OECD 인공지능 권고안

오성탁 한국정보화진흥원(NIA) 지능데이터본부 본부장

1. 머리말

인공지능이 연구수준을 넘어 범용기술(General Purpose Technology)로 발전함에 따라 기존 산업의 차별적 부가가치 창출을 견인하고, 획기적 변화를 촉발하고 있다. Statista에 따르면 인공지능 세계시장 규모는 2018년 73.5억 달러에서 2025년 898억 달러까지 성장할 것으로 전망하고 있다. 인공지능 기술은 한국에서도 이미 자율주행차, 지능형 로봇, 스마트 공장 등의 제조업과 의료, 교육, 금융, 게임, 체육 등 서비스업에 융합되어 상용화가 시작됐다.

〈표 1〉 분야별 주요 인공지능기술 기업

분야	기업
의료	루닛, 뷰노, 딥노이드 등
교육	뤼이드, 노리, 바로풀기 등
법률	로엔비, 로메니저 등
체육	APK어플킹, KT 등
게임	엔씨소프트, 넷마블게임즈 등

AI 스타트업에 대한 투자 증대, M&A 확대 등으로 인공지능 생태계가 구축되기 시작하면서 데이터, 플랫폼 등 AI 자원 확보의 중요성이 확대되고 있으며, 상용화 단계로 빠르게 성장 중인 인공지능 기술 및 산업의 저변 확대에 국가적 대응이 필요하다. 이처럼 인공지능은 기업의 성장, 삶의 질 향상 등 많은 혜택을 주는 반면, 오용과 악용 등의 위험요소 또한 갖고 있다. 마이크로소프트 창업자 빌게이츠는 핵에너지와 핵무기에 빗대어 인공지능을 잘못 사용했을 때의 위험성을 강조[1]한 바 있으며, 테슬라의 일론 머스크 역시도 인공지능의 도입 이전에 위험성을 인지하고 선규제를 만들어야 한다고 언급[2]한 바 있다. 따라서 인공지능의 위험요소를 통제하고 장점을 활용해 혜택을 얻기 위해서는 인공지능으로 인해 발생하는 이슈를 지속적으로 살피며 신뢰가능한 인공지능에 대한 합의를 구축하는 것이 중요하다.

이미 주요 선진국, 글로벌 기업, 학회, 그리고 국제 기구에서는 선제적으로 인공지능이 불러올 위험에 대비하고 인류 삶에 기여할 수 있는 신뢰가능

1) Bill Gates: "AI is like 'nuclear weapons and nuclear energy' in danger and promise", Vox, Mar 20, 2019.

2) "Why Elon Musk fears artificial intelligence", Vox, Nov 2, 2018.

한 인공지능에 대한 논의를 진행하는 중이다. 특히 OECD는 인간 중심의 가치를 추구하는 인공지능에 관해 지속적으로 합의를 이끌어내고 있으며, 2019년 5월 개최된 OECD 각료이사회에서 신뢰가능한 인공지능을 위한 가이드라인인 'OECD AI 권고안'을 회원국의 만장일치로 공식 채택한 바 있다. 이는 국제 수준의 합의가 이루어진 최초의 AI 권고안이라는 데 의의가 있다. OECD AI 권고안은 신뢰가능한 인공지능에 대한 포괄적인 기본원칙과 이러한 원칙을 지키기 위한 구체적인 국가별 인공지능 정책 수립 가이드를 제시한다. 본고에서는 OECD AI 권고안에서 제시하는 신뢰가능한 인공지능을 구현하기 위한 5가지 원칙과 5가지 정책 및 국제협력 방안에 대해 세부적으로 살피고 신뢰가능한 AI를 구현하기 위한 국내의 정책방향을 조망한다.

2. OECD AI 권고안

2.1 추진배경

주요 선진국, 글로벌 기업, 학회 그리고 국제기구 등에서는 인공지능으로 인해 예견될 수 있는 위험에 대응하고 신뢰가능한 인공지능을 구현하기 위한 방안을 지속적으로 논의해왔다. 인공지능 전략 및 윤리 원칙을 마련하고, 인공지능 발전 노력과 함께 윤리 의제를 폭넓게 다루고 있으며, 이와 관련한 인공지능 권고안 및 가이드라인 등을 발표했다.

