과학기술정보통신부
Ministry of Science and ICT
www.msit.go.kr

보도자료

4차산업혁명의 큰길로 대한민국이 달려갑니다

한국판뉴딜 디지털뉴딜

보도일시	2020. 12. 23.(수) 11:00 이후 부터 보도해 주시기 바랍니다.		
배포일시	2020. 12. 22.(화) 14:00	담당부서	인공지능기반정책과
담당과장	김경만(044-202-6270)	담 당 자	유주연 사무관(044-202-6276)

과기정통부, 사람이 중심이 되는 「인공지능(AI) 윤리기준」마련

-공개 공청회(12.7) 등 각계 전문가·시민 공개 의견수렴 거쳐 발표-

-'인간성(Humanity)을 위한 인공지능(AI)'의 3대 원칙·10대 요건 담아 -

□ 과학기술정보통신부(장관 최기영, 이하 '과기정통부')와 정보통신정책연구원은 2020년 12월 23일 대통령 직속 4차산업혁명위원회 전체회의에서 **인공지능 시대 바람직한 인공지능 개발·활용 방향을 제시**하기 위한 사람이 중심이 되는 「**인공지능(AI) 윤리기준**」**을 마련**했다.('붙임')

○ 이는 윤리적 인공지능을 실현하기 위해 **정부·공공기관, 기업, 이용자 등 모든 사회구성원이 인공지능 개발~활용 전 단계**에서 함께 지켜야 할 **주요 원칙과 핵심 요건**을 제시하는 기준으로서,

○ 그간 인공지능·윤리학·법학 등 **학계·기업·시민단체**를 아우르는 **주요 전문가들이 자문과 의견수렴 과정에 참여**했으며 **11.27 초안 발표** 이후 **12.7 공개 공청회 등 시민 의견수렴을** 거쳤다.

□ 인공지능 기술의 발전·확산과 함께 **인공지능 기술의 윤리적 개발·활용** 역시 **세계 각국과 주요 국제기구의 관심 대상**이 되어 왔으며,

○ 지난해 우리나라가 주도적으로 참여한 **경제협력개발기구(OECD) 인공지능 권고안('19.5)을 비롯하여 OECD, 유럽연합(EU) 등 세계 각국과 국제기구, 기업, 연구기관 등** 여러 주체로부터 **다양한 인공지능 윤리 원칙이 발표**되었다.

○ 이에 과학기술정보통신부는 이러한 글로벌 추세에 발 맞추어 지난해 발표된 **「인공지능 국가전략('19.12)」 주요 과제로 '인공지능 윤리기준' 마련**을 추진해왔다.

□ 과기정통부는 올해 4월부터 인공지능·윤리 전문가로 구성된 **인공지능 윤리연구반**을 통해 **국내외 주요 인공지능 윤리원칙을 분석**하고, 그 결과를 윤리철학의 이론적 논의와 연계하여 **'인간성을 위한 인공지능(AI for Humanity)'를 목표**로 하는 윤리기준 **초안을 마련**하였으며, 3개월에 걸쳐 학계·기업·시민단체 등 **각계의 다양한 전문가로부터 의견을 수렴**하였다.

□ 이러한 과정을 거쳐 마련된 「인공지능 윤리기준」은 '사람 중심의 인공지능'을 위한 최고 가치인 **'인간성(Humanity)'**을 위한 **3대 기본원칙과 10대 핵심 요건을 제시**하고 있으며, 주요내용은 다음과 같다.

○ (**목표 및 지향점**) ① 모든 사회 구성원이 ② 모든 분야에서 ③ 자율적으로 준수하며 ④ 지속 발전하는 윤리기준을 지향한다.

- ①인공지능 **개발에서 활용에 이르는 전 단계**에서 정부·공공기관, 기업, 이용자 등 **모든 사회 구성원이 참조**하는 기준
- ②**특정 분야에 제한되지 않는 범용성**을 가진 일반원칙으로, 이후 **각 영역별 세부 규범이 유연하게 발전해나갈 수 있는 기반** 조성
- ③구속력 있는 '법'이나 '지침'이 아닌 **도덕적 규범이자 자율규범**으로, **기업 자율성을 존중**하고 인공지능 **기술발전을 장려**하며 **기술과 사회변화에 유연하게 대처**할 수 있는 윤리 담론을 형성
- ④사회경제, 기술 변화에 따라 **새롭게 제기**되는 **인공지능 윤리 이슈**를 **논의하고 구체적으로 발전**시킬 수 있는 **플랫폼**으로 기능

○ **(최고 가치)** 윤리기준이 지향하는 최고가치를 '**인간성(Humanity)**'로 설정하고, **'인간성을 위한 인공지능(AI for Humanity)'**을 위한 3대 원칙·10대 요건 제시

○ **(3대 기본원칙)** ‘인간성(Humanity)’을 구현하기 위해 인공지능의 개발 및 활용 과정에서 **❶ 인간의 존엄성 원칙, ❷ 사회의 공공선 원칙, ❸ 기술의 합목적성 원칙**을 지켜야 한다.

○ **(10대 핵심요건)** 3대 기본원칙을 실천하고 이행할 수 있도록 인공지능 개발~활용 전 과정에서 ① **인권 보장**, ② **프라이버시 보호**, ③ **다양성 존중**, ④ **침해금지**, ⑤ **공공성**, ⑥ **연대성**, ⑦ **데이터 관리**, ⑧ **책임성**, ⑨ **안전성**, ⑩ **투명성**의 요건이 충족되어야 한다.

