

유네스코 AI 윤리 권고 주요 내용 및 시사점

제10호(2021. 12. 9.)

목 차

- I. 유네스코 AI 윤리 권고의 추진 배경 및 경과 / 1
- II. 유네스코 AI 윤리 권고 특징 및 한계점 / 3
- III. 4대 가치와 10대 원칙 / 6
- IV. 유네스코 AI 윤리 권고의 정책적 시사점 / 16

NIA 한국지능정보사회진흥원

「IT & Future Strategy(IF Strategy)」 보고서는 21세기 한국사회의 주요 패러다임 변화를 분석하고 이를 토대로 미래 지능화 시대의 주요 이슈를 전망, IT를 통한 해결방안을 모색하기 위해 한국지능정보사회진흥원(NIA)에서 기획, 발간하는 보고서입니다.

「IF Strategy」는 미래의 '만약을 대비한 전략'을 담은 보고서를 의미합니다.

NIA의 승인 없이 본 보고서의 무단전재나 복제를 금하며, 인용하실 때는 반드시 NIA, 「IT & Future Strategy 보고서」라고 밝혀주시기 바랍니다. 보고서 내용에 대한 문의나 제안은 아래 연락처로 해 주시기 바랍니다.

▶ 발행인 : 문 용 식

▶ 작 성

- 한국지능정보사회진흥원(NIA) 정책본부 AI·미래전략센터
오연주 책임연구원(053-230-1295, oyeonj@nia.or.kr)
- 김소미 주임연구원(053-230-1283, somikim@nia.or.kr)
- 권태균 주임연구원(053-230-1296, gyun4015@nia.or.kr)

▶ 보고서 온라인 서비스

- www.nia.or.kr

요약

◇ 국제적 AI 윤리 가이드라인의 필요성 대두

- AI 윤리와 관련한 국가 및 지역 수준의 전략과 프레임워크는 개발되어 왔으나, 현재까지 국제적 수준의 기준은 마련되지 않음
- 이에 국가의 발전 정도, 문화적 차이, 공공과 민간의 다중 이해관계자 등 다양한 요인을 고려한 국제적 AI 윤리 가이드라인의 필요 부상

◇ 유네스코 AI 윤리 권고의 특징

- 기존 권고안 및 가이드라인이 구속력 없는 원칙을 제시한 것의 한계를 인식하고, 국가별 상황에 맞게 규제 프레임워크의 도입과 개선을 요구
- AI 시스템 전주기 및 관련 행위자 모두를 적용 대상으로 정의하고 있어, 기존 권고안과 비교하여 가장 포괄적인 수준의 적용 범위를 가짐
- 유네스코의 특화 분야인 교육, 문화, 커뮤니케이션 영역에서의 AI 시스템 활용에 대한 심도 있는 관심과 이해를 보임
- 인간의 존엄성, 기본적 자유, 젠더 평등, 지속가능한 발전, 환경과 생태계의 변영과 같은 인권 관련 규범이 강조
- 권고의 특성상 구속력이 없으나, 유네스코측은 다자체제에 따른 국가 간 압력, 배움, 벤치마킹 등을 통해 영향력을 행사를 전망

◇ 4대 가치 및 10대 원칙

4대 가치	10대 원칙
<ul style="list-style-type: none"> ▪ 인권, 기본적 자유, 인간 존엄성의 존중·보호·증진 ▪ 환경 및 생태계의 변영 ▪ 다양성 및 포용성 보장 ▪ 평화롭고 정의로우며 상호연결된 사회 구축 	<ul style="list-style-type: none"> ▪ 비례성 및 무해성 ▪ 안전 및 보안 ▪ 공정성 및 비차별성 ▪ 지속가능성 ▪ 프라이버시권 및 데이터 보호 ▪ 인간의 감독 및 결정 ▪ 투명성 및 설명가능성 ▪ 책임 및 의무 ▪ 의식 및 리터러시 ▪ 다자적이고 조정 가능한 거버넌스 및 협력

◇ 유네스코 AI 윤리 권고의 정책적 시사점

- (단기) 유네스코 AI 윤리 권고를 반영한 지능정보 서비스의 사회적 영향평가 내용 및 방법을 개발하고, 그 운영 매커니즘 마련에 대한 논의를 신속하게 시작할 필요
- (장기) 국제적 추세에 맞게 윤리적 관점을 포함하면서도 기술 및 산업 경쟁력을 확보할 수 있는 균형적 정책의 발굴 필요

I

유네스코 AI 윤리 권고의 추진 배경 및 경과

□ 추진 배경

- AI 윤리와 관련한 국가 및 지역 수준의 전략과 프레임워크는 개발되어왔으나, 현재까지 국제적 수준의 기준은 마련되지 않음
- 이에 국가의 발전 정도, 문화적 차이, 공공과 민간의 다중 이해관계자 등 다양한 요인을 고려한 국제적 AI 윤리 가이드라인의 필요 부상
- 유네스코의 강점인 다수의 회원국(2021년 12월 현재 193개), 다학제적 전문성, 인권에 대한 이해를 바탕으로 모든 인류, 지속가능한 발전, 평화에 이바지할 수 있는 AI 개발을 위한 윤리 권고안 마련에 착수

□ 추진 경과

- ('18.08) 유네스코 사무총장은 세계과학기술윤리위원회*에 AI 윤리 관련 유네스코의 역할에 관한 예비 연구 준비를 요청

* COMEST(World Commission on Scientific Knowledge and Technology) : 1998년 설립된 과학 및 기술 관련 윤리 규범을 다루는 자문기관으로 그동안 로봇 윤리, 사물인터넷 윤리 등 새로운 디지털 기술의 윤리적 함의에 관한 연구와 보고서 작성을 수행해 옴

- ('19.03) 유네스코 세계과학기술윤리위원회 작업반은 '권고'*의 형태로 AI 윤리에 관한 국제 기준 마련을 제안하는 예비조사 보고서 제출

* 유네스코에서 결의하는 국제 기준설정 문서(standard-setting text)는 아래와 같이 크게 세 가지로 분류되며, 선언 → 권고 → 협약의 순으로 구속력이 높음

