
사람의 성장을 지원하는

「교육분야 인공지능 윤리원칙」(시안)

2022. 1. 27.



교 육 부

순서

I. 추진 배경 및 경과	1
II. 윤리원칙 개요	3
III. 주요 내용	6
IV. 윤리원칙별 설명 자료	7
1. 서문	7
2. 대원칙	8
3. 3대 기본원칙	9
4. 9대 세부원칙	12
5. 실천과제(안)	21
【참고1】 「교육분야 인공지능 윤리원칙」 구성 체계 및 근거 ...	22
【참고2】 국내외 주요 AI 윤리원칙 주요내용	23
【참고3】 사람이 중심이 되는 「인공지능 윤리기준」	26
【참고4】 (요약) 유네스코 인공지능 윤리 권고문	29
【참고5】 AI와 교육에 대한 베이징 합의	33

I. 추진 배경 및 경과

□ 추진배경

- 첨단 디지털기술이 사회 전반에 빠르게 도입·확산되면서 인공지능 기술에 의해 발생하는 윤리적 이슈의 증대

《인공지능 관련 주요 윤리적 이슈 사례》

인공지능 챗봇 '이루다' 논쟁(21.1.)	영국 대입시험 알고리즘 차별 논쟁 (20.8.)	미국 교사평가 알고리즘 투명성 논란(17.5.)
혐오 발언, 성적인 대화 등으로 출시 1달여 만에 서비스 중단	영국 대입 시험에서 학습자 거주지역에 따른 알고리즘 차별로 교육 불평등 강화 논란	공립학교 교사 고용에 있어 비밀 알고리즘 평가를 실시하여 교사 연맹이 소송 제기

- 세계 각국 및 주요 국제기구는 인공지능(AI)의 윤리적 개발·활용과 신뢰성 확보를 위한 다양한 기준과 원칙을 마련 중
 - ※ (EU) 신뢰할 수 있는 인공지능윤리 가이드라인(19) (OECD) 인공지능 권고안(19.5.) (일본) 인간중심의 인공지능사회 원칙(19.3.), (UNESCO) 인공지능윤리에 대한 권고문(21.11.)
 - 국내에서도 인공지능 윤리기준* 등을 마련하여 AI의 윤리적 개발 및 활용을 도모하고 있으나, 교육분야의 특수성을 고려한 논의는 부족
 - * 「인공지능(AI) 윤리기준」(과기정통부, '20.12.), 「알고리즘 윤리헌장」(카카오, '18) 등
 - 2020년 원격수업이 전면 실시된 이후, 교육분야에 인공지능 도입* 등 디지털 전환이 본격화됨에 따라 교육계가 준수할 규범 필요 증대
 - * 서울,경남교육청에서 "AI·에듀테크 기반 교육 지원" 등을 22년도 업무계획으로 발표, 충남교육청은 '2021 인공지능교육 성과보고회' 추진(21.12.)
 - ※ UNESCO는 모니터링, 능력 평가, 학습자 행동 예측 등에 관한 것이라면, 학습에 사용되는 AI는 엄격하게 관리되어야 함을 권고(21.11.)
- ☞ 인공지능이 교육현장에서 안전하게 학습자의 성장을 지원할 수 있도록 사회적 담론 형성 및 자율규제의 촉매제로서 윤리원칙 마련 추진

□ 추진경과 및 향후 일정(안)

- 전문가TF 운영(21.9월~11월) → 교원, 연구자, 관계부처 등 의견수렴(21.12월~22.1월) → 시안 발표(22. 1. 27.) → 교육당사자들의 논의를 통한 보완(~22년)
- * 시도교육청, 대학, 전문가·기관·학회, 관계부처, 교사연구회 등과 의견 수렴 계속

참고 : 교육분야 인공지능(AI) 활용의 가능성

① 교육에서 인공지능 활용 가능 분야

분류	기능	예시
학습 관리	- 출석, 과제 제출 관리 - 학업 중단 위기 학생 관리	- 교육용 챗봇
학습 / 평가	- 학생 맞춤형 학습/평가 지원 - 양질의 교육에의 접근 보장	- 지능형 교육 시스템(ITS) - 대화 기반 교육 시스템(DBTS)
교사 지원	- 자동 과제 관리 시스템 - 자동 평가 시스템 - 자동 토론 시스템	- 인공지능 교사, 인공지능 교실* * 시간표 편성, 과제 제출 등 교실 관리에 인공지능 시스템 도입
평생 교육	- 인공지능을 통한 평생교육 시스템 구축	- 인공지능 평생학습 안내자 - 인공지능 평생학습 포트폴리오

※ AI and Education(Guidance for Policy-maker) (UNESCO, '21.4.) 참고

② 교육현장의 움직임

- 학교현장 : (똑똑! 수학탐험대) 인공지능 활용 초등수학수업 지원시스템('20.9), (AI톡톡) 미래교육지원플랫폼으로 맞춤형 학습평가 지원('20.9), (책열매) 한 학기 한 권 읽기 지원 웹서비스('21.9), (체리팻) 문제풀이 진단 등 학생맞춤형 AI튜터('21.5, 서울대부설학교진흥원)
- 교육청-기업 협약 : 경남교육청 - (주)매스프레소 등 4개기업('20.12), 서울시교육청 - LG CNS('21.6.7), 서울 부산시교육청 - 웅진씽크빅·클래스팅 업무협약 체결 등(서울 '21.9.9, 부산 '21.11)
- 대학 : (고려대 AI선배) 20년간 수강 데이터를 학습시켜, 교과·비교과·자율활동 추천, 워크숍 안내 등을 통해 학생생활 관리 및 진로 선택 지원('20.7), (아주대) AI기반 학생 맞춤형 프로그램 제공('20) (한림대) AI활용 맞춤형 수업 모델 HHL 개발('21) 등

③ 인공지능 활용의 교육적 효과성

- (실증연구 사례)

Bloom(1984)	Corbett(2001)	교육부 정책연구(2021)
전통적 교실수업 대비 1:1 튜터링을 받은 학생들의 성취도 평균이 2-표준편차(2-sigma)만큼 향상	'Cognitive Computer Tutor'를 사용하여 맞춤형으로 문제를 푼 경우 평가 결과 향상	20년 연구학교 대상 인공지능 활용 초등수학수업 지원시스템 사용 결과, 학생 학습 성취도 향상

- (해외 사례)

분야	내용
학습	(MindSpark) 인도의 AI기반 맞춤형교육시스템으로, 수학과 힌디어에서 각각 표준편차의 37%와 23% 성적향상 효과
정서-태도	(SMILE) 미국 스탠포드대의 질문-학습 프로그램으로 학생이 만든 질문을 AI가 5등급으로 분류 및 평가, 장기간(1년 이상) 학습시 비판적 사고능력 향상
사회성	(Kiwi) 자폐스펙트럼장애 학생의 수학학습을 돕는 로봇으로, 성취분석과 행동 관찰을 통해 자폐아동 17명의 수학 학습능력 뿐만 아니라 92%는 사회성 향상

II. 윤리원칙 개요

《 미래교육 핵심가치 》

- 1기·2기 미래교육위원회 운영('19~'20), '포스트 코로나 교육 대전환 대화'('20.6~8월) 등 현장·전문가 등과의 의견수렴을 바탕으로 배움과 성장의 포용성, 지속성, 자율성을 미래교육 방향으로 도출

※ 미래교육 전환을 위한 10대 정책과제('20.10.5.) 발표 등

핵심 가치	포용성	지속성	자율성
논의 내용	전 생애 걸친 교육 받을 권리 보장 강화	공동체와 상생하는 교육 생태계 조성	자율과 책임에 기반한 유연한 교육체제로의 전환

- 미래교육 핵심가치에 기반하여 첨단기술의 교육적 활용 모색 등 교육의 디지털 전환을 체계적으로 준비 필요

☞ 특히, 미래사회의 핵심기술인 인공지능을 사람이 주체적으로 통제하면서 교육적으로 활용할 수 있도록 방향 설정 필요

1 기본 방향

- (목적) 인공지능을 교육현장에서 윤리적이고 안전하게 활용함으로써 사람의 성장을 지원할 수 있도록 교육분야 윤리원칙 천명

☞ 세계적으로 교육분야만의 인공지능 윤리원칙 제정 사례를 찾기 어려운 상황에서, 인공지능의 교육적 활용 확대에 대비하여 교육부 차원의 선제적 규범 마련 추진

- (역할) 인공지능의 교육분야 도입 및 교육현장(에듀테크 기업 포함) 적용 등에 대한 가이드라인 역할 수행

☞ 교육현장에서 인공지능의 안전한 활용에 대한 사회적 담론 형성 및 학계의 연구, 자율적 논의 촉진 기대

2 검토 근거

○ 교육기본법 제17조의3

제17조의3(학습윤리의 확립) 국가와 지방자치단체는 모든 국민이 학업·연구·시험 등 교육의 모든 과정에 요구되는 윤리의식을 확립할 수 있도록 필요한 시책을 수립·실시하여야 한다.

○ 원격교육법 제4조 및 제21조 등

※ 2022.3.25. 시행

제4조(국가와 지방자치단체의 책무) ① 국가와 지방자치단체는 원격교육에 관한 정책을 수립·시행하며 원격교육의 질을 향상시키기 위한 정책을 추진하기 위하여 필요한 예산상의 조치를 하여야 한다.

제21조(교원의 원격교육 전념을 위한 환경 조성) 교육부장관·교육감 및 대학등의 장은 교육기관의 교원이 질 높은 원격교육을 위하여 전념할 수 있도록 필요한 지원을 할 수 있다.

