의 그 가치가 강조되고 있습니다.

유네스코, OECD, 유럽평의회를 비롯한 국제 및 지역 기구들은 AI 거버넌스의 지침이 되는 프레임워크를 도입해 왔는데, 이는 권리 기반 접근법(rights-based approaches)과 기업의 책임을 강조하고 있습니다. 유럽연합 AI법에는 AI 위험 분류 체계가 포함되어 있으며, 남미 공동시장(Southern Common Market)과 같은 지역 기구들은 거버넌스에 인권 원칙을 통합하고 있습니다. 다양한 국가별 AI 규정 및 전략들이 이러한 광범위한 노력들을 보완하고 있습니다.

B. AI 혁신과 인권의 결합: 기업이 인권을 존중할 책임

8. <기업과 인권 이행원칙>은 모든 기업이 인권 존중의 책임을 지며, 기업 활동으로 인권에 부정적 영향을 야기하거나 이에 기여하지 않을 것을 요구합니다. 또한 기업은 사업 관계로 인해 운영, 제품과 서비스와 직접 연관된 부정적 인권 영향이 발생할 경우, 해당 영향에 기여하지 않았더라도 이를 예방하거나 완화하기 위해 노력해야 합니다.
9. 따라서 AI 개발 및 배치에 관여하는 기업들은 운영 전반에 걸쳐 인권을 보호하기 위한 조치를 시행해야 합니다. 무엇보다도, <기업과 인권 이행원칙>은 기업들이 그 의사결정, 위험 평가 및 운영 과정에 인권 고려 사항을 통합할 수 있도록 보장하기 위하여, 인권 정책을 마련할 것을 요구합니다. 이행원칙은 또한 기업들이 인권에 대한 잠재적 위험을 식별, 완화 및 해결하기 위한 지속적인 업무 관행으로 인권실사(due diligence)를 실시해야 한다고 밝히고 있습니다.

1. AI 제품과 서비스에 대한 인권실사

10. 인권실사는 기업이 잠재적이고 실제적인 부정적 인권 영향을 적극적으로 관리하는 데 있어 필수적입니다. 핵심 기대사항에는 인권 위험의 식별 및 평가, 기업 절차에 대한 평가 결과의 통합, 적절한 조치 시행, 조치의 효과성 추적, 영향이 해결된 방식에 대한 조직 소통 등이 포함됩니다. <기업과 인권 이행원칙>에 따르면 인권 영향은 주기적으로 평가해야 하며, 이는 예를 들어 새로운 AI 제품과 서비스를 도입하기 전이나 AI 제공 방식에 중대한 변경이 있기 전과 같은 때를 말합니다. 이는 특히 이러한 변경이나 새로운 제품이 예측 불가능한 인권 영향을 야기할 수 있기 때문입니다. OHCHR는 기술 분야에서 이행원칙을 적용하기 위한 실무 지침을 마련한 바 있습니다.

11. 인권실사는 조직과 상황에 따라 다르게 나타날 수 있습니다. 그럼에도 불구하고, AI 제품과 서비스와 관련된 사람들에게 미치는 위험을 다룰 때 특히 중요한 몇 가지 기본 요소가 있습니다.

(a) 시점 및 영향력: 인권실사는 제품 설계, 개발 및 사용 전반에 걸쳐 조기에 그리고 자주 실시되어야 하며, 기술 기업들이 다른 행위자들의 행동과 관행에 영향을 미치기 위해 영향력을 구축하고 행사할 것을 불가피하게 요구할 것입니다.

(b) 인식과 입증: 기업은 자사 AI 제품과 서비스 사용으로 발생할 수 있는 부정적 영향에 어떻게 대응해야 하는지 이해하고 입증할 수 있어야 합니다.

(c) 관여 프레임워크: 기술 기업은 자사 AI 제품과 서비스가 최종적으로 사용되는 상황에서 인권침해를 야기하거나 기여하거나 연관될 수 있습니다. 실제로 기술 기업의 영향에 대한 관여도는 자체적인 행동, 소홀함 및 모범 사례의 기준 발전에 따라 시간이 지나면서 변화할 수 있습니다.

(d) 가능한 영향에 대한 포괄적 관점 유지: 기업의 모든 사업 활동 및 관계 전반에 걸쳐 인권 관련 위험을 식별하는 일은 제품과 서비스의 설계, 개발, 홍보, 판매, 라이선싱, 계약 체결 및 사용이 인권에 부정적 영향을 야기할 수 있는지 여부와 그 방식을 포괄합니다.

(e) 심각도에 따른 우선순위 설정: 초점의 우선순위 설정은 기업의 사업 활동이 사람들에게 가장 심각하고 광범위하며 지속적인 피해를 야기하거나 야기할 수 있는 지점을 기준으로 해야 합니다. 기업이 영향에 대응하는 조치를 취할 때 집중해야 할 우선순위를 설정하는 경우, 사람에게 미치는 위험의 심각성에 초점을 맞춘 원칙적 접근법을 취해야 합니다. <기업과 인권 이행원칙>에 따르면, 영향의 심각성은 그 규모, 범위 및 회복 불가능성에 따라 판단되어야 합니다. 규모는 영향이 얼마나 중대하거나 심각한지를, 범위는 영향이 얼마나 광범위하거나 영향 받은 사람들의 수가 얼마인지를, 회복 불가능성은 영향 받은 사람들을 영향 발생 전의 상황과 동일하거나 동등한 상태로 복원할 수 있는지를 의미합니다.

12. AI 연구 활동에서 잠재적 인권 영향을 고려하는 것은 AI를 개발하거나 배치하는 기업이 제품 개발 과정에서 예방적 관점을 적용할 수 있도록 합니다. 모든 인권실사는 상황에 따라 달라지므로, 생성형 AI 및 AI에이전트와 같은 기술은 배치 전에 모든 위험을 예측할 수 없고 소위 기술 스택 생태계 전반에 걸쳐 확산될 수 있으므로 상황에 맞는 테스트 및 위험 관리 절차가 필요합니다. 이러한 위험을 해결하기 위해서는 다른 조치들과 더불어 가치사슬 간 협력이 필요한데, 이는 개발자가 상황별 위험을 간과하거나 배치자가 AI모델의 기능을 오해할 수 있기 때문입니다.

13. 인권 전문가의 자문을 반영한 독립적 감사는 감독과 책무성을 강화합니다. 성별 영향 평가를 비롯하여 인권 성과 및 위험에 대한 공개적 보고를 통해 투명성이 유지됩니다. 내부 역량 강화는 AI 개발팀이 전문 교육을 받고 인권 문제에 대한 전문성을 함양하도록 보장합니다.