구 분	특 성
선언(Declaration)	<ul style="list-style-type: none"> ▪ 비준의 대상이 아니며 특정 사안에 대한 국제적 원칙과 규범을 제시 ▪ 협약 및 권고와 달리 유네스코 헌장 상에 투표를 통한 채택 과정과 후속 조치 절차가 명시되어 있지 않음
권고(Recommendation)	<ul style="list-style-type: none"> ▪ 비준의 대상은 아니며 특정 사안에 대한 국제적 원칙과 규범을 공식화 하고, 회원국이 입법 또는 기타 조치를 통해 권고를 적용할 것을 요청 ▪ 유네스코 총회는 회원국에 권고의 이행상황을 보고하도록 요청 가능
협약(Convention)	<ul style="list-style-type: none"> ▪ 비준 및 승인의 대상으로, 회원국이 준수키로 동의한 원칙

- ('19.05) 이사회는 위 보고서 결과를 토대로 제40회 총회('19.11)에서 권고안 작성에 대해 논의하고, 제41회 총회('21.11)까지 권고안 제출을 요구할 것을 결정
- ('20.3~5) 유네스코 사무총장은 권고안을 작성할 24명의 국제전문가 그룹을 임명했으며, AI 윤리에 관한 권고안 초안 작업 착수
- ('20.7~8) 권고안 초안에 대한 의견을 수렴하기 위해 11개 지역별 화상회의를 포함한 온라인 협의를 추진
- ('20.9~12) 예비 보고서 및 권고안 초안을 회원국에 전달했으며, 최종 보고에 반영하기 위해 회원국으로부터 서면 의견 수렴
- ('21.4~6) 권고안에 대한 회원국의 의견을 수렴하기 위해 “유네스코 AI 윤리에 관한 권고안 정부 간 특별위원회” 개최
- ('21.11) 11월 23일에 열린 제41회 총회에서 193개 회원국이 만장일치로 AI 윤리 권고안 채택

II

유네스코 AI 윤리 권고 특징 및 한계점

□ 권고의 특징

- 기존의 권고안 및 가이드라인이 일반적이고 추상적이며 구속력 없는 원칙을 제시한 것의 한계를 인식하고, 권고의 형식을 통해 국가별 상황에 맞게 규제 프레임워크의 도입과 개선을 요구
 - 유네스코 사무총장은 권고의 형식이 ① 회원국에 책임감을 부여하고, ② 관련 법제도와 규제를 도입하는데 있어서 레퍼런스를 제시하며, ③ 이행상황을 모니터링 할 수 있고, ④ 모니터링 결과를 회원국 모두가 공유할 수 있다는 점에서 강점이 있음을 설명
 - 더불어 인권 존중, 생태계 번영, 다양성 보장 등 추상적인 가치들을 구체적인 정책조치를 통해 매우 구체적이고 명확한 언어로 표현함으로써, 기존의 권고안 및 가이드라인과 비교해 정책 수립에 있어 실용성이 높음
- AI 시스템의 수명주기*, 그 주기 중 한 단계라도 관련된 행위자** 모두를 권고의 적용 대상으로 정의하고 있어, 기존의 권고안 및 가이드라인과 비교하여 가장 포괄적인 수준의 적용 범위를 가짐
 - * 연구, 설계, 개발부터 출고 및 사용을 포함하는 주기로서 유지 거버넌스, 운용, 교역, 자금조달, 모니터링 및 평가, 유효성 검사, 사용 종료, 분해, 폐기에 이르는 모든 과정을 포함
 - ** 연구자, 프로그래머, 엔지니어, 데이터 과학자, 최종 사용자, 기업, 대학, 민간 및 공공 단체 등을 비롯한 자연인과 법인 전체를 포함
- 유네스코의 전문 분야인 교육, 문화, 커뮤니케이션 영역에서의 AI 시스템 활용에 대한 심도 있는 관심과 이해를 보이며, 해당 분야에 대해 별도의 정책조치 내용을 제시
- 인간의 존엄성, 기본적 자유, 젠더 평등, 지속가능한 발전, 환경과 생태계의 번영과 같은 인권 관련 규범이 강조되는 한편, 관련 정책 개발 및 이행과정에서 국제인권법의 준수를 반복하여 역설

□ 권고의 한계점에 대한 유네스코 입장¹⁾

- 권고의 특성상 구속력이 없다는 한계를 가짐
 - ⇒ 실질적 구속력은 없으나 다자체제(multilateral system)에 따른 국가 간 압력, 배움, 벤치마킹 등을 통해 영향력을 행사. 유네스코는 근거에 기반해 AI의 혜택과 위험에 대해서 지속적으로 기록하고 알람으로써 회원국의 실천을 장려
- 기타 국제기구 및 협의체(유럽위원회, OECD, GPAI*)의 작업과 중복
 - ⇒ 유네스코는 유럽위원회, GPAI와 같은 기구 및 협의체와 협력하고 있음. 유네스코의 권고가 중요한 이유는 193개라는 많은 회원국이 승인한 만큼 영향력을 행사한다는 것. 아울러 위의 조직들이 AI 개발 역량을 가진 국가들만의 모임이라면, 유네스코는 현재는 그 역량이 없으나 향후 투자 가능성이 있는 국가까지 포섭(UNESCO, 2021)
- * Global Partnership on Artificial Intelligence (인공지능에 대한 글로벌 파트너십)
- 인권 존중 수준이 낮은 회원국에서의 이행은 어려울 수 있음
 - ⇒ 중국, 이란, 러시아 등 인권 존중 수준이 낮다고 알려진 국가에서 서명했다는 것은, 권고가 제재 없는 장려책(incentive mechanism)으로 기능한다는 것을 의미. 또한 AI가 매우 중요한 현시점에서 잘못된 방향으로 나아가는 국가로 인식되지 않기를 원한다는 점을 보여줌.
- AI 리더 국가인 미국은 유네스코를 탈퇴*하여 서명하지 않음
 - * 미국과 이스라엘은 유네스코가 반이스라엘 편견을 이유로 2017년 탈퇴를 선언했으며, 2018년 12월 31일부터 효력 발생
 - ⇒ 다자체제에 따른 국가 간 압력은 미국에 영향을 미치는 강력한 도구이며, 윤리적 AI에 관한 국제 규범 만들기에 동참하지 않더라도 자체적으로 유사한 노력을 기울일 것임. 실제로 미국에서 EU보다 앞서 AI 규제 관련 법안이 만들어짐.