3 적용 대상 및 권리·책임 주체

- (대상) 교육기관* 및 교육활동을 지원하는 행정기관에서 **활용되는 인공지능**
* 유·초·중등·고등 및 평생교육 단계에서 정규 및 비정규의 교육을 실시하는 모든 형태의 교육기관(시설)

- (권리·책임 주체) **교육당사자 및 관계자**

- ① **이용자** : 교육기관 등에서 인공지능 서비스와 제품을 이용하는 교수자 및 학습자(미성년 학습자의 경우 학부모 포함)
- ② **개발자** : 교육기관 등에서 활용하는 인공지능 학습지원시스템·플랫폼 및 콘텐츠·프로그램 등의 개발자
- ③ **관리자** : 교육기관 등의 인공지능 도입·활용을 기획·관리하고 행·재정 정책의 수립 및 시행을 담당하는 업무수행자

4 원칙의 성격

- (형식) **교육부장관의 고시**(처분성 없음)
- (내용) **자발적 실천과 준수를 독려하는 도덕적 규범**이면서 **자율 규제**

- (기존 원칙과의 관계) 범정부 인공지능 윤리기준('20.12.)과 유네스코 윤리 권고('21.11.25.) 등을 준수하면서 가소성이 큰 학습자를 다루는 교육분야 인공지능에 대한 특수 규범

< 교육분야에 적용가능한 보편적 인공지능 윤리기준 >

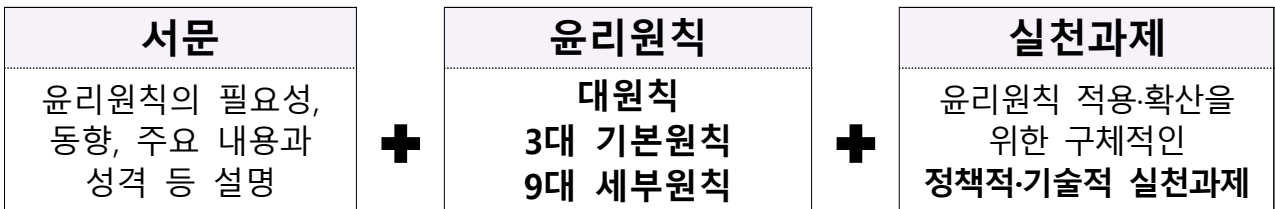
❖ 사람이 중심이 되는 「인공지능(AI)윤리기준」('20.12. 범정부 합동)>

1. 3대 기본원칙 ①인간 존엄성 원칙 ② 사회 공공선 원칙 ③기술의 합목적성 원칙
2. 10대 핵심요건(기본원칙을 실현할 수 있는 세부요건)
 - ①인권보장 ②프라이버시 보호 ③다양성 존중 ④침해금지 ⑤공공성 ⑥연대성
 - ⑦데이터 관리 ⑧책임성 ⑨안전성 ⑩투명성

❖ 유네스코 인공지능 윤리 권고('21.11, 제41차 총회 채택)

1. 적용범위(교육 관련) : 디지털 사회에서의 삶은 새로운 교육적 실행, 윤리적 영향, 비판적 사고, 고용가능성, 시민참여에 영향을 주는 책임 있는 실천을 요구
3. 정책 조치(교육 관련)
 - 충분한 AI 리터러시 교육 제공
 - 교수(교사) 연수 및 이러닝에서 AI 기술의 윤리적이고 책임있는 사용을 지원
 - AI 윤리 커리큘럼을 개발하고, AI기술교육과 인문·윤리·사회교육간 협력 도모

5 원칙의 구성

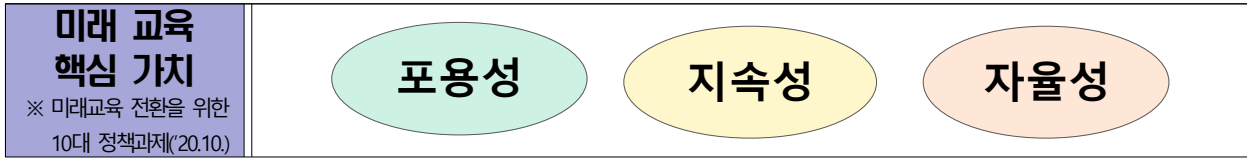


6 주기적 개선·보완

- 인공지능기술의 발전, 교육현장에서의 활용 실태, 논의의 성숙 등을 점검하여 원칙의 타당성 검토 등의 조치를 주기적으로 실시
 - ※ 첫 번째 개선 등의 조치는 3년 이내(2024년 12월 31일까지) 실시(훈령·예규 등의 발령 및 관리에 관한 규정 제7조)
- 동 윤리원칙의 법률적 근거 마련*을 통해 자율규범으로서의 권위 부여 및 위상 강화 추진
 - * 「인공지능교육진흥법안(안민석의원, '21.5.17.)」 등 국회 논의시 근거 마련 추진

Ⅲ. 주요 내용

교육분야 인공지능 윤리원칙



비전 인간다움과 미래다움이 공존하는 교육 패러다임 실현

대원칙 사람의 성장을 지원하는 인공지능

기본 원칙 학습자의 주도성 강화 교수자의 전문성 존중 기술의 합목적성 제고

세부 원칙	사람 + 공동체 + 기술	교육분야 인공지능은 1. 인간성장의 잠재가능성을 이끌어낸다. 2. 모든 학습자의 주도성과 다양성을 보장한다. 3. 교육당사자 간의 관계를 공고히 유지한다. 4. 교육의 기회균등 실현을 통해 공정성을 보장한다. 5. 교육공동체의 연대와 협력을 강화한다. 6. 사회 공공성 증진에 기여한다. 7. 모든 교육당사자의 안전을 보장한다. 8. 데이터 처리의 투명성을 보장하고 설명 가능해야 한다. 9. 데이터를 합목적적으로 활용하고 프라이버시를 보호한다.
--------------	--	---

실천 과제

(정책적 과제) ①AI윤리이슈 발굴 ②AI윤리교육 강화 ③AI윤리 적용 지원
④학술연구 지원 ⑤소통체계 구축

(기술적 과제) ①안전성 판단도구 개발 ②현장수요맞춤형 AI개발 지원 ③교육분야 데이터관리 표준 개발 ④교육분야 데이터 윤리를 위한 보안시스템 개발



IV. 윤리원칙별 설명 자료

1 서 문(안)

코로나19로 전 사회의 디지털 전환이 가속화됨에 따라 인공지능이 우리 생활에 폭넓게 스며들고 있다. 그간 첨단 과학기술의 연구 영역이고 제조업 등 산업현장에서 주로 논의되던 인공지능을 현재는 문화·예술, 경제·금융, 교통·물류, 공공·행정, 보건·의료, 치안·국방, 재난·안전 등 우리를 둘러싼 사회의 각 영역에서 접할 수 있다. 한편, 미래세대의 삶에서는 인공지능이 인간의 사고방식과 의사결정, 사회적 참여, 노동과 고용형태, 교육의 실천 방식 등에서 현재보다 더욱 광범위하게 영향을 미칠 것이다.

원격수업의 전면 도입을 계기로 교육 분야에서 디지털 기술 사용이 늘어나면서, 교수·학습의 과정에 인공지능을 활용하는 사례도 증가하고 있다. 교수자와 학습자 간 고차원 지적 활동인 교육의 과정에 인공지능 기술이 활용되기 시작한 것이다. 교육에서 인공지능의 활용은 개별 맞춤형 학습, 교육격차 해소, 정책 입안자의 합리적 의사결정 등에 이바지할 것으로 기대되지만 예상치 못한 오류, 오용과 역기능, 데이터 편향성 등의 윤리 문제에 대한 우려도 존재한다.

국내외 정부, 국제기구, 기업, 단체 등에서는 인공지능 기술이 인류에게 이로운 방향으로 발전하도록 개발 과정에서 윤리를 강조하고 사용 규범을 마련하는 등 노력하고 있다. 이런 맥락에서 최근 OECD의 「인공지능 권고안(’19.5.)」, EU의 「인공지능 윤리 가이드라인(’19)」, UNESCO의 「인공지능 윤리 권고(’21.11.)」 등이 발표되었다. 또한 국내에서도 관계부처가 함께 「사람이 중심이 되는 인공지능 윤리기준(’20.12.)」을 제안하여 사안별 또는 분야별로 인공지능에 대한 자율 규범을 마련할 수 있는 근거를 제공하고 있다.

교육 분야에 활용되는 인공지능은 ‘포용성’, ‘지속성’, ‘지울성’이라는 미래교육의 핵심가치를 토대로 인간다움과 미래다움이 공존하는 교육을 실현하는 데 이바지해야 한다. 인공지능은 ‘사람의 성장을 지원’하기 위한 목적에 맞게 활용되어야 하며, 결코 인간의 존엄성과 교육당사자들 간의 관계를 훼손해서는 안 된다. 이를 위해 ‘학습자의 주도성 강화’, ‘교수자의 전문성 강화’, ‘기술의 합목적성 제고’의 3대 기본 원칙을 세우고 9개의 윤리원칙과 정책적·기술적 실천과제를 제시한다.

이 윤리원칙은 교육계의 자발적인 실천과 준수를 독려하는 도덕 규범이면서 자율 규제의 성격을 지닌다. 교육 분야의 디지털 대전환을 위한 지침 역할도 수행한다. 아울러, 인공지능 기술의 발전 정도와 교육현장에서의 활용 실태를 점검하고 교육당사자와 연구자들의 숙의 과정을 거쳐, 변화하는 미래세대에 맞도록 이 원칙을 지속해서 보완하고 개선해 나가야 할 것이다.

■ 사람의 성장을 지원하는 인공지능

교육분야 인공지능은 사람의 전 생애에 걸쳐 전인적 성장 지원을 최고 가치로 삼으며, 사람의 인격을 존중하고 개성을 중시하여 사람의 능력이 효과적으로 발휘될 수 있도록 제공되어야 한다.

- 교육분야 인공지능을 개발하고 활용하는 것은 궁극적으로 사람의 성장을 지원하기 위한 것임을 분명히 해야 함
- 교육분야 인공지능이 사람의 성장을 지원하는 것이라는 대원칙은 교육을 위하여 인공지능이 활용되는 모든 국면에 적용됨

검토 근거

교육기본법

제2조(교육이념) 교육은 홍익인간(弘益人間)의 이념 아래 모든 국민으로 하여금 인격을 도야(陶冶)하고 자주적 생활능력과 민주시민으로서 필요한 자질을 갖추게 함으로써 인간다운 삶을 영위하게 하고 민주국가의 발전과 인류공영(人類共榮)의 이상을 실현하는 데에 이바지하게 함을 목적으로 한다.

제12조(학습자) ① 학생을 포함한 학습자의 기본적 인권은 학교교육 또는 사회교육의 과정에서 존중되고 보호된다.

② 교육내용·교육방법·교재 및 교육시설은 학습자의 인격을 존중하고 개성을 중시하여 학습자의 능력이 최대한으로 발휘될 수 있도록 마련되어야 한다.

① 학습자의 주도성 강화

교육분야 인공지능은 학습자의 개성을 교육의 모든 과정에 걸쳐 존중하고 학습자 개인의 꿈을 실현하는 일을 지원하기 위해 활용되어야 한다.

- 인공지능은 인간의 삶을 이롭게 하는 도구의 일종일 뿐, 인간을 대체하거나 인간 고유의 존엄성을 해칠 수 없음
 - 인공지능의 발전에 따라 자동화된 의사결정에 의해서 인간의 권리 구제에 공백이 발생할 우려가 있으므로,
 - 교육당사자가 인공지능을 주도적으로 활용하고 제어함으로써 인간의 존엄성과 기본 권리를 보장할 수 있도록 노력해야 함
- 또한, 교육분야에서 인공지능의 활용은 교육의 획일화를 지양하고 학습자의 필요에 대한 맞춤형 교육을 제공하기 위한 것이어야 함

검토 근거

교육기본법

제12조(학습자) ① 학생을 포함한 학습자의 기본적 인권은 학교교육 또는 사회교육의 과정에서 존중되고 보호된다.

② 교육내용·교육방법·교재 및 교육시설은 학습자의 인격을 존중하고 개성을 중시하여 학습자의 능력이 최대한으로 발휘될 수 있도록 마련되어야 한다.

③ 학생은 학습자로서의 윤리의식을 확립하고, 학교의 규칙을 준수하여야 하며, 교원의 교육·연구활동을 방해하거나 학내의 질서를 문란하게 하여서는 아니 된다.