14. 머신러닝과 같은 데이터 기반 AI 시스템은 본질적으로 부정확성의 위험을 내포하고 있으며, 이를 수정하는 일이 복잡할 수 있습니다. 많은 일이 모델 학습에 사용된 데이터에 달려 있습니다. <기업과 인권 이행원칙>에서는 비례성을 강조합니다. 위험이 커질수록 절차가 더 복잡해야 합니다. AI 혁신은 주로 민간 부문이 주도하고 있으므로, 민간 부문은 책임 있는 AI를 보장하는 데 있어

핵심적인 역할을 합니다. 범용 AI도 인권적인 우려를 낳고 있는데, 이는 다양한 적용 분야와 불투명한 의사 결정, 급속한 배치, 내재된 데이터 편향성, 최소화된 인간 감독, 유해 콘텐츠 생성 및 개발과 사용 과정에서 나타난 불분명한 책무성 문제로 인한 것입니다.

2. AI 생태계의 비즈니스 모델

15. 기업은 모든 사업 활동과 관계에서 인권실사를 실시해야 합니다. 여기에는 비즈니스 모델 중심의 관행과 AI 설계 결정이 인권 위험을 발생시키거나 악화시키는 상황을 해결하는 일도 포함됩니다. 경영진과 기업가는 인권실사를 비즈니스 모델의 회복탄력성을 혁신하고 강화하는 기회로 인식해야 합니다. 인권 보호를 최우선으로 하는 AI 비즈니스 모델에 대한 투자가 여전히 부족합니다. AI 제품이 때때로 인권 문제의 해결책으로 제시되기도 하지만, 그 효과성은 인프라가 인권을 보호하도록 설계되었는지에 달려 있습니다.

16. 보다 책임 있는 접근법은 AI 투자 전략과 그 결과로 도출된 비즈니스 모델을 인권과 조화시킬 것을 요구합니다. 비즈니스 모델의 위험을 해결하기 위한 모범 사례는 다음과 같습니다. (a) 최고 경영진 및 핵심 기능에 대한 성과 인센티브를 검토하여 인권 존중 행동을 보상하였습니다. (b) 수익 극대화만을 위한 최적화의 대척점으로 인권침해 위험을 최소화하기 위한 스트레스 테스트를 AI 제품과 서비스 설계에 대하여 실시하며, 신규 시장에서 AI 제품과 서비스를 테스트하고 확장하는 계획을 면밀히 검토하였습니다.

3. AI 제품과 서비스에 대한 이해관계자 참여

17. 의미 있는 이해관계자 참여는 인권실사의 초석이며, 다양한 상황별 위험을 고려하고 지리적 맥락 및 지역 현실에 맞게 조정되어야 합니다. 직원, 지역사회 구성원, 공급망 노동자, 관련 전문가 및 소비자를 포함한 광범위한 이해관계자의 참여는 AI를 개발하고 배치하는 기업이 자사 AI 제품과 서비스와 관련된 구체적인 우려 사항을 더 잘 이해하고 해결할 수 있도록 합니다. 참여는 포용적이며 문화적·언어적인 감수성을 갖추어야 하며, 소외되고 취약한 집단이 유의미하게 참여할 수 있도록 보장해야 합니다. 그들의 관점은 AI와 연관된 잠재적 또는 실제적 피해를 식별하고 완화하는 전략을 수립하는 데 핵심적입니다. 고려해야 할 다섯 가지 핵심 측면은 다음과 같습니다.

- (a) 인권실사 및 구제 조치의 모든 측면에 걸쳐 외부 이해관계자와의 협력
- (b) 이해관계자의 의견이 비즈니스 관행 및 결정에 어떻게 반영되었는지에 대한 피드백 루프 구축
- (c) 거래가 아닌 관계의 구축 및 육성
- (d) 글로벌 남반구(Global South)의 목소리를 참여시키기 위한 협력
- (e) 기업의 인권실사 과정에 이해관계자가 참여할 수 있도록 만드는 자원의 지원. 적절한 안전장치가 있다면 AI 자체가 이러한 참여를 지원할 수 있습니다.

4. AI와 관련하여 기업 관행에 <기업과 인권 이행원칙>을 반영하기 위한 현재의 노력

18. 현재까지 기술 기업들이 AI 제품과 서비스에 <기업과 인권 이행원칙>을 어떻게 반영하고 있는지,

그리고 해당 제품과 서비스에 대한 인권실사 절차가 얼마나 견고한지 평가하는 포괄적인 분석은 존재하지 않습니다. 이는 부분적으로 기업들의 데이터 공개가 부족한 까닭이기도 하고, 인권 영향이 본질적으로 맥락적이기 때문입니다. 이러한 맥락성과 관련된 방법론적 함의는 다양한 기업 모델과 조치를 서로 비교하는 데 어려움을 초래할 수 있습니다. 기업 행태를 벤치마킹하려는 계획들이 존재합니다.

19. 점점 더 많은 기술 기업들이 <기업과 인권 이행원칙>을 활용하여 디지털 기술(AI 등 다양한 수준의 기술)이 인권에 미치는 영향을 식별, 평가 및 완화하고 있지만, 그 전체적인 범위는 아직 알려져 있지 않습니다. 이행원칙의 이행은 기술 분야별 원칙 및 이니셔티브로 보완되고 심화되는 경우가 많은데, 이를테면 글로벌 네트워크 이니셔티브의 표현의 자유 및 사생활 보호에 관한 글로벌 원칙과 같은 것이 있습니다. 구글의 유명인 인식 API 인권 평가도 하나의 사례인데, 이는 이행원칙에 기반한 회사의 포괄적 인권 정책에 반영되었습니다. 마이크로소프트는 최근 이행원칙에 기반하여 자사의 기업용 클라우드 서비스 및 AI 기술에 대한 평가를 의뢰하였습니다. 한국 기술 기업 카카오는 이행원칙에 기반한 책임 있는 AI를 구현하기 위하여 자체 평가 도구를 개발하였습니다.

20. 이러한 노력들은 올바른 방향으로 향하는 단계이지만, AI를 개발하고 배치하는 기업들이 <기업과 인권 이행원칙>을 채택하는 것은 여러가지 이유로 아직 초기 단계에 머물러 있습니다.

(a) 기업 내부에서 AI를 규율하는 프레임워크가 중층적으로 존재하며, 이는 대체로 안전과 보안에 중점을 두고 인권을 완전히 반영하고 있지 않습니다.

(b) 인권에 대한 고려가 AI 제품 개발 단계에서 너무 늦게 이루어지는 경우가 많아, 절차 초기에 인권 요소를 반영하기 위한 개선이 필요합니다.