1) 권고 채택 후에 열린 유네스코 측의 공식 기자회견을 토대로 작성. UNESCO. (2021, 11, 25). Press conference: UNESCO presents a global agreement on the ethics of artificial intelligence. YouTube. <https://www.youtube.com/watch?v=9tp1FfUzBg/>

- 인터넷에 대한 접근성 조차 가지고 있지 않은 국가들도 존재
 - ⇒ 유네스코는 AI를 더욱 포용적으로 만들기 위해서 노력하고 있으며, 이를 위해 AI 역량을 갖추고 있지 않은 국가에서도 AI 발전에 기여할 수 있도록 투자 예정. 유네스코는 AI 기술이 소수의 국가 또는 기업에 의해서만 개발된다는 점에 유의하며, 인터넷 보편성(Internet universality)을 주도하는 조직으로서 더 많은 국가가 AI 개발에 참여할 수 있는 방안을 권고에 담았음.
- 테크기업들은 이미 자율 규제를 하고 있으나 실질적 효력 없음
 - ⇒ 이번 권고가 전달하는 핵심적인 메시지 중의 하나는 더이상 기업의 자율 규제, 또는 민간 차원에서의 자발적인 합의나 규칙으로 AI 윤리를 다루지 않을 것이라는 점. 이는 안티 AI 입장이 아닌 AI로 인한 문제를 해결하기 위해 더 많은 투자가 있어야 한다는 메시지이며, 정부는 이를 주도할 역량을 가지고 있음.

III

4대 가치와 10대 원칙

□ 4대 가치

인간에 대한 존중과 더불어 환경 및 생태계의 보호와 번영, 국가 간 자원 및 발전 수준 차이에 따른 불균형 문제의 극복 등 유엔과 유네스코에서 강조하는 가치 강조

① 인권, 기본적 자유, 인간 존엄성의 존중·보호·증진

- 국제인권법을 포함한 국제법에 따라 확립된 인간의 존엄성 및 권리의 존중, 보호, 증진이 AI 시스템 수명주기 전반에서 실현되어야 함
- AI 시스템 사용에 있어서 삶의 질이 향상되어야 하며 인간의 대상화, 존엄성 훼손, 인권 및 기본적 자유의 침해 또는 탄압이 없어야 함

② 환경 및 생태계의 번영

- 환경 및 생태계가 AI 시스템 수명주기 전반에서 인식, 보호, 증진되고, AI의 발전으로 인한 혜택을 누릴 수 있어야 함
- AI 시스템 수명주기에서 환경 및 생태계의 보호·복원, 지속가능한 개발을 위해 제정된 국제법과 각국의 규범·기준·관행을 준수해야 함

③ 다양성 및 포용성 보장

- 개인의 특성과 관계없이 AI 시스템 수명주기 전반에서 국제법에 따라 다양성 및 포용성의 존중 보호, 증진이 보장되어야 함
- 특히 필수적인 기술 인프라, 교육, 역량, 법적 기반 등이 부족한 국가(저중소득국, 최빈개도국, 내륙개도국, 군서도서개도국)와의 노력을 통해 부족한 부분을 극복하는 자세가 필요

④ 평화롭고 정의로우며 상호연결된 사회 구축

- AI 행위자들은 인권 및 기본적 자유의 가치에 근거해 모두의 이익을 위하여 상호 연결된 미래를 위한 평화롭고 정의로운 사회 구축에 적극적으로 참여하고 이를 실현해야 함

- AI 시스템 수명주기 전반에서 평화, 포용성, 공정성을 추구하는 한편, 공동체와 상호 공존의 가치를 위협해서는 안 됨

□ 10대 원칙

제시된 원칙은 기존에 발표된 권고안 및 가이드라인과 큰 차별성이 없으나, 국제법(특히 국제인권법)에 기반하여 실질적인 법제도·규정의 준수를 요구하는 특징을 가짐

① 비례성 및 무해성

- AI 시스템 수명주기 전반에서 합법적 목적을 달성하기 위해 필요 이상의 수단을 써서는 안 되며, 각 상황에 비례한 결정을 내려야 함
- AI 시스템의 선택과 방식은 다음의 방법으로 정당성을 갖추어야 함
(1) 합법적 목적 달성에 있어서 적합성비례성 보장 (2) 권고에 명시된 기본 가치를 침해하지 않음
(3) 각 상황에 대한 적합성과 엄격한 과학적 근거를 갖추며, 중대한 의사결정은 인간의 판단 적용

② 안전 및 보안

- AI 시스템의 수명주기 전반에서 원치 않는 안전과 관련한 피해와 공격에 대한 취약성과 관련된 요소를 예방·방지, 해결, 제거해야 함

③ 공정성 및 비차별성

- AI 행위자들은 사회 정의 증진, 국제법에 따른 공정성과 반차별성을 수호하는 포용적 접근을 통해 모두에게 AI의 혜택을 보장해야 함
- AI 시스템 수명주기 전반에서 차별적이고 편향된 관행 및 결과를 최소화하며, 피해에 대한 구제 방안을 갖추어야 함

④ 지속가능성

- AI 기술이 지속가능성 목표 실현에 UN의 지속가능한 개발 목표에 명시된 지속가능성이 시사하는 바에 대한 충분한 이해 필요

⑤ 프라이버시권 및 데이터 보호

- AI 시스템의 수명주기 전반에서 인간의 존엄성, 자주성, 주체성을 보호하는 데 필수적 요소인 프라이버시가 존중, 보호, 증진되어야 함

- AI 시스템 전반에서 데이터 보호 프레임워크 및 거버넌스 매커니즘은 다자적 접근방식을 통해 국제적 원칙 및 기준에 따라 보장해야 함

⑥ 인간의 감독 및 결정

- AI 시스템 수명주기의 모든 단계 및 AI 시스템 관련 구체 조치에서 자연인 혹은 법인에 윤리적·법적 책임을 물을 수 있어야 함
- 인간이 효율과 관련된 이유로 AI 시스템에 의지할 수 있지만 AI 시스템은 최종적으로 인간의 몫인 책임 및 의무를 대신할 수 없음