2 교수자의 전문성 존중

교육분야 인공지능은 교수자의 전문성을 존중해야 하며, 교수자가 가진 전문성이 효과적으로 발휘될 수 있도록 개발·활용되어야 한다.

- 교육분야 인공지능은 교수자를 보조하고 지원하는 학습도구로서 교수자의 부담을 줄이고 교수·학습의 질을 높이는 방향으로 활용될 필요가 있음
 - 교육분야 인공지능은 교수자 고유의 전문성 및 양질의 교수·학습 내용과 방법을 지원해야 함
- 또한, 교수자는 인공지능이 분석한 학습자의 데이터 등을 바탕으로 전문성을 발휘해 교육의 효과성을 높일 책무를 가짐

검토 근거

헌법

제31조 ④교육의 자주성·전문성·정치적 중립성 및 대학의 자율성은 법률이 정하는 바에 의하여 보장된다.

교육기본법

제14조(교원) ①학교교육에서 교원(敎員)의 전문성은 존중되며, 교원의 경제적·사회적 지위는 우대되고 그 신분은 보장된다.

②교원은 교육자로서 갖추어야 할 품성과 자질을 향상시키기 위하여 노력하여야 한다.

③교원은 교육자로서 지켜야 할 윤리의식을 확립하고, 이를 바탕으로 학생에게 학습 윤리를 지도하고 지식을 습득하게 하며, 학생 개개인의 적성을 계발할 수 있도록 노력하여야 한다.

UNESCO 윤리권고 104

회원국은 다른 사안보다도 교육, 교사 연수, 온라인 교육에서 인공지능의 윤리적이고 책임 있는 사용에 대한 연구 계획을 장려하여 이 분야와 관련된 기회는 향상시키고 어려움 및 위험성은 완화해야 한다.

3 기술의 합목적성 제고

교육분야 인공지능은 교육적 가치를 저해하지 않는 선에서 자율적으로 개발·활용될 수 있으며, 교육이 이루어지는 모든 과정에서 안전성을 확보해야 한다.

- 교육분야 인공지능의 개발, 활용, 평가는 정부의 일방적인 주도보다 다양한 분야 이해당사자들의 자발적이고 적극적인 관심과 참여로 이루어질 필요가 있음
 - 교육적 가치를 저해하지 않는 선에서 민간의 자율성을 존중하면서 사용할 수 있도록 해야 함
- 다만, 교육분야 인공지능의 활용은 교육이 이루어지는 모든 과정에서 안전성을 확보하는 것을 전제 조건으로 해야 함

검토 근거

원격교육법

제17조(원격교육콘텐츠 품질 관리)

② 교육부장관은 원격교육콘텐츠의 품질 관리를 위하여 품질 진단·평가, 개선 지원 등 필요한 정책을 수립·추진할 수 있다.

범정부 윤리기준

3대 기본원칙 ①인간 존엄성 원칙 ② 사회 공공성 원칙 ③기술의 합목적성 원칙

- 인류의 삶에 필요한 도구인 인공지능 기술은 사용자의 목적과 의도에 따른 고유한 목적과 수단적 가치를 지닌다.
- 인공지능은 궁극적으로 인간에게 도움이 되어야 한다는 목적에 맞도록 개발 및 활용되어야 한다.
- 인류의 삶과 번영을 위한 인공지능 개발 및 활용을 장려하여 진흥해야 한다.

① 인간성장의 잠재가능성을 이끌어낸다.

교육분야 인공지능은 인간 존엄성에 대한 존중을 바탕으로 인간성장의 잠재가능성을 이끌어낼 수 있도록 제공되어야 한다.

- 교육분야 인공지능을 활용하여 학습자의 데이터를 분석하는 경우, 인공지능의 진단 또는 예측 결과가 개별 학습자에 대한 **낙인효과**를 갖지 **않도록 노력해야 함**
- 인공지능의 지원을 받는 부분에 있어 **과도한 진단이나 예측은 자제** 하고, 인공지능을 활용한 **학습 진단이나 예측 결과의 고지에 있어** 서도 **신중을 기해야 함**

참고 사례

- **(미 경찰국 안면인식기술의 과도한 활용 사례('19))** 미국 경찰국은 범죄 수사 사건에 안면인식 기술을 활용하였으나 무고한 시민을 범죄자로 인식하여 일부 주 또는 시에서 안면인식 기술을 규제하고 있음
- **(학업중단 학생 예측 시스템)** 학생들의 출결정보, 성적 등을 활용하여 학업중단의 가능성을 예측하는 시스템이 개발될 경우 학업중단 위험도를 고정된 점수, 백분율보다는 저위험/중간 위험/고위험 등의 수준으로 교사에게 고지하여 학생의 개선 가능성을 존중하여야 함

검토 근거

교육기본법

제9조(학교교육) ③ 학교교육은 학생의 창의력 계발 및 인성(人性) 함양을 포함한 전인적(全人的) 교육을 중시하여 이루어져야 한다.

제12조(학습자) ② 교육내용·교육방법·교재 및 교육시설은 학습자의 인격을 존중하고 개성을 중시하여 학습자의 능력이 최대한으로 발휘될 수 있도록 마련되어야 한다.

제13조(보호자) ① 부모 등 보호자는 보호하는 자녀 또는 아동이 바른 인성을 가지고 건강하게 성장하도록 교육할 권리와 책임을 가진다.

제14조(교원) ③ 교원은 교육자로서 지녀야 할 윤리의식을 확립하고, 이를 바탕으로 학생에게 학습윤리를 지도하고 지식을 습득하게 하며, 학생 개개인의 적성을 계발할 수 있도록 노력하여야 한다.

② 모든 학습자의 주도성과 다양성을 보장한다.

교육분야 인공지능은 학습자의 자기주도성을 신장하는 범위에서 활용되고, 학습자의 개성과 다양성을 존중해야 한다.

- 인공지능의 교육적 활용은 학습자가 학습에 대한 자기주도성을 적극적으로 발현할 수 있도록 지원하는 것이어야 함
 - 교육분야의 인공지능은 교육내용·방법, 학습 진단과 진로 판단 등에 인공지능의 사용으로 학습자의 주도성이 억압되지 않도록 유의해야 함
- 인공지능의 활용은 교육의 획일화를 지양하고 학습자의 개별적인 필요에 대한 맞춤형 교육을 시도하기 위한 것이어야 함

참고 사례

- (다문화학생을 위한 AI 번역) 다문화가정이 증가함에 따라 학교에서 제공되는 한글 공지를 제대로 이해하기 어려운 다문화가정 학부모를 위해 학부모 맞춤형 번역을 제공하는 자연어처리 인공지능기술을 적용할 수 있음

검토 근거

교육기본법

제3조(학습권) 모든 국민은 평생에 걸쳐 학습하고, 능력과 적성에 따라 교육 받을 권리를 가진다.

제12조(학습자) ② 교육내용·교육방법·교재 및 교육시설은 학습자의 인격을 존중하고 개성을 중시하여 학습자의 능력이 최대한으로 발휘될 수 있도록 마련되어야 한다.

제14조(교원) ③ 교원은 교육자로서 지녀야 할 윤리의식을 확립하고, 이를 바탕으로 학생에게 학습 윤리를 지도하고 지식을 습득하게 하며, 학생 개개인의 적성을 계발할 수 있도록 노력하여야 한다

UNESCO 윤리권고 105

회원국은 모든 수준의 인공지능 교육 프로그램에서 여성, 다양한 민족 및 문화, 장애인, 소외·취약 계층 등 모든 사람의 참여 및 주도를 장려함과 더불어 ...

법정부 윤리기준

2. 10대 핵심요건 ①인권보장 ②프라이버시 보호 ③다양성 존중 ④침해금지 ⑤공공성 ⑥연대성 ⑦데이터 관리 ⑧책임성 ⑨안전성 ⑩투명성 .

- 인공지능 개발 및 활용 전 단계에서 사용자의 다양성과 대표성을 반영해야 하며, 성별연령장애인종종교국가 등 개인 특성에 따른 편향과 차별을 최소화하고, 상용화된 인공지능은 모든 사람에게 공정하게 적용되어야 한다.

- 사회적 약자 및 취약 계층의 인공지능 기술 및 서비스에 대한 접근성을 보장하고, 인공지능이 주는 혜택은 특정 집단이 아닌 모든 사람에게 골고루 분배되도록 노력해야 한다.

③ 교육당사자 간의 관계를 공고히 유지한다.

교육분야 인공지능은 일련의 교수·학습 활동 과정에서 교수자의 전문성을 바탕으로 교수자와 학습자, 학습자와 학습자 등 사람 간의 긍정적 관계 형성을 도울 수 있게 제공되어야 한다.

- 교육분야 인공지능은 사회공동체 일원의 권리 보호 및 배려를 통한 공동체 형성과 유지를 돕는 방향으로 적용되어야 함
 - 교육분야 인공지능이 학습자의 개별학습 경험을 강화함에 따라 사회적 관점에서 교육공동체 유지와 지속에 대한 논의가 필요함
 - UNESCO에서 강조하는 바와 같이 교육분야 인공지능은 교수자와 학습자, 학습자와 학습자 등 인간적 관계의 확대를 지원해야 함

참고 사례

- (학습자-인공지능의 1:1 관계) 인공지능을 활용한 개별학습 과정에서 학습자와 인공지능의 관계가 1:1로 고정되어 학습자와 교수자, 학습자와 다른 학습자 사이의 관계가 느슨해지고, 교육당사자 간 상호작용 저해될 수 있다는 우려가 있음
- (학습자의 정보 공유) 수업 과정에서 인공지능이 학습자를 진단하고 평가하여 자동화된 맞춤 서비스를 제공할 경우, 교수자에게 학습 정보를 공유하여 교수자가 학습자의 학습상태를 이해하고 판단할 수 있도록 지원할 필요가 있음

검토 근거

UNESCO 윤리권고 104

... 회원국은 전통적 교육 형태의 관계·사회적 측면 및 가치가 교사-학생 및 학생-교사 관계에서 매우 중요하며 이것이 교육으로의 인공지능 기술의 도입을 논할 때 고려되어야 한다는 점을 유념한 채로, 인공지능 기술이 학생 및 교사의 역량을 강화하고 경험을 증대시킬 수 있도록 보장해야 한다.

④ 교육의 기회균등 실현을 통해 공정성을 보장한다.

교육분야 인공지능은 모든 사회구성원이 지역·경제적 배경 등의 조건에 상관없이 교육 기회를 공정하게 보장받을 수 있도록 제공되어야 한다.

- 교육의 기회균등 실현을 위하여 인공지능을 활용한 교육에 대한 사회적 약자의 접근성 강화가 필요함
 - 교육의 기회균등은 「교육기본법」에서 정하는 기본원칙으로, 교육분야 인공지능을 활용함으로써 누리는 혜택으로부터 소외되는 사람이 없도록 노력해야 함

참고 사례

- (인공지능의 소수집단 차별) 네덜란드 정부는 복지혜택 부정수급 및 세금 사기 단속을 위해 인공지능 기반 위험탐지시스템(SyRI)를 개발·활용하였으나 소수 및 취약집단 차별 적용이 우려되어 2020년 2월 법원 판결에 의해 해당 시스템 활용을 철회하였음

검토 근거

헌법

제31조 ① 모든 국민은 능력에 따라 균등하게 교육을 받을 권리를 가진다.