(c) AI를 개발하는 중소기업들도 인권에 상당한 영향을 미칠 수 있지만, 이들은 이행원칙을 적용하는데 필요한 인식을 갖추고 지원을 받기 어려운 경우가 많습니다.

(d) 데이터 출처, 큐레이션 및 후속 AI모델 출력으로 인한 인권 문제를 해결하는 데 필요한 기존 기업 운영 관행에 대한 투명성 부족 문제가 계속됨. 투명성 증대는 업계 간 모범 사례 공유를 가능하게 하고 선순환 경쟁을 촉진할 수 있습니다.

21. 또한 현재의 지정학적 환경은 혁신과 이를 지연시키는 조치 간에 대립을 유발할 위험이 있습니다. 제품 개발 초기 단계에서 인권 보호 장치를 포함하는 것은, 위험을 감소시킬 뿐 아니라 다양한 사용자 집단과 그 환경과의 잠재적 상호작용을 고려함으로써 혁신을 강화할 수 있다는 사실이 인식되지 못하고 있습니다.

22. 기술 기업들이 <기업과 인권 이행원칙> 채택하는 양상이 파편적이라는 점, 특히 표면적으로 인권 및 국제 인도법 측면에서 고위험을 초래하는 사용 사례에 관여하는 기업들이 관련된 분야(예를 들어 무기, 감시, 이민, 군사 분야)의 경우 더욱 우려스럽습니다. 분쟁 지역이나 취약한 국가 환경에 여러 AI 기술을 배치하는 것은 특히 세심한 주의와 더 높은 안전 장치가 필요합니다. 대부분의 AI 기술은, 초기 설계시 민간용이든 군사용이든 본질적으로 이중용도(dual-use)의 속성을

가지고 있으며 쉽게 용도가 변경되어 민간 사용과 군사적 적용 사이의 경계를 모호하게 만듭니다. 예를 들어, 얼굴 인식과 같이 군사적 적용을 위해 처음 개발된 AI 기반 기술들은 법 집행 및 국경 통제 작전과 같은 민간 영역에서 사용되어 왔습니다. 군사 영역에서는 AI 시스템이 민간인을 전투원으로 오인할 수 있어 민간인 보호에 위험을 초래할 수 있습니다.

C. AI 시대의 권리 보호: 기업 행위에 대한 국가의 역할

23. 국제법에 부합하여, <기업과 인권 이행원칙>은 국가의 인권 보호 의무를 재확인하며, 기업 활동에서 비롯되거나 이와 연관된 영향을 포함하여 제3자에 의한 부정적 영향으로부터 인권을 보호할 의무를 강조합니다. 해당 의무는 기업의 AI 기술 개발, 배치 및 사용으로도 확장됩니다. 이행원칙은 기업이 인권을 존중하도록 요구하기 위해 의무적 조치와 인센티브 기반 조치의 '스마트한 조합' 개념을 도입합니다. 이 접근법은 국가적, 지역적, 국제적 프레임워크 전반에 걸쳐 정책적인 일관성을 달성하고 기업 행위에 대한 기대사항을 맞추는 것을 목표로 합니다.

1. AI 관련 규제 동향

24. 생성형 AI를 비롯한 AI에 대한 규제 접근법이 국제 인권 체계에 기반할 경우 개인의 인권이 보호되고 글로벌 차원의 정책적인 일관성이 구축될 수 있습니다. AI 개발, 배치 및 사용 과정에서 민간 부문의 인권 존중 역할과 관련된 조항을 포함하는 최근 규제 노력으로는 브라질, 대한민국 및 유럽연합의 사례가 있으며, 이에 대해서는 아래에서 더 자세히 설명합니다.

25. 규제의 출발점은 보안이나 안전에 초점을 맞춘 일반적인 위험 모델이 아니라, 사람에게 미치는 영향에 초점을 맞추어야 합니다. 일부 규제 체계는 민간 부문과 공공 부문의 AI 기술 사용을 이분법적으로 구분하지만, 양 부문의 상호 연관성을 고려한다면 양측 모두 인권 보호에 주력해야 합니다. 또한 AI 규제는 공공 및 민간 부문 모두에서 투명성과 책무성을 보장하고 인권에 미치는 영향을 완화해야 합니다. 투명성과 밀접하게 연관된 핵심적인 고려사항은 데이터 거버넌스, 특히 사생활과 개인정보 보호에 관한 것입니다.

26. 설계, 개발, 배치부터 최종 사용에 이르는 AI 수명주기에 대한 인권 영향 평가를 포함한 인권실사 절차가 강력한 규제 안에 포함되어야 하며, 이는 모든 범위의 인권을 고려하여 실시되어야 합니다. AI의 광범위한 적용 분야를 고려할 때, 위험을 식별, 평가 및 완화하고 심각한 인권침해를 최우선적으로 완화하기 위해서는 일관된 인권실사 적용이 필수적입니다. AI 가치사슬 전체에 초점을 맞추면 전체론적 접근이 보장될 것이며, AI로 생성된 아동 성 착취물과 같은 특정 피해에 대해서는 표적 개입이 필요할 수 있습니다.

27. 규제의 효과를 높이기 위해서는 규제 노력과 함께 기업을 대상으로 하는 포괄적 인센티브 기반 정책 수단, 집행 규정 및 준수 지원이 동반되어야 합니다.

28. <기업과 인권 이행원칙>에 따라, 투명성과 절차 중심의 법률이 AI와 관련된 다양한 비즈니스 모델과 현실에 민감하게 대응하면서 기업에 기대하는 사항을 명확히 규정해야 합니다. 적절한 통지,

적법 절차 및 차별금지 요건을 통해 기업은 관련 기업 시스템 및 AI 제품과 서비스와 관련된 절차에서 권리를 존중하는 방식으로 해당 조항들을 통합할 수 있을 것입니다.