⑦ 투명성 및 설명가능성

- AI 시스템의 투명성 및 설명가능성은 인권, 기본적 자유, 윤리 원칙을 존중, 보호, 증진하고 법적 책임 제도를 보장하기 위한 필수 전제
- ※ (투명성) AI 알고리즘을 기반으로 한 결정을 통보받거나 그 결정이 내려질 때 이에 대해 충분한 정보가 제공되어야 하며, 관련 AI 행위자에 정보를 요청할 수 있는 기회가 주어져야 함
- ※ (설명가능성) AI 시스템이 도출한 결과에 대한 이해와 추적이 가능해야 하며, 알고리즘의 입력값, 결과값과 기능, 그리고 이것이 AI 시스템에 미치는 영향을 이해할 수 있어야 함

⑧ 책임 및 의무

- AI 행위자 및 회원국은 국내외 법에 따른 윤리적·법적 책임에 따라 인권, 기본적 자유, 환경 및 생태계를 보호하고 증진해야 함
- AI 시스템과 관련된 모든 의사결정 및 행위의 윤리적·법적 책임은 AI 행위자에게 귀속되며, 의무 이행 메커니즘이 개발되어야 함

⑨ 의식 및 리터러시

- 열린 교육 및 접근 가능한 교육, 관련 역량 및 리터러시 증진, 시민 참여를 통해 AI 기술과 데이터의 가치에 관한 의식과 이해를 증진
- AI 시스템의 영향에 관한 학습은 AI 시스템이 인권, 인권 행사, 환경 및 생태계에 미치는 영향에 기반을 두어야 함

⑩ 다자적이고 조정 가능한 거버넌스 및 협력

- AI 거버넌스에 대한 포용적 접근을 통해 모두가 혜택을 공유하고 지속가능한 발전을 이루기 위해 AI 시스템 수명주기 전반에서 다양한 이해관계자의 참여가 필요

참고

AI 윤리 규범 비교 (유네스코, 우리나라, OECD)

AI 윤리 8대 테마*	UNESCO (21년) 「AI 윤리에 관한 권고 초안」 4대 가치 및 10대 원칙	우리나라 (20년) 「인공지능(AI) 윤리기준」 3대 기본원칙 10대 핵심요건	OECD (19년) 「AI 권고안」 5대 원칙
Privacy (프라이버시)	· (원칙5) 프라이버시권 및 데이터 보호	· (요건2) 프라이버시 보호 · (요건7) 데이터 관리	-
Accountability (책임성)	· (가치2) 환경 및 생태계의 반영 · (원칙8) 책임 및 의무	· (요건4) 침해금지 · (요건8) 책임성	· (원칙5) 책임성
Safety and Security (안전 및 보안)	· (원칙2) 안전 및 보안	· (요건9) 안전성	· (원칙4) 견고성, 보안 및 안전
Transparency and Explainability (투명성 및 설명가능성)	· (원칙7) 투명성 및 설명가능성	· (요건10) 투명성	· (원칙3) 투명성 및 설명가능성
Fairness and Non-discrimination (공정성 및 차별금지)	· (가치3) 다양성 및 포용성 보장 · (원칙3) 공정성 및 비차별성	· (요건3) 다양성 존중	· (원칙2) 인간 중심 가치, 공정성
Human Control of Technology (기술에 대한 인간의 통제)	· (원칙6) 인간의 감독 및 결정	· (요건4) 침해금지	-
Professional Responsibility (전문직 책임)	· (원칙10) 다자적이고 조정 가능한 거버넌스 및 협력	-	-
Promotion of Human Values (인간 가치 증진)	· (가치1) 인권, 기본적 자유, 인간 존엄성의 존중·보호·증진	· (원칙1) 인간 존엄성 원칙 · (원칙2) 사회의 공공선 원칙 · (요건1) 인권보장 · (요건5) 공공성	· (원칙1) 포용적 성장, 지속가능한 발전, 안녕(well-being) · (원칙2) 인간 중심 가치, 공정성
기타	· (가치2) 환경 및 생태계의 반영 · (가치4) 평화롭고 정의로우며 상호연결 된 사회 구축 · (원칙1) 비례성 및 무해성 · (원칙4) 지속가능성 · (원칙9) 의식 및 리터러시	· (원칙1) 기술의 합목적성 원칙 · (요건6) 연대성	· (원칙1) 포용적 성장, 지속가능한 발전, 안녕(well-being)

* 하버드대학교 버크만 클라인 센터는 '16년~'19년 발표된 AI 윤리 원칙에 관한 대표문서 36개를 종합 분석하여 8개의 테마를 도출. 위의 표는 유네스코, 우리나라, OECD의 AI 윤리 규범을 8대 테마와 비교한 것으로, 각 규범이 2개 이상의 테마를 포함하는 경우 그 의도와 핵심적 내용을 고려하여 일부 중복 배치.

□ 정책조치 분야 (우리나라 관련 정책현황과 비교)²⁾

① 윤리영향평가

유네스코	한국
· AI 시스템의 해택위험을 식별 평가하고, 모니터링 조치를 개발하기 위한 영향평가 프레임워크 도입	· (과기부, '21~) 지능정보 서비스 사회적 영향평가
· AI 시스템의 영향을 식별·예방·완화하고 그 의무를 이행하기 위한 실사 및 감독 메커니즘 개발	· (과기부, '21~) AI 제품·서비스 신뢰 확보 체계 마련
· AI 시스템이 반권, 디지털 격차 등에 미치는 사회경제적 영향평가 및 관련 정보에 대한 접근성 확보	· (과기부, '21~) 지능정보 서비스 사회적 영향평가
· 의사결정에 사용된 알고리즘, 데이터, 관계된 AI 행위자 등 AI 시스템 수명주기 모든 단계 모니터링	· (과기부, '21~) AI 제품·서비스 신뢰 확보 체계 마련
· AI 기반 결정이 인간의 의사결정 주권에 미치는 사회적·심리적 영향 조사	· (과기부, '21~) 지능정보 서비스 사회적 영향평가
· 공공당국과 같은 분야에서 사용하는 AI 시스템 윤리 영향평가 절차를 제시하는 규제 프레임워크를 채택	· (행안부, '21) 공공분야 AI 도입을 위한 실무자 안내서 마련 · (법제처, '21) AI 활용한 자동화 행정 근거 마련