교육기본법

제4조(교육의 기회균등) ① 모든 국민은 성별, 종교, 신념, 인종, 사회적 신분, 경제적 지위 또는 신체적 조건 등을 이유로 교육에서 차별을 받지 아니한다.

제12조(학습자) ② 교육내용·교육방법·교재 및 교육시설은 학습자의 인격을 존중하고 개성을 중시하여 학습자의 능력이 최대한으로 발휘될 수 있도록 마련되어야 한다.

법정부 윤리기준

1. 3대 기본원칙 ①인간 존엄성 원칙 ②사회의 공공선 원칙 ③기술의 합목적성 원칙

- 공동체로서 사회는 가능한 한 많은 사람의 안녕과 행복이라는 가치를 추구한다.
- 인공지능은 지능정보사회에서 소외되기 쉬운 사회적 약자와 취약 계층의 접근성을 보장하도록 개발 및 활용되어야 한다.
- 공익 증진을 위한 인공지능 개발 및 활용은 사회적, 국가적, 나아가 글로벌 관점에서 인류의 보편적 복지를 향상시킬 수 있어야 한다.

㉔ 교육공동체의 연대와 협력을 강화한다.

교육분야 인공지능은 그 활용에 있어 민·관·학·연의 협력을 지향하고 지속 가능한 교육생태계를 구축할 수 있도록 제공되어야 한다.

- 인공지능을 활용하여 교육을 바꾸어나가기 위해서는 교수자, 행정기관 등 교육당사자와 연구자 및 개발자, 에듀테크 기업 등의 다양한 관계자가 협력하고 함께 도전할 수 있는 기반이 마련돼야 함
- 더불어, 교육은 학습자를 미래사회의 일원으로 양성하는 것으로, 중장기적인 영향을 고려하는 등 미래세대에 대한 배려를 바탕으로 교육분야 인공지능의 활용이 이루어져야 함

참고 사례

- (민간기업과의 계약) 민간기업에서 개발한 인공지능 기반 학습지원시스템 서비스 및 제품에 대해 교육목적과 상황에 따라 당사자와 협의를 거쳐 계약 조건을 개선하는 속의 과정이 필요할 수 있음

검토 근거

교육기본법

제23조(교육의 정보화) ① 국가와 지방자치단체는 정보화교육 및 정보통신매체를 이용한 교육을 지원하고 교육정보산업을 육성하는 등 교육의 정보화에 필요한 시책을 수립·실시하여야 한다.

UNESCO 윤리권고 109

회원국은 개발도상국 등에서 민간기업으로 하여금 과학계가 연구를 위해 데이터에 쉽게 접근하는 것을 허용하도록 장려해야 함

범정부 윤리기준

2. 10대 핵심요건 ①인권보장 ②프라이버시 보호 ③다양성 존중 ④침해금지 ⑤공공성 ⑥연대성 ⑦데이터 관리 ⑧책임성 ⑨안전성 ⑩투명성

- 다양한 집단 간의 관계 연대성을 유지하고, 미래세대를 충분히 배려하여 인공지능을 활용해야 한다.
- 인공지능 전 주기에 걸쳐 다양한 주체들의 공정한 참여 기회를 보장하여야 한다.
- 윤리적 인공지능의 개발 및 활용에 국제사회가 협력하도록 노력해야 한다.

⑥ 사회 공공성 증진에 기여한다.

교육분야 인공지능은 학습자가 민주시민으로서 필요한 자질을 갖추므로써 개인의 행복 추구하고 함께 사회의 공익 추구에도 긍정적인 영향을 미치도록 적용되어야 한다.

- 교육은 공동체의 일원으로서의 민주시민을 양성하는 과정으로, 교육의 디지털 전환에 대응하여 미래세대를 위한 새로운 윤리의식의 정립이 필요하다는 공감대가 형성되고 있음
 - 이에 따라 교육분야 인공지능은 사회를 긍정적인 방향으로 이끌고 공공성 증진하는 방향으로 활용되어야 함

참고 사례

- (교육당사자의 의사결정 보조) 학습자 진로 탐색, 교수학습모델 개발 등 교육당사자의 의사결정을 보조하는 데 활용될 인공지능 시스템은 업무의 사회적 영향을 평가하고 진단할 수 있도록 개발되어야 함

검토 근거

교육기본법

제2조(교육이념) 교육은 홍익인간(弘益人間)의 이념 아래 모든 국민으로 하여금 인격을 도야(陶冶)하고 자주적 생활능력과 민주시민으로서 필요한 자질을 갖추게 함으로써 인간다운 삶을 영위하게 하고 민주국가의 발전과 인류공영(人類共榮)의 이상을 실현하는 데에 이바지하게 함을 목적으로 한다.

범정부 윤리기준

2. 10대 핵심요건 ①인권보장 ②프라이버시 보호 ③다양성 존중 ④침해금지 ⑤공공성 ⑥연대성 ⑦데이터 관리 ⑧책임성 ⑨안전성 ⑩투명성

- 인공지능은 개인적 행복 추구 뿐만 아니라 사회적 공공성 증진과 인류의 공동 이익을 위해 활용해야 한다.
- 인공지능은 긍정적 사회변화를 이끄는 방향으로 활용되어야 한다.
- 인공지능의 순기능을 극대화하고 역기능을 최소화하기 위한 교육을 다방면으로 시행하여야 한다

㉞ 모든 교육당사자의 안전을 보장한다.

교육분야 인공지능은 일련의 교수·학습 과정에서 나타날 수 있는 잠재적 위험을 방지하고 안전을 보장하는 방향으로 구현되어야 하며, 인공지능의 활용에 있어 책임주체가 명확히 설정되어야 한다.

- 교육은 학습자의 가치관과 인격 형성의 주된 요인이므로 교육분야 인공지능의 안전성을 보장하는 것은 매우 중요함
 - 특히, 교육 현장에서의 학습자는 대부분 미성년으로 인공지능 데이터의 불건전·편향 등 인공지능 안전성이 확보되지 않을 경우, 미래세대에 미칠 부정적 영향이 매우 클 수 있음
- 따라서 교육분야 인공지능의 잠재적 위험으로 인한 피해를 최소화하고 실효성 있는 피해구제를 위한 책임 주체와 범위를 명확히 해야함

참고 사례

- (인공지능의 데이터 편향 가능성) 인공지능이 편향된 데이터로 모형 훈련이 이루어지거나 데이터 수집·처리 과정에서 해킹 등의 외부 공격으로 부적절한 데이터가 포함되는 경우 학습자에게 편향된 출력 메시지를 생성하여, 학습자의 가치관과 인격 형성에 부정적 영향을 미칠 수 있음
- (인공지능의 오판 및 범죄 가능성) 자율주행차의 사고, 미국 경찰이 안면인식기술을 활용해 부당 체포한 사례 등에 대하여 책임소재 여부가 불분명해지는 경우가 있을 수 있음

검토 근거

교육기본법

제12조(학습자) ① 학생을 포함한 학습자의 기본적 인권은 학교교육 또는 사회교육의 과정에서 존중되고 보호된다.

제13조(보호자) ① 부모 등 보호자는 보호하는 자녀 또는 아동이 바른 인성을 가지고 건강하게 성장하도록 교육할 권리와 책임을 가진다.

EU의 개인정보보호법(GDPR:General Data Protection Regulation)

제5조(f) (f) 개인정보의 적절한 보안을 보장하는 방식으로 처리해야 한다.(‘무결성과 기밀성’).

법정부 윤리기준

2. 10대 핵심요건 ①인권보장 ②프라이버시 보호 ③다양성 존중 ④침해금지 ⑤공공성 ⑥연대성 ⑦데이터 관리 ⑧책임성 ⑨안전성 ⑩투명성

- 인공지능 개발 및 활용과정에서 책임주체를 설정함으로써 발생할 수 있는 피해를 최소화하도록 노력해야 한다.
- 인공지능 설계 및 개발자, 서비스 제공자, 사용자 간의 책임소재를 명확히 해야 한다.
- 인공지능 개발 및 활용 전 과정에 걸쳐 잠재적 위험을 방지하고 안전을 보장할 수 있도록 노력해야 한다.
- 인공지능 활용 과정에서 명백한 오류 또는 침해가 발생할 때 사용자가 그 작동을 제어할 수 있는 기능을 갖추도록 노력해야 한다.

⑧ 데이터 처리의 투명성을 보장하고 설명 가능해야 한다.

교육분야 인공지능은 데이터의 수집·정제·선택 등의 과정이 투명해야 하고, 알고리즘과 데이터의 처리 과정이 교육당사자가 이해할 수 있는 언어로 설명 가능해야 한다.

- 교육이 이루어지는 과정에서 인공지능의 영향을 예측하고 진단하기 위해 데이터 수집·정제·선택 등의 처리 과정이 투명해야 함
- 또한, 교수·학습과정 등에서 활용되는 인공지능이 적재적소에 활용되고 교육적 효과를 발휘하기 위해서는,
 - 인공지능이 분석한 결과에 대한 근거와 판단 자료가 교육당사자가 이해할 수 있는 언어나 형태로 제공되어야 함

참고 사례

- (인공지능 평가 알고리즘 검증의 필요성) 인공지능 기반 평가 알고리즘이 검증되지 않을 경우, 평가 결과에 대한 타당성, 공정성 논란이 발생할 수 있음
- (알고리즘 블랙박스 문제) 다양한 분야에서 활용되는 인공지능 알고리즘의 블랙박스 성격으로 인해, 오류 발생 여부, 발생 지점 등을 판단하기 어려워 이용자에게 피해가 발생해도 책임소재 규명이 어려울 수 있음

검토 근거

교육기본법

제23조(교육의 정보화) ② 제1항에 따른 정보화교육에는 정보통신매체를 이용하는 데 필요한 타인의 명예·생명·신체 및 재산상의 위해를 방지하기 위한 법적·윤리적 기준에 관한 교육이 포함되어야 한다.

EU 「신뢰할 수 있는 인공지능 윤리원칙」

인공지능 시스템에 의해 내려진 결정은 인간이 이해하고 추적할 수 있어야 한다는 점 강조

범정부 윤리기준

2. 10대 핵심요건 ①인권보장 ②프라이버시 보호 ③다양성 존중 ④침해금지 ⑤공공성 ⑥연대성 ⑦데이터 관리 ⑧책임성 ⑨안전성 ⑩투명성

- 사회적 신뢰 형성을 위해 인공지능의 투명성과 설명 가능성을 높이고, 타 원칙과의 상충관계를 고려하여 활용 상황에 적합한 수준의 투명성과 설명 가능성을 높이려는 노력을 기울여야 한다.
- 인공지능기반 제품이나 서비스를 제공할 때 인공지능의 활용 내용과 활용 과정에서 발생할 수 있는 위험 등의 유의사항을 사전에 고지해야 한다.

㉠ 데이터를 합목적적으로 활용하고 프라이버시를 보호한다.