29. 의무적 실사 및 보고 기준이 등장하면서 기술 기업들이 위험을 식별하고 해결하도록 요구하는 추세로, 이는 <기업과 인권 이행원칙>과 유사한 접근법입니다. 예를 들어, 유럽연합 AI법은 '인권 및 기본적 자유 보호에 관한 협약(유럽인권협약)'에 따라 AI 시스템의 배치 및 사용에서 인권을 보호할 것을 강조합니다. 이 법은 국가 안보나 범죄 수사를 위한 정당한 사유가 없는 한 공공장소에서 실시간 생체 인식 감시와 같은 고위험 AI의 적용을 금지합니다. 고위험 AI 시스템을 배치하는 자는 개인의 권리에 대한 잠재적 위험을 평가하기 위한 평가를 실시하여 책무성과 투명성을 보장해야 합니다. 유럽연합의 '디지털 서비스법'에 따르면 플랫폼 사업자는 콘텐츠 거버넌스 접근법을 포함한 운영 관련 위험을 평가하고 완화해야 합니다. 마찬가지로, 2026년 1월 시행 예정인 한국의 「AI 발전과 신뢰 기반 조성 등에 관한 기본법」 역시 위험 기반 접근법을 채택했습니다. 브라질 하원에서 현재 심의 중인 법안 제22338/2023호는 AI의 개발, 구현 및 사용을 규제하는데, 위험 기반 접근법에 기반한 구성으로 "과도하거나" "고위험"한 응용 프로그램에 대한 추가 규정을 포함하고 있습니다. 다른 일부 국가들도 부문별로 의무적인 인권실사 관련 법률이 제정되었거나 추진 중입니다.

2. AI 관련 정책적 조치

30. 규제는 기업이 권리를 존중하는 AI 접근법을 채택하도록 장려하는 정책적 조치와 밀접하게 상호연관되어 있습니다. 예를 들어, 권리를 존중하는 AI 연구 개발에 대한 정부 자금 지원이 이에 해당합니다. 일부 전문가들은 AI 개발을 촉진하는 공공 자금을 명확히 정의된 공익적 목표를 위해 활용해야 한다고 강조하며, 어떤 유형의 AI가 필요한지, 그리고 어떤 구체적인 사용 사례를 위한 것인지에 투명한 근거를 제시할 것을 요구합니다. 이를 달성하는 한 가지 방법은 책임 있는 기술 및 관행 개발을 장려하겠다는 약속을 산업 정책의 기반으로 삼는 것입니다. 여기에는 책임 있는 AI 및 디지털 인프라 개발을 지원하기 위한 자금 배분도 포함됩니다. 인권 기반 접근법을 통합한 국가 AI 전략에는 책임 있는 기업 활동을 위한 다른 정책 조치들도 통합될 수 있습니다.

31. 또 다른 지렛대는 국가가 민간 부문 및 보건, 교통, 전자정부, 교육, 치안, 안전, 복지, 형사, 사법, 감시 기술 등 다양한 분야에서 AI 제품과 서비스를 조달하는 주체로서의 역할입니다. 공공 조달 절차에 인권 요구사항을 반영하면 기업이 자사 관행을 인권 기준에 맞추어 조정하게끔 유도할 수 있습니다. 또한, 개발금융기구에서 활동하는 국가의 기능으로 디지털 전환 사업에 인권이 포함되도록 보장하는 인센티브를 추가적으로 제공할 수 있습니다.

32. 2022년 네덜란드 정부가 도입한 알고리즘 등록부와 같은 정책적 조치도 책임 있는 행동을 촉진할 수 있습니다. 이 등록부는 공공 기관이 알고리즘 사용에 대해 공개하도록 요구하며, 여기에는 의도된 사용 목적에 대한 설명과 잠재적 인권 영향에 대한 평가가 포함됩니다.

영국과 북아일랜드의 AI 보증 프레임워크(AI Assurance Framework)도 AI 시스템의 신뢰성을 측정하고 평가하기 위한 사이버 보안 및 품질 관리와 함께 유사한 인권 기준을 도입하였습니다.

33. 정부는 국제전기통신연합(ITU)과 같은 기술 표준 기구와 협력하면서 정책 일관성 증진에 핵심적 역할을 수행하는데 더불어 기술 표준 설정 과정 및 AI 상호운용 표준 개발에 인권을 통합할 것을 촉구할 수 있습니다. 인권 원칙을 기술 용어로 정책화하면 전문가들이 이를 AI 개발 전 과정에 반영하는 데 도움이 될 수 있습니다. 이와 관련하여 표준 기구들은 AI 기술 표준에 인권을 통합하는 일의 중요성을 점차 인식하고 있습니다. 예를 들어 국가적 차원에서는 미국 국립표준기술연구소(NIST)가 신뢰할 수 있고 책임 있는 AI(Trustworthy and Responsible Artificial Intelligence) 이니셔티브에서 인권의 중요성을 강조하며 기술 표준을 인권 프레임워크와 조화시키는 데 기여하고 있습니다.

D. AI 등 기술 사용으로 인한 인권침해에 대한 책무성과 구제

34. AI 등 기술 사용으로 인한 인권침해에 대응하는 구제 메커니즘, 절차 및 전략에 인권을 통합하는 일은 필수적입니다. <기업과 인권 이행원칙>의 제3원칙은 AI 제품과 서비스 등 기술로 인한 인권침해를 구제하기 위한 체계를 제공합니다. 이 체계는 국제 인권법으로 확립된 바 있는 효과적으로 구제받을 권리에 기반을 두고 있습니다.

35. AI 제품과 서비스 등 기술 사용으로 인한 인권침해에 대한 구제수단을 마련하는 데에는 다양한 메커니즘과 절차가 기여합니다. <기업과 인권 이행원칙>에서는 효과적인 구제를 보장하기 위하여 사법 제도, 국가적 비사법적 제도, 비국가적 고충 처리 제도 등 세 가지 유형의 메커니즘을 제시하고 있습니다. 이행원칙에서는 국가가 관할권 내에서 발생하는 이러한 침해를 예방, 조사, 처벌 및 구제하기 위하여 적절한 조치를 취할 것을 촉구합니다. 기업 또한 AI 제품과 서비스와 관련된 경우를 비롯하여 자사가 부정적 영향을 초래하거나 기여한 경우 구제 절차에 협력하거나 이를 제공해야 합니다.

36. 영향 받는 개인 및 이해관계자 집단이 자신들의 다양한 경험에 부응하고 <기업과 인권 이행원칙>의 기준에 부합하는 효과적 구제수단에 접근할 수 있도록 보장하는 일이 가장 중요합니다. 구제수단에 대한 접근성 확보는 산업 분야나 운영 환경을 불문하고 늘 어려운 문제일 수 있지만, 특히 디지털 기술과 관련해서 더욱 어려운 문제가 나타나고 있습니다. 복잡하고 서로 얽혀 있는 디지털 생태계의 특성상 책임주체를 규명하고 새로운 구제 제도를 마련하거나 기존 제도를 활용하는 일이 어려울 수 있습니다. 디지털 기술은 종종 복잡한 기술적 문제를 수반하는데, 여기에 기술 혁신의 빠른 속도가 결합하면서 구제수단 모색 과정에 추가적인 난관이 발생합니다.