② 윤리적 거버넌스 및 책무

유네스코	한국
· 포용적이고 투명하며 다원적, 다각적, 다자적인 AI 거버넌스 메커니즘 보장	
· AI 시스템에 의한 피해가 디지털 및 물리적 환경에서 조사·방지되도록 집행 메커니즘 및 시정조치 마련	· (개보위, '21~) 설명요구권 및 이의제기권을 포함하도록 개인정보보호법 개정 · (금융위, '21) 금융분야 AI 가이드라인 마련 · (과기부, '23~) AI가 개입된 범죄에 대한 제재 다양화
· 국가지역 수준에서 인권, 환경, 생태계 등을 고려한 AI 시스템 인증 메커니즘 및 상호 인증체계 채택	
· AI 도입의 적절성 여부 평가를 포함한 AI 시스템에 대한 투명한 자체평가 실시	· (과기부, '21~) AI 제품·서비스 신뢰 확보 체계 마련
· AI 시스템의 윤리적 지침을 준수하도록 독립적인 AI 윤리 책임기관을 개설하거나 기타 메커니즘 도입	
· 국가적 차원에서 윤리적·포용적 AI 시스템으로 이뤄진 디지털 생태계의 개발 및 접근을 증진	
· AI 거버넌스의 국제 논의에서 국제기구, 다국적기업, 학술기관, 시민사회 등과 협력, 모든 회원국 참여 보장	· (과기부, '22~) 시민 의견 수렴 플랫폼 구축운영
· AI 시스템 규범에 대한 개정 혹은 제정은 AI 시스템 수명주기 전반에서 인권 및 기본적 자유를 증진	

2) 우리나라 정책의 경우 각 부처에서 발표한 정책자료를 기반으로 작성하였으며, AI 윤리와 관련된 계획이더라도 유네스코 권고의 정책조치와 일치 정도가 낮은 경우 포함되어 있지 않을 수 있음.

· 독립적인 데이터 보호 책임기관을 통해 민감한 분야에서 사용되는 AI 시스템의 사회경제적 영향을 모니터링	
· 법치의 원칙과 국제법 기준에 따라 AI 시스템 관련 의사결정을 내리도록 사법부의 역량 강화	
· 안전 및 투명성 수준을 측정하는 국제 기준을 개발하여 시스템을 객관적으로 평가하고 규범 준수 여부 점검	
· AI 시스템 수명주기 전반에서 AI 행위자의 행동이 국제 인권법, 기준, 원칙과 일치하도록 보장하는 정책 시행	
· AI 행위자가 AI 시스템 및 데이터 결과에서 발견되는 모든 고정관념을 공개·삭제하도록 하는 메커니즘 도입	
· AI 개발팀과 데이터셋이 인구 구성을 반영할 수 있도록 다양성과 포용성을 증진하는 정책 시행	
· AI 시스템 수명주기 전반에서 인간 감독의 원칙을 따른 규제 프레임워크를 개발, 검토, 조정	
· 법적 프레임워크 개정 시 모든 AI 행위자의 참여 증진, 규제 및 정책 공신력 확보 전 안전한 테스트 환경 마련	· (과기부, '20~) ICT 규제 샌드박스 도입
· AI 시스템의 투명성, 설명가능성, 신뢰성을 보장하기 위한 분명한 요건 설정	· (과기부, '21~) AI 윤리기준 실천을 위한 자율점검표 마련

③ 데이터 정책

유네스코	한국
· 데이터 수집 및 선택 과정의 적절성, 적절한 데이터 보안 및 보호 조치 보장	· (과기부, '21~) 원천 데이터의 다양성, 충분성, 사실성, 공정성을 확보하고, 획득과정에서 저작권, 개인정보보호 등 법제도 정비
· AI 시스템 트레이닝 데이터의 지속적인 품질 평가를 보장하는 데이터 거버넌스 전략 개발	· (과기부, '21~) 학습용 데이터 구축 전과정 신뢰성 확보
· 국제법에 따라 프라이버시권 보호와 감시 문제 등을 해결하기 위한 적절한 안전장치 마련	· (개보위·방통위·금융위, '20) 「개인정보 보호법」 등 데이터 3법 개정
· 개인이 개인정보에 대한 권리를 행사하고, 그 권리를 적절한 프레임워크를 통해 보호받을 수 있도록 보장	· (개보위·방통위·금융위, '20) 「개인정보 보호법」 등 데이터 3법 개정
· 범죄 관련 정보, 생체 데이터, 의료 데이터 등 민감 정보의 보안을 위한 정책 프레임워크 확립	· (개보위·방통위·금융위, '20) 「개인정보 보호법」 등 데이터 3법 개정 · (복지부, '21) 「보건의료 데이터 인공지능 혁신전략」 수립
· 정보 접근 및 열린 정부를 비롯한 분야의 정책·규제 프레임워크 검토하고 개방형 데이터 증진	· (행안부, '18~) 공공기관이 보유한 공개 가능한 공공 데이터 전면 개방
· AI 시스템의 트레이닝, 개발, 사용을 위한 양질의 데이터셋의 사용을 장려하고 데이터 수집 및 사용 감독	
· 데이터에 대해 디지털 공유재 방식의 접근을 채택, 데이터의 상호운용성 강화, 민간기업의 공익을 위한 데이터 공유 장려 등 협업 증진	· (행안부, '18~) 분산된 공공분야 데이터의 통합 및 분석 활용을 위한 범정부 데이터 플랫폼 구축