교육분야 인공지능의 개발 및 활용을 위해 수집되는 데이터는 활용 목적에 적합한 정도로 수집되고 본래 목적에 부합하도록 활용되어야 하며, 데이터의 처리 과정에서 교육당사자의 프라이버시를 보호해야 한다.

- 교육 분야에 인공지능에 활용되는 데이터는 합목적성이 강조되어야 하며 불필요한 데이터를 과도하게 수집하지 않도록 해야 함
- 비식별화되기 어려운 개인정보 및 민감정보의 공개와 활용에 대하여 당사자에게 적절히 고지하고 선택할 수 있도록 해야 함
 - 교육과정에서 인공지능에 제공되는 데이터는 교육당사자의 학습이력, 건강·상담기록 등 민감한 정보일 경우가 많을 수 있어 보다 엄격한 관리가 필요함

참고 사례

- (디바이스 제조사와의 상충관계 우려) SI 서비스 사용을 위해 사용자의 결정과 상관없이 제조사가 과도한 개인정보를 제공하도록 하는 상충관계가 발생할 수 있음
- (얼굴인식 인공지능) 공개된 알고리즘 모형을 이용해 인공지능이 얼굴인식 프로그램을 학습하여, 학습활동 과정에서 교실 또는 학교의 동료학생 얼굴 이미지를 촬영하거나 훈련데이터로 사용하는 경우 윤리적 논란이 발생할 수 있음(중국 항저우의 학교 사례)

검토 근거

교육기본법

제23조(교육의 정보화) ① 학교생활기록 등의 학생정보는 교육적 목적으로 수집·처리·이용 및 관리되어야 한다.

제23조의3(학생정보의 보호원칙) ② 제1항의 규정에 의한 학생정보는 법률이 정하는 경우를 제외하고는 당해 학생(학생이 미성년자인 경우에는 학생 및 학생의 부모 등 보호자)의 동의 없이 제3자에게 제공되어서는 아니된다.

원격교육법

제20조(개인정보 등의 보호) ① (생략) 원격교육 과정에서 조사하거나 제공받은 개인 또는 법인·단체의 정보는 이 법과 관련 법률에 근거하지 아니하고 처리되어서는 아니 된다.

② 교육부장관 및 교육감은 원격교육의 운영 과정에서 교원 및 학생의 개인정보를 보호하기 위하여 필요한 정책을 마련하여야 한다.

법정부 윤리기준

2. 10대 핵심요건 ① 인권보장 ② 프라이버시 보호 ③ 다양성 존중 ④ 침해금지 ⑤ 공공성 ⑥ 연대성 ⑦ 데이터 관리 ⑧ 책임성 ⑨ 안전성 ⑩ 투명성

- 인공지능을 개발하고 활용하는 전 과정에서 개인의 프라이버시를 보호해야 한다.
- 인공지능 전 생애주기에 걸쳐 개인 정보의 오용을 최소화하도록 노력해야 한다.
- 개인정보 등 각각의 데이터를 그 목적에 부합하도록 활용하고, 목적 외 용도로 활용하지 않아야 한다.
- 데이터 수집과 활용의 전 과정에서 데이터 편향성이 최소화되도록 데이터 품질과 위험을 관리해야 한다.

【 정책적 과제 】

- ❖ **(AI윤리이슈 발굴)** 교육분야에 인공지능기술이 도입·활용될 경우 발생할 수 있는 윤리이슈를 선제적으로 발굴하고 대응 방안을 마련한다.
- ❖ **(AI윤리교육 강화)** AI윤리 커리큘럼을 개발하고, 교육당사자에게 AI윤리교육 프로그램을 제공한다. 특히, 경제적·사회적 취약계층 학생 및 장애 학생 등의 AI윤리교육에 대한 접근을 보장한다.
- ❖ **(AI윤리 적용 지원)** 인공지능기술의 윤리적 활용에 필요한 교수자 역량 제고를 위해 교수자 연수 기회를 제공하며, 현장의 실제적인 문제해결 지원 서비스 등을 강화한다.
- ❖ **(학술연구 지원)** 전통적 형태의 교육가치를 인정하면서, 교육분야에 인공지능기술이 윤리적이고 책임감 있게 활용될 수 있도록 관련 학술연구를 지원한다.
- ❖ **(소통체계 구축)** 교육분야 인공지능의 윤리적 활용에 대한 모범사례 홍보와 관련 정부부처·기관·학교 등과의 협력을 위한 소통체계를 구축한다.

【 기술적 과제 】

- ❖ **(안전성 판단도구 개발)** 인공지능기술이 교육분야에 적절하고 안전하게 도입될 수 있도록 도입주체별 현장 가이드라인과 판단도구 개발을 지원한다.
- ❖ **(현장수요맞춤형 AI개발 지원)** 산업계의 인공지능기술을 교육분야에 선택·도입하는 수준을 넘어, 교육계 수요에 최적화된 인공지능기술을 발굴·개발할 수 있도록 지원한다.
- ❖ **(교육분야 데이터 관리 표준 개발)** 교육AI 윤리원칙을 충족하는 인공지능의 개발·관리에 필요한 빅데이터 플랫폼, 테스트베드 구축, 데이터 표준 규격 개발 등 인공지능기술의 교육적 활용을 위한 기술환경 구축을 지원한다.
- ❖ **(교육분야 데이터 윤리를 위한 보안시스템 개발)** 산업계와 교육계는 인공지능기술 적용으로 발생할 수 있는 개인정보유출, 오남용 등 데이터 윤리문제를 방지하고 관리할 수 있는 보안시스템의 개발과 유지보수를 지원한다.

참고1

「교육분야 인공지능 윤리원칙」 구성 체계 및 근거

분류	원칙	근거법률
대원칙	사람의 성장을 지원하는 인공지능	교육기본법 제2조, 제12조
기본원칙	학습자의 주도성 강화	교육기본법 제12조
	교수자의 전문성 존중	헌법 제31조 교육기본법 제14조 UNESCO 윤리권고
	기술의 합목적성 제고	원격교육법 제17조 범정부 윤리기준
세부원칙	인간성장의 잠재가능성을 이끌어낸다	교육기본법 제9조, 제12조, 제13조, 제14조
	모든 학습자의 주도성과 다양성을 보장한다.	교육기본법 제3조, 제12조, 제14조 UNESCO 윤리권고 범정부 윤리기준
	교육당사자 간의 관계를 공고히 유지한다.	UNESCO 윤리권고
	교육의 기회균등 실현을 통해 공정성을 보장한다.	헌법 제31조 교육기본법 제4조 범정부 윤리기준
	교육공동체의 연대와 협력을 강화한다.	교육기본법 제23조 UNESCO 윤리권고 범정부 윤리기준
	사회 공공성 증진에 기여한다.	교육기본법 제2조 범정부 윤리기준
	모든 교육당사자의 안전을 보장한다.	교육기본법 제12조, 제13조 EU의 개인정보보호법(GDPR) 제5조 범정부 윤리기준
	데이터 처리의 투명성을 보장하고 설명 가능해야 한다.	교육기본법 제23조 EU 신뢰할 수 있는 AI 윤리원칙 범정부 윤리기준
	데이터를 합목적적으로 활용하고 교육당사자의 프라이버시를 보호한다.	교육기본법 제23조, 제23조의3 원격교육법 제20조 범정부 윤리기준

제 목	주 체	개 요	주요 원칙	주요 특징	
1	Preparing for the Future of Artificial Intelligence ('16)	President National Science and Technology Council Committee on Technology (정부기관)	美 정부의 입장에서 AI 기술과 관련하여 나아갈 방향 제시한 정부 보고서	공공선, 공정성, 안전, 투명성, 이해가능성, 선을 위한 AI(AI for good), 인간 가치(Human values)	<ul style="list-style-type: none"> AI를 주요 성장동력으로 보고 美 정부의 역할 강조 윤리원칙 제시보다는 국제인도법에 근거한 AI 무기체계 개발 등 다양한 AI 관련 이슈를 제시하는데 초점
2	Tenets ('16)	Partnership on AI (민간 연구소)	학계, 재계, 정책입안자 등 다양한 주체들의 협력 도모	AI 혜택 최대화, 다양한 주체들 간 협력, 사생활보호, 견고함, 해악금지, 설명가능성	<ul style="list-style-type: none"> 학계 기업 정책입안자 등 다양한 주체들 간 협력을 강조하고 이를 통해 대중 교육 등 추진할 것을 제안 기술 혜택 최대화의 전제로 사생활 보호, 연구공동체의 책임, 견고성, 해악금지 등 제시
3	AI Policy Principles ('17)	Information Technology Industry Council (민간 협회)	개발자에 대한 정부 차원의 지원 및 공적 영역과 사적 영역의 협업 강조	안전과 제어가능성 해석가능성, 인간 존엄성, 데이터의 대표성, 유연한 규제 접근 , 기회의 평등	<ul style="list-style-type: none"> 개발자의 입장을 강조, 정부의 규제나 개발자에 대한 정보공개 요구에 부정적 다만 개발자에게도 안전한 설계 데이터 대표성 등 높은 수준의 책임성 요구
4	DeepMind Ethics & Society Principles ('17)	DeepMind (기업)	사내에서 AI 연구 수행시 윤리적 고려사항 제시	사생활 침해 금지, 평등, 도덕성, 포용성, 안전과 책무성, 거버넌스규제	<ul style="list-style-type: none"> 연구자에게 필요한 윤리원칙과 체크리스트를 제시하면서도 안전과 책무성을 보장하는 거버넌스규제 필요성 제기
5	Asilomar AI Principles ('17)	Future of Life Institute (민간 연구소)	미국 보스턴의 비영리 연구단체인 삶의 미래 연구소 (Future of Life Institute) 주관으로 작성한 윤리원칙	인권보장, 개인정보보호, 해악금지, 공공성, 데이터 관리, 책임성, 통제성, 투명성, 무기경쟁 회피	<ul style="list-style-type: none"> 스티븐 호킹, 일론 머스크 등 다수의 AI학자, 마인프론트 및 신현 관료들이 사명 AI 기술 연구자, 정책 입안자, 관련 산업 종사자에게 필요한 윤리원칙 제시
6	AI at Google: Our Principles ('18)	Google (기업)	구글 AI 개발자에게 필요한 윤리원칙 제시	사회적 혜택 증진, 불공정한 편견, 지양, 설명가능, 사생활침해 방지	<ul style="list-style-type: none"> 개발을 제한해야되는 AI 어플리케이션으로 해를 끼치는 기술, 인명을 해하는 무기관련 기술, 국제 규약 위반 감시기술 등 제시
7	Microsoft AI principles ('18)	Microsoft's AETHER(AI and Ethics in Engineering and Research) (기업)	MS AI 개발자에게 필요한 윤리원칙 제시	공정성, 신뢰성 및 안전, 사생활 및 보안, 포용성, 투명성, 책무성	<ul style="list-style-type: none"> MS 사내 윤리강령 성격이 강하며, 책임질 수 있는 AI와 이를 위한 교육 강조
8	OpenAI Charter ('18)	OpenAI (민간 연구소)	AI 기술 연구자에게 필요한 윤리적 태도와 원칙 제시	공공선, 해악금지, 안전 담보, AI개발 선두주자, 타 연구단체 협력,	<ul style="list-style-type: none"> 연구자의 자유로운 연구 증진에 초점 고도로 자율적인 AGI(artificial general intelligence) 상정
9	Principles for Trust and Transparency ('18)	IBM (기업)	IBM 직원들을 대상으로 AI 연구를 위해 제시된 윤리원칙	인간 지능 증강(augment) , 데이터 소유권, 국경간 데이터 이동, 투명성	<ul style="list-style-type: none"> AI는 인간을 대체하는 것이 아니라 증강(augment)하기 위한 것임을 명시 AI 사용 여부시기, 학습 데이터 출처 고지 등 규정
10	The Montreal Declaration for a Responsible Development of AI ('18)	University of Montreal (민간 대학)	몬트리올 대학에서 개발된 사회적으로 책임 있는 AI 연구를 위한 윤리원칙	복지(well-being), 자율성 존중, 사생활 보호와 친밀성, 연대성, 민주적 참여, 공평, 다양성 포용, 사려	<ul style="list-style-type: none"> 친밀성(intimacy), 사려(prudence), 지속가능한 발전 등 다른 가이드라인에 잘 등장하지 않는 원칙 제시 윤리원칙 제시와 함께 서명으로 선언에 동참하도록 장려
11	지능정보사회 윤리가이드라인 ('18)	정보문화포럼 (정부기관)	인간 중심의 지능정보사회 구현	이용자 주도성, 이용자/시민참여 , 공익, 공정성, 위험예방, 프라이버시 보호	<ul style="list-style-type: none"> 지능정보기술 관련 개발자 및 공급자의 윤리 의식 고취 및 이용자의 오남용 방지 지침 주체별(개발자, 공급자, 이용자) 세부지침 마련
12	AI in the UK: Ready, Willing and Able? ('18)	영국 정부 (정부기관)	영국 정부 차원에서 정책적으로 접근할 수 있는 제언 제시	데이터 접근과 제어, 이해가능한 AI, 디지털 이해력 증진, 공중보건 관리	<ul style="list-style-type: none"> 영국이라는 특정 국가 입장에서 공중 보건 데이터 관리, AI 디지털 이해력 제고 등 구체적으로 취할 수 있는 AI 관련 정책을 제시