1. AI으로 인한 피해와 관련된 구제 접근에 관한 주요 고려사항

37. AI 기술이 초래하는 인권 위험은 편향성, 차별, 사생활 침해에 그치지 않고 건강 위험, 복지 문제 및 표현의 자유와 정보접근권 등 여타의 인권 문제까지 포괄합니다. 모든 위험을 AI 배치 전에 완전히 예측할 수 있는 것이 아니며, 의도하지 않았거나 예견할 수 없는 위험도 존재합니다. 이를 해결하기 위해서는 포괄적이고 다각적인 접근이 필요합니다. AI 관련 피해자들이 구제수단에 접근할 때는 복잡성과 불투명성 등 AI 시스템의 기술적 특성으로 인하여 추가적인 어려움을 겪을 수 있습니다. 이는 의사결정 과정과 다양한 이해관계자의 관여 방식을 이해하기 어렵게 만들어 책임 소재에 대한 판단을 극도로 복잡하게 만듭니다.

38. 이 문제와 관련한 주요 고려 사항은 다음과 같습니다.

- (a) AI 의사결정 과정의 투명성과 책무성: 영향 받은 개인들은 현재의 투명성 부족으로 인해 AI와 관련하여 자신이 입은 피해를 인지하지 못할 수 있으며, 이는 해당 결정에 이의를 제기하는 일을 복잡하게 만들고 입증 책임을 가중시킵니다. 영향 받은 개인들이 문제를 신고할 수 있는 명확한 소통 채널이 부재하면 이러한 문제가 더욱 악화됩니다.
- (b) 설명가능성: AI 의사결정 과정을 이해하기 쉬운 방식으로 문서화하고 설명할 수 있는 능력은 책무성을 촉진하고 영향 받은 사람들의 권리를 보호합니다. AI의 '블랙박스 효과'를 해결하려면 위험 평가 보고의 의무화 및 파운데이션 모델에 대한 연구자 접근권 부여 등의 조치가 필요합니다.
- (c) 기술적 전문성: 알고리즘 편향을 입증하려면 기술적 전문성이 필요한데, 많은 피해자와 사법부는 이를 갖추지 못하고 있습니다. 소셜 미디어 플랫폼에서 AI에 기반한 차별을 당한 피해자들은 차별을 받았다는 사실을 기술적 수준에서 입증하지 못하여 구제 실현에 종종 어려움을 겪습니다.
- (d) 사생활 및 개인정보 보호: AI 시스템으로 영향 받는 개인의 사생활과 개인정보를 보호하는 일이 꼭 필요하며, 데이터 오용을 방지하기 위한 조치가 구제에 포함되어야 합니다. 예를 들어 생체 인식 데이터는 유출 후 변경할 수 없기 때문에, 생체 인식 기술과 관련된 침해는 특히 해결이 어렵습니다.
- (e) 이해관계자 집단별 특별한 장벽: 예를 들어 사회복지 서비스에 사용되는 AI도구는 편향을 지속시켜 소외된 공동체가 구제수단에 접근하는 일을 어렵게 만듭니다. 온라인 그루밍이나 유해 콘텐츠 노출 등 AI와 관련된 피해에 노출된 아동은 구제수단에 접근하는 데 상당한 어려움을 겪습니다. 법 제도에서 AI 기반 착취의 복잡성을 처리하거나 자녀 보호 방법을 이해하기 어려운 부모를 지원하는 체계가 부족할 수 있습니다. 여성과 여아들은 AI으로 유발된 성별 기반 폭력을 신고하는 데 특정한 장벽을 만나게 되는데, 특히 신고 체계에 대한 인식 부족이나 법적 체계의 미비 등에 따른 문제는 신고율 저하로 이어집니다.

2. AI가 조장하는 피해에 대응하기 위한 모범 사례

39. AI 기반 결정으로 영향 받은 개인이 직면한 문제를 해결하는 데 있어 대중적 인식이 핵심적인 역할을 합니다. 많은 피해자들은 자신의 권리를 인식하거나 효과적인 구제수단을 찾는 데 어려움을 겪습니다. 대중 인식 캠페인과 교육 프로그램은 이러한 개인들이 사법적 접근에 필요한 지식을 갖추 수 있게끔 하는 중요한 수단으로 기능합니다.

40. 법관 및 기타 법률 전문가의 역량을 강화하여 국가 기반 사법 구제 제도에서 발생하는 디지털 기술 및 인권 관련 사건을 처리할 수 있도록 하는 일이 중요합니다. 사법 제도의 전문성을 강화함으로써 기술이 제기하는 새로운 문제들에 효과적이고 공정하게 대응할 수 있을 것입니다.
41. 국가 차원의 비사법적 고충 처리 제도는 AI 관련 인권침해 사건에서 사법 제도를 보완하고 보조하는 핵심적인 역할을 합니다. 이러한 제도에는 규제 기관, 옴부즈맨, 감사 기관, 공공 민원 처리 기관, 책임 있는 기업 행동을 위한 OECD 국가 연락처 및 국가인권기구가 포함될 수 있습니다. <기업과 인권 이행원칙>에 따라, 기업은 자사 AI 제품과 서비스가 피해를 야기하거나 기여한 것으로 확인된 경우 등에 적극적으로 구제 조치에 참여해야 합니다. 구제를 가능하게 하는 효과적인 수단 중 하나는 기업 고충 처리 메커니즘이 이행원칙의 핵심 기준을 충족하도록 만드는 것입니다. 필요한 인권 보호 장치가 마련되어 있다면, AI 기능을 갖춘 데이팅 앱에서 기술로 유발된 성별 기반 폭력을 기록하는 일로부터 소셜 미디어 플랫폼에서 발생한 전쟁 범죄를 기록하는 일에 이르기까지 AI가 구제수단에 대한 접근을 지원할 수 있습니다.