④ 발전 및 국제협력

유네스코	한국
· 국제포럼, 정부 간 포럼, 다자적 포럼 등에 AI 윤리 문제에 대한 논의를 포함함으로써 AI 윤리를 우선시	· (정부, '20) OECD 「AI 권고안」 작성에 주도적 참여 · (외교부, '19~) 유엔 인권이사회 '신기술과 인권' 인권 주도
· 교육, 정보통신, 보건, 농업/식량 공급, 환경 등의 개발 분야에서 AI 사용 시 본 권고의 가치 및 원칙 준수	
· 저중소득국 등의 발전과제를 해결하기 위해 국제기구를 통해 AI 개발을 위한 국제협력 플랫폼을 제공	
· 저중소득 등의 참여 확대와 리더십 증진을 위해 AI 연구혁신에 대한 국제협력 증진	
· 국제기구, 연구기관, 다국적 기업의 참여에 기반한 AI 윤리 연구 증진	

⑤ 환경 및 생태계

유네스코	한국
· AI 시스템 수명주기 전반에서 직간접적 환경영향을 평가, AI 시스템 및 데이터 인프라의 환경영향 감소	· (산업부·중기부, '21) 스마트공장 통합플랫폼 구축과정에서 '중소기업형 ESG 가이드라인' 마련
· 윤리적 AI 기반의 재난 극복, 환경 보호, 지구 보존을 보장하기 위한 인센티브 제공	· (환경부, '20~) 물관리, 오존 예보, 생태계 조사·연구 등을 위해 AI 도입 · (환경부, '20~) 디지털 기반 환경영향평가 기술개발사업
· AI 방법 채택 시 행위자들이 데이터·에너지·자원 효율적인 방법을 채택하도록 장려	

⑥ 젠더

유네스코	한국
· 디지털 기술과 AI의 잠재력이 젠더 평등을 극대화하도록 장려하고 젠더 관점을 윤리영향평가에 포함	
· 국가 디지털정책이 젠더실행계획을 포함하고, 여아와 여성이 디지털 경제에서 소외되지 않도록 제도 지원	
· 젠더 고정관념과 차별적 편견이 AI 시스템에 반영되지 않도록 선제적 조치	· (여가부, '21) AI 활성화로 인한 사회적 편견과 차별 방지를 위한 윤리기준 마련 및 정책 개선 권고
· AI 시스템 수명주기 전반에서 여성의 참여, 참여, 고용을 장려하고 젠더 대표성 확보를 위한 정책 개발	· (여가부, '21) 특정 성별영향평가에 따른 정책 개선 권고안에서 AI 전문인력의 성별다양성 권고
· 여아와 여성이 해당 분야에 진출할 수 있도록 인센티브 제공, 연구 공동체에서의 차별 근절 메커니즘 시행	
· 학술기관 및 민간기관의 젠더 다양성을 증진하기 위한 모범사례를 공유, 학계와 산업에서의 젠더 다양성 증진	

⑦ 문화

유네스코	한국
· 회원국은 유무형 문화유산의 보존, 강화, 증진, 관리, 접근에 적절한 경우 AI 시스템을 적용	
· 자동번역 및 음성지원 등 자연언어처리 시스템과 같은 AI 시스템의 문화영향을 평가	
· 예술 및 창작 분야에서 AI 기술 사용의 지속가능성 평가, 예술가·창작가 대상 AI 교육 및 디지털 트레이닝 장려	
· 문화 시장의 리스크를 피하기 위해 지역 문화기관과 중소기업의 AI 도구에 대한 인식 개선과 평가 증진	
· 문화적 다양성을 확대하도록 알고리즘의 추천이 지역 콘텐츠를 가시화하고, 발견성을 높일 수 있도록 지원	· (과기부, '20~) 한국어 방언 발화 AI 학습용 데이터 구축
· AI 기술로 생산된 작품의 지적재산권(IP) 적용 및 보호 여부 등에 관한 연구 장려	· (문화부·특허청, '21~) AI 창작물의 권리에 관한 법체계 개편안 마련 · (문화부, '21~) '신기술 환경 지식재산권 협의체' 운영
· 국립박물관, 미술관, 도서관 자료의 아카이빙과 접근성 강화에 AI 시스템을 사용하도록 장려	

⑧ 교육 및 연구

유네스코	한국
· 국민의 자율성 강화와 정보 격차 해소를 위해 모든 국민들에게 AI 리터러시 교육 제공	· (교육부, '25) 초·중·고교 AI 교육 전면 도입
· AI 교육에 앞서 기초 리터러시, 산술 능력, 미디어 및 정보 리터러시, AI 윤리 등 '필수 역량'의 습득 장려	
· AI 기술로 인한 기회와 도전, AI 시스템이 인권에 미치는 영향 등 AI 개발에 대한 일반 인식 프로그램 증진	· (과기부, '21~) 전국민 AI 윤리 총론 및 맞춤형 교육과정 개발
· 학습, 교사 트레이닝, e-러닝 등에 있어 책임감 있고 윤리적인 AI 기술 사용에 대한 연구 이니셔티브 장려	
· 여성, 장애인 등 AI 교육에서 소외되는 이의 참여와 리더십을 장려하고 그 과정을 모니터링, 모범사례 공유	
· 모든 수준의 사람들에게 적합한 AI 윤리 커리큘럼 개발, AI 기술 교육과 인문·윤리·사회적 측면 간 상호협력 도모	· (과기부, '21~) 전국민 AI 윤리 총론 및 맞춤형 교육과정 개발
· AI 윤리 가치 및 원칙의 준수가 AI 기술개발에 기여할 수 있음을 인지하고, 관련 연구 지원	· (과기부, '22~) 사람중심 AI 핵심원천기술개발 사업
· 저·중·소득국 등의 과학 공동체가 연구를 위해 필요한 민간기업의 데이터에 접근할 수 있도록 지원	
· 과학·기술·공학·수학(STEM) 이외의 영역이 AI 학제 간 연구에 포함되도록 지원	

⑨ 정보통신

유네스코	한국
· AI 시스템을 활용하여 정보지식에 대한 접근성 개선	
· AI 행위자가 표현의 자유, 자동화된 콘텐츠의 생성·조정·큐레이션과 관련된 정보 접근을 증진하도록 지원	· (방통위, '21) 「인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙」 발표
· AI 시스템의 사용 방법과 의미를 이해하는 데 필요한 역량을 강화하기 위해 리터러시 기술에 투자	· (방통위, '22) 스타트업을 대상으로 서비스 설계 시 AI 역기능 등 위험관리 컨설팅 지원 사업 추진
· 언론이 AI 시스템의 유익성과 유해성을 효과적으로 보도할 권한과 자원을 확보	