이에 OECD에서는 인공지능이 불러올 위험에 대비하고 인류 삶에 기여할 수 있는 신뢰가능한 인

〈표 2〉 주요 인공지능 권고안 및 가이드라인

주요 선진국	
미국	A Report on Algorithmic System, Opportunity and Civil Right A Vision for safety 2.0: Automated Driving Systems
일본	Draft AI R&D Guidelines, The Japanese Society for Artificial Intelligence Ethical Guidelines
프랑스	Cedric Villani: For a Meaningful Artificial Intelligence Algorithms and AI: CNIL's report on the ethical issues
영국	Board of Loards AI Select Committee, AI Code
국제기구	
OECD	Artificial Intelligence in Society (2018), Draft AI Council Recommendation(2019. 3)
G7	Charlevoix Common Vision for the Future of Artificial Intelligence(2018. 6, G7 Summit, Canada)
G20	Artificial Intelligence for Europe(2018. 4), Draft Ethics Guidelines for Trustworthy AI(2018. 12)
학회 및 기업	
ACM	Code of Ethics and Professional Conduct
IEEE	Ethically Aligned Design(EAD), Partnership on AI
Google	Ethical AI Principles
Microsoft	AETHER
카카오	Algorithm Ethics

※ 출처: 한-OECD AI 컨퍼런스(2019. 3. 22) 'Toward Trustworthy AI' 발표자료 재구성

공지능을 구현하기 위한 가이드라인인 'OECD AI 권고안'을 2019년 5월 개최된 OECD 각료이사회(MCM)[3]에서 공식 채택했다. 국가 대표 및 국제기구 대표들이 동의하고 채택한 국제적 AI 권고안이라는 점에서 향후 인공지능 분야에 미칠 영향력이 클 것으로 예상된다. OECD AI 권고안의 주안점 중 하나는 인공지능 연구·개발 및 적용 현황을 측정하고 신뢰가능한 인공지능 구현의 진전을 평가하기 위한 지표를 개발하는 것이다. 이와함께 OECD는 AI 정책 관측기구(AI Observatory)를 만들어 신뢰가능한 인공지능을 위한 지속적인 모니터링을 진행할 예정이다.

2.2 신뢰가능한 인공지능

신뢰가능한 인공지능이란 윤리적 원칙을 형성하는 기반으로 인간의 기본권을 존중하며 기본원칙 · 기본가치를 존중하는 인공지능을 의미한다. 여기서 의미하는 기본권 존중에는 인간의 존엄성 존중, 개인의 자유 존중, 민주주의 · 정의 · 법규 존중, 평등 및 비차별 및 소수집단 보호가 포함되어 있다. 기본원칙 · 기본가치 존중에는 인간의 자율성이 인공지능 시스템에 종속되지 않을 자율성의 원칙, 인공지능이 인류에게 해를 끼치지 않게 디자인해야 하는 무해성 원칙이 포함된다. 또한 인공지능은 반드시 공평하게 개발 · 이용 · 통제되어야 하는 공평성의 원칙뿐만 아니라, 인공지능이 의사결정을 할 때 어떠한 근거로 판단을 내리는지에 대한 '설명 가능성'의 원칙도 포함된다.

신뢰가능한 인공지능은 관점에 따라 다양하게 해석할 수 있으나, 〈표 3〉의 구성요소를 만족하는 인공지능을 신뢰가능한 인공지능으로 해석하기도 한다.

〈표 3〉 신뢰 가능한 인공지능의 주요 구성요소

주체	구성요소	세부내용
EU	적법성(Lawful)	모든 관련 법률 및 규정을 준수하고 합법적이어야 함
	윤리성(Ethical)	윤리적 원칙과 가치에 순응해야 함
	견고성(Robust)	좋은 의도로 설계된 AI 시스템도 의도치 않은 부작용이 있을 수 있으므로 기술적·사회적으로 견고해야 함
OECD	포용 및 지속 가능	모든 이해관계자는 인류의 포용 성장, 지속 가능한 발전 및 복지 증진에 힘써야 함
	인간중심	AI 활동주체는 AI 시스템 수명주기 전반에 걸쳐 법률, 인권 및 민주적 가치 등 인간 중심 가치를 존중하고 지키기 위해 힘써야 함
	투명성 및 설명가능성	AI 활동주체는 AI 시스템에 대한 이해를 증진시키고 감춰진 것이 없도록 투명성과 설명 가능성을 확보하여야 함
	견고성 및 안전성	AI 시스템은 전 수명주기에 걸쳐 견고하게 작동되어야 하며 바람직하지 못한 조건을 견딜 수 있거나 극복할 수 있어야 함
	책임 완수	AI 활동주체는 자신들의 역할, 상황의 토대 위에 최신성을 유지하면서 위의 원칙을 존중하며 AI 시스템이 적절히 기능하도록 하는 데 책임을 다해야 함