제 목	주 체	개 요	주 요 원 칙	주 요 특 징
13	카카오 (기업)	카카오 내 AI 관련 연구시 지향되어야 할 윤리원칙 제언	사회윤리 준수, 차별 경계 학습데이터 운영 알고리즘 독립성 및 설명 기술 포용성 아동·청소년 보호	<ul style="list-style-type: none"> 국내기업 최초 AI 윤리현장으로, 알고리즘과 데이터에 대한 관리, 아동과 청소년에 대한 보호 필요성 등 강조
14	Microsoft's AETHER (기업)	AI가 가져올 미래의 변화에 대응하기 위해 MS의 Aether 연구소에서 책자 제작	시에 의한 진보 공정성, 신뢰성 및 안전, 사생활보안, 포용성, 투명성	<ul style="list-style-type: none"> AI가 경제사회적 진보를 이끌고 지역적·전지구적 문제를 해결할 것이라는 관점 AI가 직업과 직장에 미치는 영향에 공공부문과 민간부문이 협력해 대응할 필요성 제시
15	AI Now (민간 연구소)	작업환경에서 다양성을 확보하기 위해 고려할 사항 제시	다양성, 해악금지, 개방성, 투명성	<ul style="list-style-type: none"> 급여지급 기준의 인종별, 성별 공개, 직원 채용시 투명성 준수 등 제시 특히 AI 시스템 사용시 투명성·편견·해악에 대한 철저한 점검·감시·추적·공개를 강조
16	The IEEE Global Initiative on Ethics of Autonomous and Intelligent System (민간 학회)	IEEE에서 Ethics in Action 캠페인과 함께 아울러 공개된 보고서	인권 복지우선, 책무성, 투명성, 오용의 인식	<ul style="list-style-type: none"> 각 원칙별로 이론적 배경, 참고 자료를 제시하고 윤리원칙뿐만 아니라 관련 분야들에 대한 자료 수록
17	방통위-KISDI (정부기관)	안전한 지능정보서비스 환경조성 및 이용자의 권리와 자유에 근거한 윤리원칙 제시	사람중심 서비스, 투명성과 설명가능성, 책임성, 안전성, 차별금지, 참여, 프라이버시와 데이터거버넌스	<ul style="list-style-type: none"> 안전한 지능정보서비스 환경조성 및 이용자 보호를 위해 모든 주체 사이의 협력 강조 기업과 연구자들의 의견을 폭넓게 수렴하여 작성 이용자 보호의 관점 강조
18	산업통상자원부 (정부기관)	2007년에 만들어진 로봇윤리현장을 수정 보완	인간의 존엄성 보호, 공공선, 행복추구, 투명성, 제이가능성, 책무성, 안전성, 정보보호	<ul style="list-style-type: none"> 로봇산업계에 종사하는 연구원, 개발자, 및 사용자가 로봇과 AI를 설계·제작·공급·사용·관리하는 데 기준으로 삼는 가이드라인 제시
19	일본 총무성 (정부기관)	25명의 산학연 전문가로 구성된 '인간 중심의 AI 사회 원칙 위원회'를 통해 제안	인간중심, 교육교양, 개인정보 보호, 보안, 공정경쟁, 공정성, 책임성, 투명성, 혁신	<ul style="list-style-type: none"> 저출산, 고령화, 지방소멸, 재해 재난 등 일본이 처한 어려움을 AI가 해결할 수 있을 것으로 상정 AI를 공공재로 활용하여 사회의 근본적인 변화와 혁신을 달성하여 지속 가능한 발전 추구
20	EU (국가 정부기관)	EU 산하의 50여명으로 구성된 AI 전문가 그룹 주도	인간 권리자율성 보장, 기술적 견실성, 사생활, 데이터 관리, 투명성, 다양성, 차별금지, 복지, 책무성	<ul style="list-style-type: none"> 법국가 차원의 협업을 통해 신뢰할 수 있는 AI를 위한 윤리원칙 정립에 초점을 맞춤 각 원칙의 평가 리스트를 구체적으로 제시
21	OECD (범국가 정부기관)	OECD 디지털 경제 정책 위원회 주관하에 제작	포용적 성장, 지속가능 발전, 인간중심 가치, 공정성, 투명성, 설명가능성, 견고성, 보안 및 안전, 책무성	<ul style="list-style-type: none"> 윤리원칙 뿐 아니라 정책 입안자들에게 대한 제언 제시, 국가별 정책 수립과 국제적 협력 도모
22	Jobin, A., Ienca, M. & Vayena, E. (개인)	전 세계의 주요 84개의 AI 가이드라인을 분석	투명성, 정의, 해악금지, 책임, 사생활보호, 혜택, 추구, 자유, 신뢰, 지속가능성, 연대성	<ul style="list-style-type: none"> 주요 윤리원칙을 빈도수별로 분석하고 주로 선진국을 중심으로 발표되고 있음을 밝힘
23	The Alan Turing Institute (국영 연구소)	영국의 국영 연구소인 Alan Turing 연구소에서 제작	존중, 연결 보호, 돌봄, 공정성, 책임성, 지속가능성, 투명성	<ul style="list-style-type: none"> AI 기술이 데이터를 처리할 때 발생할 수 있는 위험이나 문제점을 예방하는 데 필요한 윤리원칙에 초점
24	UNESCO	SDG를 달성하기 위해 신기술을 효율적으로 활용하겠다고 약속한 '칭다오 선언'을 시대에 맞게 결의 변경	교육정책, 교육관리시스템, 교수과정, 교사지원, 학습·학습평가, 삶과 일을 위한 기회, 기술·발명, 평생학습기회에 활용, 공평하고 평등한 AI사용, 데이터·알고리즘에서 윤리, 투명성 등 확보 등	<ul style="list-style-type: none"> 2030 교육 의제를 달성하기 위해 AI 기술을 활용하는 방법에 대한 지침 및 권장 사항

제 목	주 체	개 요	주 요 원 칙	주 요 특 징	
25	Principles Artificial Intelligence: A Map of Ethical and Rights-Based Approaches (20)	Berkman Klein Center For Internet & Society (민간 연구소)	36개의 윤리 가이드라인에 등장한 윤리원칙들을 주제별로 분석	사생활보호, 책무성, 안전과 보안, 투명성과 설명가능성, 공정성과 차별금지, 인간의 기술 통제, 전문적 책임	<ul style="list-style-type: none"> 정부, 정부 기관, 사적 기관 등 다양한 주체들이 제시한 윤리원칙들을 8개의 주제로 분류하고 분석
26	Rome Call for AI Ethics (20)	로마 교황청 (민간기관)	로마 교황청에서 인간의 혁신적인 미래를 위한 AI 윤리원칙 제정	투명성, 포용성, 책임성, 불편부당성, 신뢰성, 보안과 사생활 보호	<ul style="list-style-type: none"> 종교 기관인 가톨릭교회에서 제정한 윤리 원칙으로, 인간 가족(human family)에 대한 봉사, 젊은 세대에 대한 준비, 자연의 회복 필요성 등 제시
27	자율주행차 윤리 가이드라인(20)	국토교통부	자율주행자동차의 도입으로 발생할 수 있는 부정적 효과 감소, 책임 소재를 가릴 수 있는 방향 제시, 행위주체들의 행동원칙 준수 권고	(기본가치) 인간은 존엄하기 때문에 인간의 명령에 복종, 인간의 행복 추구, 인간의 생명 보호 및 공공선 추구하고, 지속가능한 발전 (행동원칙) 투명성, 제어 가능성, 책임성, 안전성, 보안성	<ul style="list-style-type: none"> 자율주행자동차의 기본가치와 행동 원칙 제시 행위 주체인 설계자, 제작자, 관리자, 서비스 제공자, 이용자가 적용 대상
27	사람이 중심이 되는 인공지능(AI) 윤리기준(20)	범정부 합동	국내외 주요 인공지능 윤리원칙 분석-윤리 철학의 이론적 논의와 연계	(3대 원칙) 인간의 존엄성 원칙, 사회의 공공선 원칙, 기술의 합목적성 원칙 (10대 핵심요건) 인권보장, 프라이버시보호, 다양성 존중, 침해금지, 공정성, 연대성, 데이터 관리, 책임성, 안전성, 투명성	<ul style="list-style-type: none"> 인공지능 개발에서 활용에 이르는 전 단계에서 모든 사회 구성원이 참조하는 기준 특정 분야에 제한되지 않는 범용성을 가진 일반 원칙 구속력 있는 '법'이나 '지침'이 아닌 도덕적 규범이자 자율 규범
29	인공지능(AI) 개인정보보호 자율점검표(21)	개인정보 보호위원회	AI기술 서비스 환경에 적합한 자율적인 개인정보보호 이행 점검에 필요한 사항 안내 ※ 「개인정보 보호법」상 주요 준수사항 포함	적법성, 안전성, 투명성, 참여성, 책임성, 공정성	<ul style="list-style-type: none"> AI 개발·운영에 참여하는 자의 개인 정보 보호에 대한 인식 제고와 개인 정보보호를 자율적으로 실천하는데 필요한 자율 점검표 마련 AI서비스 이외 분야 개인정보처리자도 자율적으로 활용 가능 법령상 준수해야 할 의무권장 내용에 대한 점검항목(16개), 확인사항(54개)으로 구성
30	국가 인공지능 사업추진 윤리원칙(21)	한국지능정보사회진흥원(NIA)	지능정보기술로 국가사회의 혁신성장과 디지털 기반 ESG 가치 실현에 기여	(기획-사람과 책임) 법제도 마련, 장기적 비전과 실행계획 제시, 다양한 의견수렴 (실행-안전과 평등) 국민의 편익과 행복 우선 추구, 다양성 보장 및 편향 방지, 제도와 윤리 규범 안에서 개발·활용 등등 (확산-공익과 민주사회) 인공지능 교육 및 국제사회 협력 추진, 인공지능 영향력을 고려한 사업개선, 사업성과와 혜택의 고른 분배 등	<ul style="list-style-type: none"> 국가의 인공지능 사업을 추진하는 과정에서 지켜야 할 최소한의 규범 사업추진 단계별(기획, 실행, 확산) 3가지의 핵심가치와 9가지의 윤리원칙 제시
31	The Recommendation on the Ethics of Artificial Intelligence (유네스코 인공지능 윤리 권고(21.11.))	UNESCO	인공지능 기술의 개발과 사용에 있어 세계 공동으로 적용할 수 있는 권고안 마련	(교육과 연구) 관련 충분한 AI 리더십 교육 제공, 교수·교사 연수 및 이력에서 AI 기술의 윤리적 책임 있는 사용에 대한 연계·지속 AI 교육을 충실히 누리지 못하는 모든 이의 참여와 리더십 장려, AI 윤리 커리큘럼 개발 협력, 민간기업 독려 등	<ul style="list-style-type: none"> 회원국들에 법적 절차를 포함하여 인공지능 윤리 권고 적용 제안 유네스코의 중심영역(교육, 과학, 문화, 정보통신)과 관련된 AI 시스템의 광범위적 윤리적 영향 강조