⑩ 경제 및 근로

유네스코	한국
· AI 시스템이 근로 시장에 미치는 영향 및 교육과 관련된 함의점을 파악하고 조치	
· 산업의 미래와 필요에 맞춰 트레이닝 프로그램과 전략을 조율하기 위해 다자간 협력 증진	· (과기부, '21) 「민관 협력 기반의 소프트웨어 인재양성 대책」 마련 · (노동부, '21) 플랫폼 종사자 특화 직업훈련 시범사업
· 실직 위기의 근로자들이 공정한 전환 과정을 거칠 수 있도록 업스킬링, 리스킬링, 안전망 프로그램 등 시행	· (고용부, '21) 「산업구조 변화에 대응한 공정한 노동전환 지원방안」 마련 · (국세청·고용부· 기재부, '21) 일자리 환경변화에 대응하기 위한 고용보험 확대 지원
· AI 시스템의 경제적·사회적·지리적 영향에 대한 연구, 인간-로봇 및 인간-인간 상호작용에 대한 연구 장려	
· AI 기술 관련 시장 독점, 지배적 시장 구조의 남용을 예방하기 위해 시장의 경쟁력과 소비자 보호를 보장하는 적절한 조치 실행	· (특허청·공정위, '22~) 데이터의 독점 및 부당한 이용행위에 대한 규율 방안 마련

⑪ 의료 및 사회적 안녕

유네스코	한국
· 국제법 및 인권 의무에 따라 질병 발생 완화, 인간의 건강과 생명권 보호에 효과적인 AI 시스템 채택	· (복지부, '21) 「보건의료 데이터·인공지능 혁신전략」 수립 · (식약처, '22) AI 의료기기 국제 기준 개발
· 의료분야에서의 AI 시스템 개발이 과학적·의학적으로 검증되었음을 확인, 모든 단계에서 환자와 대리인의 참여 권장	
· 보건 분야에 AI 적용 시 1) 편견 최소화, 2) 개발 과정에 관련 주체 참여, 3) 프라이버시 보호, 4) 인간에 의한 최종 결정, 5) 임상 사용 전 AI 시스템 보장 등에 주의	

· AI 시스템이 정신 건강에 잠재적으로 미칠 수 있는 피해의 영향 및 규제에 관한 연구 진행	
· 인간-로봇 간 상호작용과 이것이 인간-인간 간 관계에 미치는 영향에 관한 지침 개발	
· 인간-로봇 상호작용이 AI 시스템에 적용되는 인권 및 기본적 자유 등의 가치 및 원칙을 따르도록 지원	
· 사용자가 AI 시스템과의 상호작용을 쉽게 인지하고, 중단하거나 인간의 개입을 요청할 수 있는 환경 마련	
· 인간의 감정을 모방하는 AI 기술의 의인화에 대한 경각심을 깨우는 정책 시행	
· 아동에 미칠 수 있는 심리적·인지적 영향에 집중하여, 인간과 AI 시스템의 상호작용에 따른 효과 연구 장려	
· 아동의 삶에 영향을 미치는 AI 시스템 개발의 논의와 의사결정 과정에 이들을 참여시키는 메커니즘 시행	

IV

유네스코 AI 윤리 권고의 정책적 시사점

□ 정책적 시사점

① (단기) 지능정보 서비스 사회적 영향평가 도입 및 개선

지능정보 서비스의 사회적 영향평가는 「지능정보화기본법」에 따라 '21년부터 실시하도록 되어 있어, 권고를 반영한 내용 및 방법의 개발과 그 운영 메커니즘 마련에 대한 논의를 신속하게 시작할 필요

○ 영향평가 환류체계 마련

- 권고안은 AI 시스템의 혜택과 위험을 식별하고 평가하는 것에서 나아가, 평가를 통해 도출된 결과를 이행하는 실사 및 감독 메커니즘의 개발을 영향평가의 일환으로 인식
- 「지능정보화 기본법」에서는 지능정보 서비스 사회적 영향평가에 대한 후속 조치를 소극적으로 다루고 있어* 영향평가 → 권고 → 관계 부처·기관 조치 → 모니터링으로 이어지는 환류체계 마련 필요

* 제56조(지능정보서비스 등의 사회적 영향평가)제2항 : 과학기술정보통신부장관은 사회적 영향평가의 결과를 공개하고, 해당 지능정보서비스 등의 안전성·신뢰성 향상 등 필요한 조치를 국가 기관 등 및 사업자 등에 권고할 수 있다.

○ 영향평가 시 국제인권규범 반영

- 권고안은 윤리영향평가를 비롯하여 문서 전반에서 인간의 존엄성, 기본적 자유, 젠더 평등, 환경과 생태계 변형, 지속가능한 발전 등 국제인권규범을 강조
- 영향평가 수행 시 권고안이 강조하는 국제인권규범을 반영해야 하며, 이를 위해 인권 관련 단체 및 전문가와의 협력 필요

○ 영향평가 프레임워크의 개발 및 정교화

- 권고안에서 다루고 있는 가치 및 원칙의 포괄성, 분야별 정책조치의 광범위성으로 인해 영향평가의 범위, 방법, 환류체계 등을 포함하는 프레임워크 개발은 도전적 과제

- 근거 기반의 영향평가 프레임워크를 개발하고, 변화하는 기술환경 및 정책환경을 반영할 수 있도록 지속적인 연구 및 지원 필요

○ 영향평가 실시 주체의 권한 강화

- 권고안은 AI 도입에 따른 일반적 영향과 더불어 AI 시스템 수명주기 전반에서 의사결정에 사용된 알고리즘, 데이터, 관련된 AI 행위자 등을 모니터링할 것을 요구
- 특정 시스템의 모니터링을 위해서는 정보 접근성을 가능케 하는 법제도적 절차가 선행되어야 하며, 영향평가 실시 주체가 관련 정보를 확인·분석할 수 있도록 그 권한을 명시할 필요