※ 출처: EU 신뢰할 수 있는 AI 윤리 가이드라인(2019. 4)

3] 각료이사회(Ministerial Council Meeting)는 OECD의 최고 의사결정기구이며, 36개국 회원국 고위급 인사와 유엔, 유네스코, IMF 등 국제기구의 대표가 참석

2.3 OECD AI 권고안 소개

OECD AI 권고안은 혁신적이고 신뢰할 수 있으며 인권과 민주적 가치를 존중하는 인공지능이란 원칙을 담고 있다. 본 권고안에는 신뢰가능한 인공지능에 대한 지표 개발과 그 구현의 진척도를 평가할 수 있는 근거 기반을 만들기 위한 조항도 포함되어 있다. 본 권고안은 법적인 구속력은 없지만 권고안에 동의한 국가 및 국제기관은 해당 권고안을 따르겠다는 정치적 의지를 대변하는 것으로 간주할 수 있으며, OECD 36개 회원국과 비회원국 6개국을 포함한 세계 42개국[4] 이 OECD 각료이사회에서 만장일치로 해당 권고안을 채택한 바 있다. 이에 따라 OECD AI 권고안이 전 세계의 인공지능과 로봇 관련 기술의 핵심 기준으로 활용될 수 있을 것으로 기대된다.

OECD는 AI 기술예측포럼(2016), AI 국제회의(2017) 등을 시작으로 지속해서 AI에 대한 정책적 활동을 수행했으며, AI에 대한 국제 수준의 정책 환경을 형성해야 할 필요성을 입증했다. 이에 2018년 5월 파리에서 열린 OECD 디지털경제정책위원회(CDEP, Committee on Digital Economy Policy) 정례회의에서 인공지능 발전을 위한 권고안을 수립하기로 합의하였고, OECD 이사회는 최종적으로 2019년 5월 장관급 회의에서 해당 권고안을 채택하기로 하였다.

1980년 채택된 OECD 개인정보보호 지침이 미국, 유럽 및 아시아의 많은 개인정보보호법 및 체계의 기초로 활용되었던 바와 같이 OECD AI 권고안 또한 개별 국가의 정부가 입법을 고안하는 것을 지원하여 그 영향력을 미치고 국제 표준의 기초를 형성할 수 있을 것으로 전망된다.

2.4 OECD AI 권고안 구성

OECD AI 권고안에서는 인공지능과 관련된 용어를 우선 정의했으며, 신뢰가능한 인공지능을 구현하기 위한 5가지 원칙과 5가지 국가정책 및 국제협력 방향으로 구성되어 있다.

〈표 4〉 OECD AI 권고안 핵심용어

용어	세부내용
AI 시스템 (AI system)	- 기계 시스템으로 인간이 정의한 목적에 따라 실제 환경 또는 가상 환경에 영향을 미치는 예측, 권장, 의사결정을 할 수 있는 시스템을 의미 - 다양한 수준의 자율성하에서 작동하도록 설계된 시스템
AI 시스템 수명주기 (AI system lifecycle)	- (i) 디자인, 데이터 및 모델, (ii) 타당성 및 검증, iii) 구현, iv) 운영 및 모니터링의 단계로 구성 - 이러한 단계는 종종 반복적인 방식으로 수행되며 반드시 순차적인 것은 아님
AI 지식 (AI knowledge)	- AI 시스템 수명주기를 이해하고 활동에 참여하는 데 필요한 데이터, 코드, 알고리즘, 모델, 연구, 노하우, 교육 프로그램, 관리, 프로세스 및 모범사례 등과 같은 기술 및 리소스를 의미
AI 활동주체 (AI actors)	- AI 시스템 수명주기에서 AI 시스템을 구현하거나 운영하는 조직 및 개인처럼 적극적인 역할을 수행하는 사람
이해관계자 (Stakeholder)	- AI 시스템에 직·간접으로 관련되거나 영향을 받는 모든 조직 및 개인을 포괄하며 AI 활동 주체를 포함

4) • OECD 회원국(36): Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, the United States

• OECD 비회원국(6): Argentina, Brazil, Colombia, Costa Rica, Peru, Romania

2.4.1 OECD AI 권고안 핵심용어

OECD AI 권고안에서는 AI 시스템, AI 시스템 수명주기, AI 지식, AI 활동주체, 이해관계자를 핵심용어로 제시하며 이에 대한 설명을 제시한다.