* 출처 : 사람이 중심이 되는 인공지능 윤리기준(과기정통부 등 범정부, '20.12.) 등 참고

I. 서문

… (전략) 본 윤리기준은 범용성이 있는 일반 원칙으로서 **사안별 또는 분야별 인공지능 윤리기준 제정의 근거를 제공하여** … (중략) … 본 윤리기준에서 제시하는 원칙과 요건들은 **상황에 따라 상충관계가 발생할 수 있으며**, (후략) …

II. 인공지능 윤리기준: 3대 기본원칙, 10대 핵심요건

1. 3대 기본원칙 - 인공지능 개발 및 활용 과정에서 고려될 원칙

○ ‘인간성을 위한 인공지능(AI for Humanity)’을 위해 인공지능 개발에서 활용에 이르는 전 과정에서 고려되어야 할 기준으로 **3대 기본원칙**을 제시한다.

① 인간 존엄성 원칙

- 인간은 신체와 이성이 있는 생명체로 인공지능을 포함하여 인간을 위해 개발된 기계 제품과는 교환 불가능한 가치가 있다.
- 인공지능은 인간의 생명은 물론 정신적 및 신체적 건강에 해가 되지 않는 범위에서 개발 및 활용해야 한다.
- 인공지능 개발 및 활용은 안전성과 견고성을 갖추어 인간에게 해가 되지 않도록 해야 한다.

② 사회의 공공선 원칙

- 공동체로서 사회는 가능한 한 많은 사람의 안녕과 행복이라는 가치를 추구한다.
- 인공지능은 지능정보사회에서 소외되기 쉬운 사회적 약자와 취약 계층의 접근성을 보장하도록 개발 및 활용되어야 한다.
- 공익 증진을 위한 인공지능 개발 및 활용은 사회적, 국가적, 나아가 글로벌 관점에서 인류의 보편적 복지를 향상시킬 수 있어야 한다.

③ 기술의 합목적성 원칙

- 인류의 삶에 필요한 도구인 인공지능 기술은 사용자의 목적과 의도에 따른 고유한 목적과 수단적 가치를 지닌다.
- 인공지능은 궁극적으로 인간에게 도움이 되어야 한다는 목적에 맞도록 개발 및 활용되어야 한다.
- 인류의 삶과 번영을 위한 인공지능 개발 및 활용을 장려하여 진흥해야 한다.

2. 10대 핵심요건 - 기본원칙을 실현할 수 있는 세부 요건

- 3대 기본원칙을 실천하고 이행할 수 있도록 인공지능 전체 생명 주기에 걸쳐 충족되어야 하는 10가지 핵심 요건을 제시한다.

① 인권보장

- 인공지능의 개발과 활용은 모든 인간에게 동등하게 부여된 권리를 존중하고, 다양한 민주적 가치와 국제 인권법 등에 명시된 권리를 보장하여야 한다.
- 인공지능의 개발과 활용은 인간의 권리와 자유를 침해해서는 안 된다.

② 프라이버시 보호

- 인공지능을 개발하고 활용하는 전 과정에서 개인의 프라이버시를 보호해야 한다.
- 인공지능 전 생애주기에 걸쳐 개인 정보의 오용을 최소화하도록 노력해야 한다.

③ 다양성 존중

- 인공지능 개발 및 활용 전 단계에서 사용자의 다양성과 대표성을 반영해야 하며, 성별·연령·장애·인종·종교·국가 등 개인 특성에 따른 편향과 차별을 최소화하고, 상용화된 인공지능은 모든 사람에게 공정하게 적용되어야 한다.
- 사회적 약자 및 취약 계층의 인공지능 기술 및 서비스에 대한 접근성을 보장하고, 인공지능이 주는 혜택은 특정 집단이 아닌 모든 사람에게 골고루 분배되도록 노력해야 한다.

④ 침해금지

- 인공지능을 인간에게 직·간접적인 해를 입히는 목적으로 활용해서는 안 된다.
- 인공지능이 야기할 수 있는 위험과 부정적 결과에 대응 방안을 마련하도록 노력해야 한다.

⑤ 공공성

- 인공지능은 개인적 행복 추구 뿐만 아니라 사회적 공공성 증진과 인류의 공동 이익을 위해 활용해야 한다.
- 인공지능은 긍정적 사회변화를 이끄는 방향으로 활용되어야 한다.
- 인공지능의 순기능을 극대화하고 역기능을 최소화하기 위한 교육을 다방면으로 시행하여야 한다.

⑥ 연대성

- 다양한 집단 간의 관계 연대성을 유지하고, 미래세대를 충분히 배려하여 인공지능을 활용해야 한다.
- 인공지능 전 주기에 걸쳐 다양한 주체들의 공정한 참여 기회를 보장하여야 한다.
- 윤리적 인공지능의 개발 및 활용에 국제사회가 협력하도록 노력해야 한다.

⑦ 데이터 관리

- 개인정보 등 각각의 데이터를 그 목적에 부합하도록 활용하고, 목적 외 용도로 활용하지 않아야 한다.
- 데이터 수집과 활용의 전 과정에서 데이터 편향성이 최소화되도록 데이터 품질과 위험을 관리해야 한다.

⑧ 책임성

- 인공지능 개발 및 활용과정에서 책임주체를 설정함으로써 발생할 수 있는 피해를 최소화하도록 노력해야 한다.
- 인공지능 설계 및 개발자, 서비스 제공자, 사용자 간의 책임소재를 명확히 해야 한다.

⑨ 안전성

- 인공지능 개발 및 활용 전 과정에 걸쳐 잠재적 위험을 방지하고 안전을 보장할 수 있도록 노력해야 한다.
- 인공지능 활용 과정에서 명백한 오류 또는 침해가 발생할 때 사용자가 그 작동을 제어할 수 있는 기능을 갖추도록 노력해야 한다.

⑩ 투명성

- 사회적 신뢰 형성을 위해 인공지능의 투명성과 설명 가능성을 높이고, 타 원칙과의 상충관계를 고려하여 활용 상황에 적합한 수준의 투명성과 설명 가능성을 높이려는 노력을 기울여야 한다.
- 인공지능기반 제품이나 서비스를 제공할 때 인공지능의 활용 내용과 활용 과정에서 발생할 수 있는 위험 등의 유의사항을 사전에 고지해야 한다.

III. 부록

1. 본 윤리기준에서 인공지능의 지위

- 본 윤리기준에서 지향점으로 제시한 ‘인간성을 위한 인공지능(AI for Humanity)’은 인공지능이 인간을 위한 수단임을 명시적으로 표현하지만, 인간 종 중심주의(human species-centrism) 또는 인간 이기주의를 표방하지는 않는다.
- 본 윤리기준에서 인공지능은 지각력이 있고 스스로를 인식하며 실제로 사고하고 행동할 수 있는 수준의 인공지능(이른바 강인공지능)을 전제하지 않으며 하나의 독립된 인격으로서의 인공지능을 의미하지도 않는다.

2. 적용 범위와 대상

- 본 윤리기준은 인공지능 기술의 개발부터 활용에 이르는 전 단계에 참여하는 모든 사회구성원을 대상으로 하며, 이는 정부·공공기관, 기업, 이용자 등을 포함한다.

3. 인공지능 윤리기준의 실현방안

- ‘인공지능 윤리기준’을 기본 플랫폼으로 하여 다양한 이해관계자 참여하에 인공지능 윤리 쟁점을 논의하고, 지속적 토론과 숙의 과정을 거쳐 주체별 체크리스트 개발 등 인공지능 윤리의 실천 방안을 마련한다.

참고4

[요약] 유네스코 인공지능 윤리 권고 (21.11, 유네스코)
(The Recommendation on the Ethics of Artificial Intelligence)

※ 원문은 유네스코(<https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14>) 참고

□ 논의 경과

- 제40차 총회에서 「인공지능 윤리 권고문」을 수립하기로 결정 (19.11월)
- 권고문 초안의 정교화 작업을 위한 특별 전문가집단(Ad Hoc Expert Group, 총 24명) 구성 (20.3월)
- 권고문 초안에 대한 이해관계자 협의회 (20.6~8월)
- 정부간 전문가 회의를 통한 시안 수립 (21.1월), 정부간 회의를 통한 시안 의견수렴 (21.4월, 6월), 회원국 회람 (21.8월)
- 제212차 집행이사회 검토 (21.10월) 및 제41차 총회 의제 채택 (21.11.25.)