III. <기업과 인권 이행원칙>의 적용: AI 기술과 관련된 여성 및 여아의 권리와 기업의 책임

42. <기업과 인권 이행원칙>은 국제 인권법에 따른 국가의 의무를 인정하는 데 기반하고 있으며, 여기에는 여성에 대한 모든 형태의 차별철폐에 관한 협약 및 아동권리협약이 포함됩니다. 이행원칙은 기업이 여성 및 여아에 대한 존중을 실천하는 문제에서 해당 체계를 보완하며, 이는 AI 제품과 서비스의 개발, 배치 및 사용에도 적용됩니다.
43. AI 기술은 교육 및 고용 기회에 대한 접근성이나 개인이 목소리를 키울 수 있는 플랫폼에 대한 접근성을 향상시켜 개인의 역량을 강화할 수 있습니다. 그러나 이러한 혜택은 종종 중대한 인권 위험을 동반하기도 하는데, 이는 국가적·민족적·종교적·언어적·인종적으로 소수자, 원주민, 지역사회, 농촌 지역 주민, 경제적 취약계층, 취약하거나 소외된 상황에 처한 개인 등 불리한 처지에 있는 집단에 종종 불균형적인 영향을 미칩니다. 해당 집단 내에서 여성과 여아들은 더 높은 위험에 직면해 있는데 이는 기존 성불평등 패턴의 지속성과 고정관념 강화와 연계된 차별 및 교차성으로 인한 것입니다. 또한 과학·기술·공학·수학(STEM) 분야에서 여성과 여아들의 저조한 참여는 AI 설계 및 개발 과정에서 이들의 관여를 제한합니다.
44. 여성과 여아들은 특히 스토킹, 성희롱, 성착취 등과 같은 성별 기반 폭력에서처럼 AI에 관련된 특정한 피해에 노출되어 있습니다. AI생성 딥페이크는 자주 여성을 표적으로 삼으며 괴롭힘, 협박 및 명예 훼손에 이용될 수 있으며, 그 제작 및 배포의 용이성은 학대의 규모를 확대시킵니다.

신고 제도에 대한 인식 부족과 신고 관련 법적 체계의 미비로 인하여 이러한 문제들에 대한 신고율이 낮을 수 있습니다. 또한 AI 시스템은 채용, 대출, 법 집행 등 다양한 분야에서 차별적 결과를 초래하여 기존의 불평등을 강화하고 여성과 여아를 비롯한 사람들의 기회를 제한할 수 있습니다. AI으로 인한 일자리 감소 및 감시와 데이터 악용 등 사생활과 자율성에 대한 부정적 영향도 여기에 해당되는 문제입니다. 교차적이고 다중적인 형태의 차별은 여성과 여아들이 AI 관련 피해를 경험하는 방식에 대한 차이를 낳으며, 이러한 복합적 위험을 해결하기 위해서는 정책 입안자, 기업 및 시민 사회 공동의 노력이 필요합니다.

45. 기술 기업들은 인권 존중이라는 기업 책임의 일환으로, 성인지적 제품과 서비스 설계에 대해 뛰어난 시장내 의사결정 권한을 고려하여, 조직 전반과 가치사슬에 걸쳐 성인지 정책 및 관행을 채택함으로써 AI 제품과 서비스로부터 여성과 여아의 권리가 침해되는 것을 방지해야 합니다. 예를 들어, 성평등에 초점을 맞춘 직장 내 이니셔티브를 도입하거나, 직원들에게 성인지 관행에 대한 교육을 실시하거나, AI 도구가 유해한 고정관념을 재생산하지 않도록 보장하는 것이 가능합니다. 또한 디지털 제품에 대한 포용적 디자인 관행을 촉진하거나, 기술 리더십에서 여성의 대표성을 지지하거나, 성평등 옹호 단체들과 협력할 수도 있습니다.
46. 주기적인 인권 영향 평가에서 성별 영향을 고려해야 하며, 특히 새로운 활동이나 기술 제품과 서비스의 주요 변경이 이루어지기 전에, 다양한 출신의 여성이 유의미하게 참여할 수 있도록 합니다. AI를 개발하고 배치하는 기업들은 비즈니스 모델과 알고리즘 투명성, 특히 유해 콘텐츠를 증폭시키는 추천 시스템을 성별 관점에서 검토하는 성별 영향 평가를 장려해야 합니다. 여성 단체, 시민사회단체 및 기술 전문가와의 협의는 성별 관련 피해를 완화하고자 하는 노력의 효과성을 모니터링하는 데 도움이 될 수 있습니다. 예를 들어, 일부 플랫폼에서는 기기 내 해시(on-device hashing) 기술을 활용하여 사적인 이미지가 본인 동의 없이 온라인에 유포되는 것을 탐지하고 차단하는 도구를 제공하고 있습니다.
47. 최근 인터넷 검색 엔진 기업 두 곳이 검색 결과에서 동의 없는 사적 딥페이크 콘텐츠를 감소시키기 위한 조치를 취했고, 주요 기술 기업 두 곳이 협력하여 블루투스 기기를 스토킹에 악용하지 못하도록 방지하는 신규 표준을 제안했다는 사실이 고무적입니다.
48. 일부 국가들은 디지털 영역에서 여성과 여아들의 권리를 보호하기 위한 입법적 조치를 취하기 시작했으며 AI 규제도 여기에 포함됩니다. 이러한 접근 방식은 알고리즘 영향 평가 의무화, 고위험 AI 시스템의 편향성과 차별성에 대한 주기적 감사 요구, 영향 받은 사람들에 대한 구제 등 차별적 결과에 대한 책무성 제도 수립을 포함합니다. 또한 캐나다를 비롯한 여러 국가들이 여성 권리 존중 관행에 관한 지침을 발표했습니다.

국가들은 또한 기업과 인권 이행원칙의 성별 차원의 문제에 대해서는 "초국적 기업 및 기타 사업체와 인권 문제에 관한 실무 그룹" 보고서를 활용할 수 있습니다.

49. 사법적 구제수단 접근에 대한 어려움에도 불구하고, 동의 없이 개인정보를 공유한 여성 건강 기술 기업을 상대로 한 집단 소송을 비롯하여, 기술 기업과 관련된 사법적 구제 사례가 일부 존재합니다. 성별에 기반한 영향을 더 잘 해결하고 구제하기 위해 기업 기반 고충 처리 제도가 개선되어야 할 수도 있습니다. 예를 들어, 2024년 메타 감독 위원회가 내린 100건의 구속력 있는 결정 중 성별 관련 문제에 초점을 맞춘 사례는 극소수에 불과했습니다.