□ 과기정통부는 그간 전문가·시민 의견수렴 과정에서 제기된 바와 같이 향후 윤리기준의 현장 확산을 돕기 위해 개발자·공급자·이용자 등 **주체별 체크리스트 개발, 인공지능 윤리 교육 프로그램 마련 등 구체적인 실천방안을 마련하고 추진**해나갈 계획이며,

○ 앞으로도 **‘인공지능 윤리기준’을 기본 플랫폼**으로 하여 **다양한 이해관계자 참여**하에 **인공지능 윤리 이슈**를 **지속 논의**하고 윤리기준이 **기술·사회 변화를 반영하여 계속해서 발전**될 수 있도록 노력할 예정이다.

□ 과기정통부 최기영 장관은 “지난 **11월 27일 윤리기준 초안을 발표**한 후 **공청회(12.7) 등 폭넓은 공개 의견수렴**을 거쳐 **「인공지능 윤리기준」**이 마련된 만큼, 동 윤리기준이 인공지능 윤리 이슈에 대한 **우리사회의 토론과 숙의의 시작점**이자 **사람 중심의 인공지능으로 나아가는 플랫폼**이 되도록 노력하겠다.”고 밝혔다.

붙임: 사람이 중심이 되는 「인공지능(AI) 윤리기준」

이 자료에 대하여 더욱 자세한 내용을 원하시면
과학기술정보통신부 유주연 사무관(044-202-6276)에게 연락주시기 바랍니다.

사람이 중심이 되는
「인공지능(AI) 윤리기준」

2020. 12. 23

관계부처 합동

Ⅰ. 서문

오늘날 인공지능 기술은 컴퓨팅 파워의 성장, 데이터의 축적, 5G 등 네트워크 고도화와 같은 ICT 기술의 발전을 토대로 급 성장하고 있다. 인공지능은 제조, 의료, 교통, 환경, 교육 등 산업 전반에서 본격적으로 활용·확산되고 있으며, 우리 생활에서도 쉽게 인공지능 기술을 접할 수 있게 되었다. 이러한 인공지능 기술의 발전·확산은 생산성·편의성을 높여 국가 경쟁력을 높이고 국민의 삶의 질을 높일 것으로 기대되지만, 한편으로는 기술 오용, 데이터 편향성과 같은 인공지능 윤리 이슈도 제기되고 있다. 본 윤리기준은 이러한 시대적 흐름을 고려하여 '인공지능 개발과 활용 전 단계에서 정부·공공기관, 인공지능 기술 개발자, 인공지능 기술을 활용한 제품·서비스 공급자·활용자 등 모든 사회 구성원이 사람중심의 인공지능' 구현을 위해 고려해야 할 기본적이고 포괄적인 기준을 제시하는 것을 목표로 한다.

본 윤리기준은 '사람중심의 인공지능' 구현을 위해 지향되어야 할 최고 가치로 '인간성(Humanity)'을 설정하고 있다. 이는 아래와 같은 사실을 의미한다. 모든 인공지능은 '인간성을 위한 인공지능(AI for Humanity)'을 지향하고, 인간에게 유용할 뿐만 아니라 나아가 인간 고유의 성품을 훼손하지 않고 보존하고 함양하도록 개발되고 활용되어야 한다. 인공지능은 인간의 정신과 신체에 해롭지 않도록 개발되고 활용되어야하며, 개인의 윤택한 삶과 행복에 이바지하며 사회를 긍정적으로 변화하도록 이끄는 방향으로 발전되어야 한다. 또한 인공지능은 사회적 불평등 해소에 기여하고 주어진 목적에 맞게 활용되어야 하며, 목적의 달성 과정 또한 윤리적이어야 하고, 궁극적으로 인간의 삶의 질 및 사회적 안녕과 공익 증진에 기여 하도록 개발되고 활용되어야 한다.

본 윤리기준은 산업·경제 분야의 자율규제 환경을 조성함으로써 인공지능 연구개발과 산업 성장을 제약하지 않고, 정당한 이윤을 추구하는 기업에 부당한 부담을 지우지 않는 것을 목표로 한다. 또한 본 윤리기준은 범용성이 있는 일반 원칙으로서 사안별 또는 분야별 인공지능 윤리기준 제정의 근거를 제공하여 영역별 세부 규범이 유연하게 발전해 나갈 수 있는 기반을 조성하고, 나아가 사회경제 및 기술 변화와 함께 새롭게 제기되는 인공지능 윤리 쟁점을 반영하여 지속적으로 수정되고 보완되는 일종의 '인공지능 윤리 플랫폼'으로 기능할 수 있다.

본 윤리기준에서 제시하는 원칙과 요건들은 상황에 따라 상충관계가 발생할 수 있으며, 상충하는 문제의 해결 방식은 개별 맥락과 상황에 따라 달라질 수 있다. 따라서 본 윤리기준에서는 각각 원칙들 사이에 고정된 형태의 우선순위를 제시하지는 않으며, 직간접적으로 영향을 받는 이해관계자가 지속적인 토론과 숙의 과정에 참여하여 절충점과 해결 방안을 모색하도록 권유한다.

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 과정**에서 고려되어야할 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 활용되어야 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

- 인공지능 기술은 인류의 삶에 필요한 도구라는 목적과 의도에 부합되게 개발 및 활용되어야 하며 그 과정도 윤리적이어야 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 요건**을 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·지역·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 직간접적인 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 타 원칙과의 상충관계를 고려하여 인공지능 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

o 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.

o 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

o 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

o '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.