□ 주요 내용

① 서문 (Preamble)

- 회원국들은 적절한 법적 절차를 포함하여 각 국가의 정부 구조를 바탕으로 인공지능 윤리 권고 적용을 제안함
- 회원국들은 과학적인 연구와 윤리 분석 및 평가에 의한 인공지능 기술의 개발과 사용에 있어 전 분야에서 권고문이 활용될 수 있도록 독려하길 추천함

② 적용 범위 (Scope of Application)

- 유네스코의 중심 영역(교육, 과학, 문화, 정보통신)과 관련된 AI 시스템의 광범위적인 윤리적 영향에 특별히 관심을 기울여야 함
- (교육) 디지털 사회에서의 삶은 새로운 교육적 실행, 윤리적 성찰, 비판적 사고, 책임 있는 설계 관행, 새로운 숙련을 노동 시장과 고용 가능성, 시민참여에 영향을 주는 책임 있는 디자인 실천과 새로운

기술을 요구함

- (과학) 의학, 과학, 인문학, 사회학 등 전 영역에 AI 기술은 새로운 연구 능력과 접근을 가져올 것이므로 과학적 이해와 설명에 영향을 줄 것이고, 새로운 결정의 기준을 만들 것임
- (문화) AI 기술은 문화·창조 산업을 풍요롭게 할 수도 있지만 문화 콘텐츠·데이터·시장·수익의 공급이 소수의 행위 주체에게 편중되어 언어·미디어·문화적 표현·참여·평등의 다양성 및 다원성에 부정적 영향을 줄 잠재력을 가짐
- (정보통신) AI 기술은 정보의 구조와 생산과정에 중요한 역할을 하므로 허위정보, 누락정보, 혐오발언, 차별, 표현의 자유, 개인정보, 미디어 리터러시 관련 이슈를 증가시킬 것임

③ 정책 조치 (Areas of Policy Action) - 8. 교육과 연구 (Education and Research)

101. 인간의 역량을 강화하고 인공지능 시스템의 광범위한 도입에서 파생되는 디지털 격차 및 디지털 접근불평등을 줄이기 위해서, 모든 국가의 모든 수준의 대중에게 적절한 인공지능 리터러시 교육을 제공할 수 있도록 국제기구, 교육 기관, 민간 단체, 비정부단체와 협력해야 함
102. 회원국은 인공지능 교육을 위한 '기초소양*'의 습득을 장려하며, 특히 이러한 기초소양 교육의 격차가 두드러지는 국가 및 이러한 국가 내 지역·구역에서 장려해야 함
 - * 기본언어능력, 산술능력, 코딩·디지털 능력, 미디어·정보 리터러시, 비판적 사고, 팀워크, 사회정서적 능력, 인공지능 윤리능력
103. 회원국은 데이터, 인공지능 기술이 초래하는 기회 및 어려움, 유아의 권리를 비롯한 인권에 대한 인공지능 시스템의 영향 및 함의를 비롯하여 인공지능 개발에 대한 보편 인식 프로그램을 장려해야 함
 - 이러한 프로그램은 기술전문가집단뿐만 아니라 비전문가집단에게도 이용이 쉬워야 함

104. 회원국은 다른 사안보다도 교육, 교사 연수, 온라인 교육에서 인공지능의 윤리적이고 책임 있는 사용에 대한 연구 계획을 장려하여 이 분야와 관련된 기회는 향상시키고 어려움 및 위험성은 완화해야 함
- 이 계획에는 교육의 질에 대한 적절한 평가, 인공지능 기술의 사용이 학생 및 교사에게 미치는 영향에 대한 적절한 평가가 수반되어야 함
 - 전통적 교육 형태의 관계·사회적 측면 및 가치가 학생-교사 및 학생-학생 관계에서 매우 중요하며 이것이 교육으로의 인공지능 기술의 도입을 논할 때 고려되어야 한다는 점을 유념한 채로 인공지능 기술이 학생 및 교사의 역량을 강화하고 경험을 증대시킬 수 있도록 보장해야 함
 - 모니터링, 능력 평가, 학습자의 행동 예측에 관한 것이라며, 학습에 사용되는 인공지능 시스템은 엄격한 요구조건 하에 있어야 함
 - 인공지능은 관련 개인 데이터 보호 기준에 따라 인지능력의 약화 및 민감정보의 유출 없이 학습과정을 보조해야 함
 - 학습자와 인공지능 간의 상호작용 동안 축적된 지식을 습득하기 위해 양도된 데이터는 상업적 사용을 비롯하여 오용, 유용, 또는 범죄적으로 사용되지 않아야 함
105. 모든 수준의 인공지능교육 프로그램에서 소녀 및 여성, 다양한 민족, 장애인 등 혜택을 향유하지 못하는 모든 사람의 참여 및 주도를 장려하고, 이에 대한 다른 모범 사례 모니터링·공유 증진
106. 자국 전통에 따라 전 학년에 AI윤리교육과정 개발하고 기술 숙련 교육과 AI교육의 인문학·윤리·사회적 측면 간의 교차협력 증진
- AI윤리교육의 온라인 과정 및 디지털 자료는 언어 등 환경의 다양성 고려하고, 특히 장애인이 접근 가능한 형태로 개발되어야 함
107. 회원국들은 연구 투자 혹은 공공 및 민간 분야에 투자 인센티브를 제공하는 등의 방법을 통해, AI 윤리 연구를 비롯한 AI 연구를 장려하고 지원해야 함

- 회원국들은 국제법과 본 권고안에 명시된 가치 및 원칙을 따르는 연구가 AI 기술의 개발과 개선에 크게 기여할 수 있음을 인지해야 함
 - 회원국들은 윤리적 AI 개발 모범사례를 공표하고, 윤리적 방법으로 AI를 개발하는 연구자 및 기업들과 협업해야 함
108. 회원국들은 AI 연구자들이 연구 윤리 훈련을 수료하도록 하고, AI 설계, 상품, 발표 단계, 특히 사용된 데이터베이스의 분석, 주석이 달리는 방법, 적용 가능한 결과의 질과 범주 등과 같은 부분에서 윤리적 사항을 고려하도록 해야 함
109. 회원국들은 개발도상국, 소도국(小島國) 등에서 과학 공동체가 연구를 위해 필요한 데이터에 접근할 수 있도록 민간기업을 독려해야 함. 이는 개인정보 및 데이터 보호 기준에 부합하는 방식으로 시행되어야 함
110. AI 연구에 대한 비판적 평가와 잠재적 오용 및 부작용에 대한 모니터링을 보장하기 위해, 회원국들은 향후 AI와 관련된 기술의 개발이 독립적이고 철저한 과학 연구에 기초하도록 해야 함
- 또한 문화연구학, 교육학, 윤리학, 국제관계학, 법학, 어학, 철학, 정치학, 사회학, 심리학과 같은 과학·기술·엔지니어링·수학(STEM) 이외의 영역이 AI 학제 간 연구에 포함되도록 해야 함
111. 회원국들은 AI 기술이 특히 전통 모델 중심적인 분야에서, 과학적 지식과 관행의 발전에 좋은 기회를 제공한다는 점을 인정하는 한편, 과학 공동체가 AI 사용의 혜택, 한계 및 위험을 인식하도록 해야 함
- 회원국들은 과학 공동체가 정책 수립 과정에 참여하고 AI 기술의 강점 및 약점에 대한 인식을 공표하는 역할을 수행할 수 있도록 함

※ 원문은 유네스코(<https://unesdoc.unesco.org/ark:/48223/pf0000368303>) 참고

□ 개 요

- 「AI와 교육에 대한 국제 컨퍼런스」에서 합의안 채택(19.5월, 베이징)
 - ※ 105개 유네스코 회원국 대표, 50개 정부의 부처, 학술·민간기관 등 참여
 - '교육 2030 아젠다'달성을 위한 AI기술 활용 지침과 권장사항 포함

□ 주요내용

① 교육정책에서 AI에 대한 계획

- AI가 여러 학문에 영향*을 미칠 수 있음을 주의해야 하며, 교육 정책에서 AI를 실현하는 데 필요한 투자를 염두에 두어야 함
- * 평생교육 관점에서 일관되고 통합된 교육 AI전략 계획 및 개발 필요

② 교육 관리와 실행을 위한 AI

- 증거 기반 정책을 위해 데이터 사용 가능성을 인식하고, 학생, 교사, 학부모, 지역사회에 교육과 훈련을 제공하기 위해 AI를 사용할 수 있음

③ 교수과정과 교사를 지원하는 AI

- AI가 교사의 교육 업무를 대체될 수는 없으며*, AI가 활용되는 교육환경에 맞게 교사의 역할을 검토하고 역량 개발을 지원해야 함
- * 교사와 학생 간 인간적인 소통과 협동은 교육의 핵심 요소가 되어야 함

④ 학습과 학습평가를 위한 AI

- 학습 및 평가를 지원하는 AI의 잠재력과 가능성을 인식하고 학습 방법을 변화시켜야 하며, 학습과정의 조정, 학생 역량의 평가 등을 지원하기 위한 AI를 개발해야 함

⑤ AI시대 삶과 일을 위한 가치, 기술개발

- AI를 수용함으로써 발생할 경제, 노동시장, 사회 변화에 맞는 교과 과정을 업데이트하고 개발하는 구조를 갖춰야 하며,

- 사회의 모든 면에서 인간과 기계의 협력을 위한 AI리터러시를 강화하는 제도적 조치를 취해야 하며
- 지역의 AI 인재를 양성하기 위해 고등교육과 연구기관을 지원하는 중기 및 장기 계획을 세우고 AI 인력을 양성해야 함

⑥ 모든 이를 위한 평생학습기회에 AI 활용

- SDG4를 달성하기 위한 평생학습의 원칙을 확인하고, 개별 맞춤형 학습을 위해 AI플랫폼과 데이터에 기반한 분석을 수용해야 하며,
- 노령층이 경제활동에 참여하도록 기술 개발에 주의를 기울여야 함

⑦ 교육에서 공정하고 포괄적으로 AI 사용

- 교육 내에서, 교육을 통해서, 포용성과 공정성을 실현해야 하며, AI를 교육에 활용함으로써 취약계층*이 교육받을 기회를 늘리고, AI의 개발 및 사용이 디지털 격차를 확대해서는 안 됨

* 성별, 장애, 사회경제적 지위, 민족, 문화적 배경, 지리적 위치 등

⑧ 성별에 상관없이 공정한 AI와 성 평등성을 위한 AI

- 교육분야 AI 개발 시, 성적 편견을 없애고 AI 개발에 사용되는 데이터가 성별 영향을 인지하도록 해야 함

⑨ 교육 데이터 사용 및 알고리즘에서 윤리, 투명성, 감사(勘査) 확보

- AI가 편견을 강요할 수 있음을 인식하고, 데이터 소유권, 개인정보 보호, 공익 목적 데이터 사용 등과 관련된 법적·윤리적 위험에 유의하고, 교육에서 AI의 책임있는 사용을 위한 규제를 만들어야 함

⑩ 모니터링, 평가, 연구

- 교육에서 AI가 미치는 영향에 대한 연구 및 분석을 지원하고 AI의 교육적 영향력을 평가하는 구조를 개발해야 함

⑪ 국제 조직을 위한 권장사항- 자금조달, 파트너십, 국제적 협력

- 국가 간 AI 개발 격차 해소, AI의 공정한 사용, AI개발 관련 한계점 검토 및 규제 방안 마련 등을 위해 협력해야 함