IV. 기술 부문에서 기업의 인권 존중을 장려하는 투자자의 역할

50. 기업 부문에 <기업과 인권 이행원칙>의 실제 적용에 대한 이해도를 함양하는 일 외에도, 해당 부문이 책임을 다하도록 보장하려면 더 많은 관심이 필요합니다. 투자 부문은 이 지점에서 핵심적인 역할을 수행하면서 이행원칙에 따른 자신의 책임 또한 다할 수 있습니다. 투자자들은 투자시 인권 문제를 우선시함으로써 기업들이 운영 전반과 가치사슬에 걸쳐 인권 존중을 반영하도록 장려할 수 있습니다. 현재는 투자자들, 특히 AI 투자와 관련된 책임에 대한 인식이 낮은 수준에 머물러 있습니다.
51. 투자자들은 기술 기업의 지배구조와 의사결정에 독특한 영향력을 행사합니다. 공개 주식 투자자와 사모 자본 투자자는 각기 다른 방식으로 영향력 행사 수단을 활용하여 기술 기업들이 자사 비즈니스 모델을 보다 인권을 존중하는 방향으로 전환하도록 독려할 수 있습니다.
52. 다양한 자산군에서 점점 더 많은 투자자들이 포트폴리오 기업들에 대하여 비즈니스 모델을 변경하여 심각한 인권침해 영향을 방지하거나 완화하도록 보다 적극적으로 촉구하려는 의지를 보이고 있습니다. 또한 일부 투자자들은 인권 기반 또는 지속가능성 기반 배제 기준을 유지하여 인권 위험이 높은 투자를 처음부터 하지 않으려 합니다.
53. 더 많은 투자자가 이러한 전략을 기업의 AI 개발 및 배치와 관련된 의사결정에 반영하는 일이 중요합니다. 투자 전 실사에 인권 관점을 적용하고, 포트폴리오 기업의 자체 인권실사 절차의 적절성을 평가하며, 필요한 경우 해당 절차의 개선을 촉구하는 것이 핵심적인 전략입니다. 투자자들이 이러한 전략을 실행할 때, 그들은 AI 제품과 서비스의 개발, 배치 및 사용과 관련된 인권 위험의 수준과 범위를 줄이는 데 기여할 수 있습니다. 투자자들 사이에서 이러한 관행의 활용도를 높이기 위한 진전이 이루어졌지만, 대부분의 공개 주식을 소유한 대형 기관 투자자와 영향력이 큰 AI 기술 몇몇 책임지는 첨단 스타트업에 자금을 조달하는 벤처 캐피털 투자자 등에 이러한 관행을 주류화하기 위해서는 추가적인 노력이 필요합니다.

V. 결론 및 권고 사항

54. 사회가 AI의 혜택을 최대한 얻고 활용하는 동시에 AI가 초래하는 상당한 위험을 예방하고 완화할 수 있는 최선의 방법에 대한 문제를 해결하는 데 있어, <기업과 인권 이행원칙>은 국가 규제 당국과 기업 모두에게 중요한 기초를 제공합니다. 이행원칙은 AI 가치사슬 전반에 걸쳐 인권침해를 방지하고 대응하는 방법에 대한 지침을 담고 있으며, 여기에는 국가가 인권 보호 의무를 이행하고 기업이 인권 존중 책임을 이행하면서, 구제수단 접근성을 제공하는 체제가 포함됩니다. 국가와 기업이 각자의 의무와 책임을 이행하는 데 일부 진전이 있었지만, AI 등 기술 기업의 활동에 이행원칙의 실제 적용을 강화하기 위해서는 훨씬 더 많은 노력이 필요합니다.
55. 국가는 국제 인권법 및 <기업과 인권 이행원칙>을 포함한 국제 기준에 따라 다음을 이행해야 합니다.
- (a) AI가 관여된 활동을 비롯하여 기업 활동과 연관된 인권침해로부터 개인을 보호해야 하며, 여기에는 다음과 같은 규제 조치와 정책적 인센티브를 적절히 조합하여야 합니다.
 - (i) 기업에 대한 의무적 조치의 일환으로, AI의 전체 수명주기에 걸쳐 AI와 관련된 인권 위험을 식별, 예방 및 완화하기 위한 인권실사를 실시하도록 요구함으로써 AI 제품과 서비스에 인권 존중이 반영되도록 보장하는 규제 체제를 시행합니다. 여기에는 AI 시스템의 기반이 되는 데이터 모델이 모든 사람을 대표하도록 보장하고 분쟁으로 영향 받는 지역 등 고위험 사용 사례 및 상황에 대해서는 강화된 인권실사를 요구하는 일이 포함됩니다.
 - (ii) AI모델을 개발 및 배치하는 기업들에 대하여 데이터 입력과 결과 출력을 검증하여 성별, 인종 및 문화적 다양성 측면에서의 대표성을 보장하고, 차별적 AI 출력 및 그 결과로 인한 부정적 영향으로부터 보호할 것을 요구합니다. 이러한 부정적 영향은 종종 여성과 여아 및 기타 취약 계층에게 불균형적인 영향을 미치며 교차적 차별을 악화시킵니다.
 - (iii) 기업에 대하여 AI와 연관된 인권 영향을 완화하고, 충분한 안전장치가 마련되지 않은 AI의 사용을 금지하는 등 제한 조치를 활성화할 것을 요구합니다.
 - (iv) 기업에 대한 인센티브 기반 조치의 일환으로, AI 개발 및 배치의 기반이 되는 디지털 인프라에 대해 결정하는 산업 정책에 인권 고려 사항을 포함합니다.
 - (v) 국가 AI 전략 및 연구개발 전략에 인권 고려사항을 통합하여 인권 체제에 기반한 AI 개발 및 배치를 장려하고, AI 제품 수명주기 전반에 걸쳐 인권실사를 실시할 수 있는 추가적인 인센티브를 제공하며, 여기에는 AI 개발팀 내 인종적·문화적·성적 다양성을 증진하는 방안도 포함합니다.
 - (b) AI 제품과 서비스로 인해 개인의 권리가 침해될 경우, 효과적인 구제수단과 완전한 배상을 받을 수 있도록 보장해야 하며, 여기에는 AI가 지원한 의사결정 과정에 대해 투명성을 요구하고 해당 결정에 대해 의미 있는 인간 감독을 요구하는 등의 조치가 포함됩니다.

- (c) 국가적, 지역적, 국제적 차원에서 AI 거버넌스를 다루는 다양한 정책 분야 및 행위자들 간에 일관성을 촉진해야 합니다.
- (d) 개발금융기구의 AI 제품과 서비스 관련 사업에서 인권 보호 장치를 마련해야 합니다.
- (e) 국가인권기구, 법관 및 법률 전문가들의 역량 강화 조치에 적절한 자원을 제공하여 AI 관련 인권침해 사건을 효과적으로 식별하고 처리할 수 있는 지식과 도구를 갖추어야 합니다.
- (f) 영향 받는 이해관계자의 디지털 리터러시를 지원하여, 이들이 이용 가능한 구제수단에 대한 정보에 포용적이고 명확한 언어와 형식으로 접근할 수 있도록 보장해야 합니다.
- (g) 저소득층 및 소외 계층이 구제 제도에 접근하는 데 불균형적인 영향을 미치는 비용적, 절차적 장벽을 제거해야 합니다. AI와 관련된 잠재적 피해 및 이용 가능한 구제수단에 대하여 대중적 인식 제고 및 홍보 전략에 투자하고, 가장 큰 영향을 받는 지역사회와 이를 공동으로 개발하여야 합니다.