○ 부처 간 협력체계 마련

- 권고안에 제시된 윤리영향평가 필요사항 중 일부는 기존의 영향평가제도(성별, 아동, 환경영향평가 등) 및 국가승인통계를 개선·활용함으로써 충족될 수 있음
- 중복적 영향평가 및 조사에 따른 행정적 소요를 막기 위해, 영향평가 프레임워크 개발과 실행에 있어 부처 협력체계 활성화

○ 공공에서의 AI 시스템 활용에 대한 영향평가

- 권고안은 공공영역에서 사용되는 AI 시스템의 영향평가 중요성을 별도로 강조하고 있으며, 감사가능성·추적가능성·설명가능성을 포함하는 영향평가 절차 프레임워크 요구
- 공공영역에서의 AI 시스템 조사·평가·모니터링과 관련된 법제도는 마련되어 있지 않은 상황으로, 추후 관련 제도의 도입 및 영향평가 절차 구체화 필요

② (중장기) 윤리적 관점과 경쟁력 간 균형을 갖춘 정책 개발³⁾

유네스코 권고는 향후 EU의 「인공지능법(안)」은 물론 각국에서 개발되는 AI 정책의 방향성을 설정할 것으로 예측됨에 따라, 국제적 추세에 맞게 윤리적 관점을 포함하면서도 기술 및 산업 경쟁력을 확보할 수 있는 균형적 정책의 발굴 필요

3) AI 윤리와 관련한 시민단체 전문가 1인과 법전문가 1인의 자문 내용을 반영하여 작성됨.

○ 다중 이해관계자의 참여에 기반한 정책 거버넌스 확립

- 공공, 대학, 기업 등 AI 시스템 기획과 개발에 참여하는 이해관계자 모두가 참여하는 정책 거버넌스의 확립이 요구되며, 특히 AI 공급·수요 기업의 현장성을 적극 반영할 수 있는 체계 마련 필요
- 우리나라 기업의 경우 일부 대기업을 제외하고는 AI 윤리 준수에 대한 인식이 부족한 상황으로, 모든 규모와 업종의 기업을 정책 논의에 포함함으로써 정책성 다양성과 포용성을 확보
- 아울러, ‘이루다’ 사건이 보여주듯 AI의 문제적 결과와 시민사회의 반발은 그 개발 기업에 중대한 영향을 미칠 수 있어, 시민사회와의 대화를 통해 규제·기술·산업 성장·시민사회 요구 간 합의 과정 마련

○ 국제적 모니터링 및 공조 강화를 통한 AI 리더십 발휘

- 유네스코에 따르면 AI 윤리 권고 발표와 더불어 20여 개 국가가 정책 개발 과정에 착수했으며, 그 내용과 도입 시기는 우리나라 정책에 직·간접적 영향을 미칠 영향
- 유네스코 권고에 사인을 한 것은 물론, 최근 중국이 AI 윤리 촉진을 위한 제도의 마련과 국제적 논의에 적극적인 것은 AI 윤리 문제의 해결 없이는 국제적 리더십 확보가 어렵다는 판단에 따름
- ‘AI에 대한 규제 및 정책 vs. 기술·산업 성장’이라는 이분법적 구도를 벗어나, 글로벌 AI 윤리 리더십을 통해 우리나라 AI 기술·산업 생태계의 확장을 위한 돌파구 마련 필요

○ 범부처 협력을 확보할 수 있는 AI 윤리 추진체계 구축

- 현재 과기정통부를 컨트롤 타워로 하여 AI 윤리를 다루고 있으나, 각 부처에서 관련 지침 및 가이드라인 등을 개별적으로 발표하는 한편 관련 업무가 여러 산하기관에 분산되어 있는 등 정책적 응집성 부족
- 유네스코 권고의 범위를 고려할 때 일반행정부처, 경제부처, 사회부처의 모든 업무에 있어서 AI 윤리 관점의 필요 여부를 점검해야 하며, 점검·모니터링·후속조치의 전 과정에 대한 추진체계가 요구됨

○ AI 윤리 관련 생태계 성장 전략 마련

- AI 윤리의 확보를 반드시 규제의 관점으로 볼 필요는 없으며, AI 윤리 컨설팅, AI의 위해 가능성에 따른 예방·대응 기술 산업 발전 등 새로운 산업 영역과 민간의 성장을 위한 전향적 기회로 인식
- 더불어 AI 윤리 실천을 위한 기술 등의 방안을 마련하는 기업에 인센티브를 부여함으로써, 긍정적 산업 효과를 창출할 필요

IT & Future Strategy 보고서

- 제1호(2021. 1. 15.) 「주목받는 인공지능(AI) 9대 핵심 기술 분석 및 주요 시사점」
- 제2호(2021. 2. 23.) 「초대규모 AI 모델(GPT-3)의 부상과 대응 방안」
- 제3호(2021. 3. 3.) 「데이터 기반 포스트 코로나 이슈 분석과 10대 메가트렌드」
- 제4호(2021. 4. 26.) 「코로나19 1년, 우리나라의 변화」
- 제5호(2021. 5. 26.) 「인공지능 플랫폼 글로벌 선도기업 분석」
- 제6호(2021. 7. 2.) 「데이터 기반 탈중앙화 기술의 사회적 함의 분석」
- 제7호(2021. 7. 9.) 「사람 중심의 디지털 거버넌스를 위한 정책방향」
- 제8호(2021. 11. 11.) 「보건의료 분야 데이터 인공지능 플랫폼 분석 및 발전방향」
- 제9호(2021. 12. 6.) 「데이터로 살펴본 우리나라 AI 이슈와 정책과제」
- 제10호(2021. 12. 9.) 「유네스코 AI 윤리권고 주요 내용 및 시사점」

1. 본 보고서는 방송통신발전기금으로 수행한 정보통신·방송 연구개발 사업의 결과물
이므로, 보고서의 내용을 발표할 때는 반드시 과학기술정보통신부 정보통신·방송 연구
개발 사업의 연구결과임을 밝혀야 합니다.
2. 본 보고서 내용의 무단전제를 금하며, 가공·인용할 때는 반드시 출처를「한국지능정
보사회진흥원(NIA)」이라고 밝혀 주시기 바랍니다.
3. 본 보고서의 내용은 한국지능정보사회진흥원(NIA)의 공식 견해와 다를 수 있습니다.