56. AI 제품과 서비스를 개발하고 배치하는 국가는 <기업과 인권 이행원칙>을 포함한 국제 기준에 따라 다음을 이행해야 합니다.

- (a) AI 개발, 배치 및 사용에 인권 기반 접근법을 적용하여 AI 전 수명주기에 걸쳐 인권 원칙을 반영해야 합니다.
- (b) AI 제품과 서비스를 공공 감사와 책무성을 보장하는 투명한 방식으로 개발하고 구현해야 하며, 여기에는 AI 시스템이 설계, 학습 및 배치되는 방식을 명확히 문서화하고 고충 처리 및 피해 구제 체제를 수립하는 일이 포함되어야 합니다.
- (c) AI 제품과 서비스의 개발, 배치 및 사용 전반에 걸쳐 인권실사를 실시하며, 여기에는 잠재적 인권 및 성별 관련 영향을 식별하기 위한 위험 평가와 확인된 위험을 완화하기 위한 조치들이 포함되어야 합니다. 이러한 조치들은 제품과 서비스의 상황별 구현에 대응할 수 있도록 지속적으로 점검되어야 합니다.
- (d) 영향 받는 지역사회 및 시민사회단체를 포함한 이해관계자와의 협의를 실시하여 AI 시스템의 개발 및 배치 과정에 이들을 참여시켜야 합니다.
- (e) AI 기술을 제3자로부터 조달할 경우, AI 제품과 서비스 사용에 대한 인권실사를 자체적으로 실시하고 이 의무를 외부에 위탁하지 않아야 합니다.

57. AI를 개발하고 배치하는 기업은 <기업과 인권 이행원칙>을 포함하여 해당되는 국제 기준에 따라 다음을 이행해야 합니다.

- (a) 자사 AI 제품과 서비스로 인한 부정적 영향으로부터 인권을 보호하기 위한 접근법을 담은 정책을 수립하고, 관련 정책, 위험 평가 및 의사결정 과정을 포함한 AI 수명주기 전반에 인권 고려 사항을 반영해야 합니다.

- (b) 자사 AI 제품과 서비스가 야기하거나 기여할 수 있는 부정적 인권 영향을 식별, 예방, 완화하고 책무를 이행하기 위하여 인권실사를 실시하고, 특히 AI 제품과 서비스의 신규 활동이나 주요 변경이 이루어지기 전에 주기적으로 인권 영향을 평가하여야 하며, 인권 전문가의 의견을 반영하여 그 평가를 감사해야 합니다.
- (c) AI 제품과 서비스에 대한 효과적인 고충 처리 메커니즘을 운영 수준에서 수립하거나 이에 참여하고, 국가 기반 사법 제도와의 협력을 포함하여 영향 받는 개인 및 지역사회에 효과적인 구제수단을 제공해야 합니다.
- (d) 영향 받는 집단 및 시민사회단체와 유의미한 이해관계자 협의를 수행하여 AI 제품과 서비스에 관련된 인권 위험을 식별하고 해결해야 합니다.
- (e) 정성적·정량적 지표를 활용하여 성별 기반 영향에 대한 완화 조치의 효과성을 모니터링하고, 영향 받는 집단과 주기적으로 협의해야 합니다.
- (f) 투명성 보고서 및 성별 영향 평가 등을 통해 인권 영향과 관련된 성과와 위험에 대해 소통해야 합니다.
- (g) AI 제품의 개발팀을 포함한 내부 인권 역량을 확대하고, 인권 위험에 대한 성인지 및 교차적 접근법에 대한 교육을 실시해야 합니다.

58. AI 제품 또는 서비스를 조달하는 기업은 <기업과 인권 이행원칙>에 따라 인권에 미치는 부정적 영향에 대한 상황별 이행도를 평가하고 인권실사 절차를 통해 피해를 완화해야 합니다.

59. 투자자들에 대하여 <기업과 인권 이행원칙>에 따라 다음 사항을 권장하여야 합니다.

- (a) 인권 존중을 위한 책임을 다하기 위해, 인권을 존중하는 AI를 개발 및 배치하는 기업에 대한 투자를 조성합니다.
- (b) 투자에 앞서 투자 조건서에 기업 실사 요구사항을 명시하는 등 기술 기업 창업자에 대한 기대사항을 명확히 제시합니다.
- (c) 인권 위험과 연관될 수 있는 사업 활동을 식별하고, AI를 개발하거나 배치하는 기업 내 투자 관계 담당 및 기타 부서와 협력하여 잠재적 피해를 완화합니다.
- (d) 기업이 자사 AI 제품과 서비스에 대한 인권 영향 평가를 실시하도록 권고합니다.
- (e) AI를 개발하고 배치하는 기업이 자사 제품과 서비스와 연계된 인권침해 문제를 해결하지 못하는 경우, 투자를 검토하고 투자 철회도 고려합니다.

60. 시민사회단체는 다음을 이행해야 합니다.

- (a) 보호에 공백이 있을 가능성이 있고 인권에 부정적 영향을 미치는 AI 제품과 서비스를 지속적으로 식별하고 주목해야 합니다.

(b) <기업과 인권 이행원칙>을 효과적으로 활용하는 등의 방법으로 국가 및 기업 행위자에 관여하여 이러한 공백을 해결해야 합니다.

61. 유엔은 AI를 개발하고 배치하는 기업, AI를 구매하는 기업, 국가, 시민사회 행위자, 국가인권기구 및 기타 주요 행위자와 협력하는 등의 방법으로 <기업과 인권 이행원칙>을 AI 수명주기 전반에 걸쳐 지속적으로 홍보하고 확산시켜야 합니다.

62. 인권이사회에 대하여 <기업과 인권 이행원칙>의 실제 적용과 관련하여 다음과 같은 방식으로 계속해서 관여할 것을 권장합니다.

(a) 다음과 같은 지침의 개발을 지원합니다.

(i) AI 제품 개발 과정에서의 인권실사

(ii) 디지털 기술, 특히 AI도구 및 서비스와 관련된 인권침해에 대응하는 기술 분야의 구제수단 접근

(iii) AI 수명주기별로 성과 여아에게 미치는 특정한 인권 영향

(b) AI 개발과 배치가 야기하는 인권 위험에 대하여 지속적으로 대응하고 기록하고 식별합니다.

(c) AI 관련 활동에 대하여 이행원칙을 실제 적용하는 문제에 관련하여, 국가, 기술 기업 및 기타 관련 이해관계자들에게 추가적인 규제, 운영별, 부문별, 사용 사례별 지침을 제공할 것을 검토합니다.

(끝)