2.4.2 신뢰가능한 AI 구현을 위한 5가지 원칙

OECD 이사회는 신뢰가능한 AI를 구현하기 위해 AI 활동주체가 각자의 역할에 맞게 다음의 5가지 원칙을 이행할 것을 권고한다.

첫째, 포용성장·지속가능 발전·복지 증진이다. 모든 이해관계자는 인류의 포용성장, 지속가능한 발전 및 복지 증진을 위해 힘써야 한다. 인간의 능력과 창의력을 향상시키고 소수집단의 포용을 진전시키는 방향으로 AI를 구현하려 노력해야 한다. 아울러 성차별과 같은 사회적 불평등을 감소시키고 자연환경을 보호하는 방향으로 AI를 구현하려 힘써야 한다.

둘째, 인간중심 가치 지향과 공정성 지향이다. AI 활동주체는 AI 시스템 수명주기 전반에 걸쳐 자유, 존엄성, 자치, 프라이버시, 평등, 다양성 등의 인간중심 가치를 존중하고 지키기 위해 힘써야 한다. 이를 위해 AI 활동주체는 인간의 결정 능력을 모방한 메커니즘과 안전장치를 최신 기술을 반영하여 구현해야 한다.

셋째, 투명성과 설명가능성 확보이다. AI 활동주체는 AI 시스템에 대한 이해를 증진시키고 감춰진 것이 없도록 투명성과 설명 가능성을 확보해야 한다. AI 시스템의 전체 맥락을 잘 설명하는 최신의 의미가 있는 정보를 제공해야 한다.

넷째, 견고성과 보안 및 안전성 확보이다. AI 시스템은 전 수명주기에 걸쳐 견고하게 작동되어야 한다. AI 시스템은 안전해야 하며, 어떤 상황에서도 시스템의 안전성을 해치는 위험을 노출하지 말아야 한다. 이를 위해 AI 활동주체는 AI 시스템의 수명주기에 따라 만들어진 데이터 세트, 프로세스 및 의사결정 등을 추적 가능토록 하여, AI 시스템이 최신 상태를 일관성 있게 유지하고 있는지 분석 가능하게 해야 한다. AI 활동주체는 자신의 역할, 상황, 역량을 토대로 AI 시스템의 수명주기 전반에 걸쳐 지속적으로 프라이버시, 디지털 보안, 안전성과 편향성(bias) 등과 같은 위험을 체계적으로 관리할 수 있는 방법을 구현하고 강화해야 한다.

마지막으로 책임의 완수이다. AI 활동주체는 자신의 역할 및 상황을 토대로 최신성을 유지하면서 위의 원칙을 존중하며 AI 시스템이 적절히 기능하도록 하는데 책임을 다해야 한다.

2.4.3 신뢰가능한 AI 구현을 위한 국가정책과 국제협력

OECD는 앞 절의 원칙을 준수하며 특히 중소기업(SME)에 주의를 기울여 다음과 같은 국가정책과 국제협력을 이행할 것을 권고한다.

첫째, AI 연구개발(R&D) 투자이다. 정부는 장기적인 공공투자를 고려하고 민간투자, 학제 간 연구를 포함한 R&D를 장려해야 한다. 이를 통해 AI의 기술적 문제뿐만 아니라 사회적, 법적, 윤리적 함의에도 초점을 맞춘 신뢰가능한 AI의 혁신을 촉진해야 한다. 또한, 정부는 공공투자만이 아니라 개인정보를 보호하는 가운데, 민간 분야의 데이터를 공개하도록 민간투자도 장려해야 한다. 부적절한 편향성으로부터 자유로운 AI R&D 환경을 지원하고 상호운용성 및 표준 사용을 개선해야 한다.

둘째, AI 디지털 생태계 육성이다. 이러한 생태계에는 특히 디지털 기술 및 인프라와 AI 지식 공유 메커니즘이 적절하게 포함되어야 한다. 정부는 안

5) 데이터 트러스트(data trust)란 두 개 이상의 조직이 데이터를 안전하고 공정하며 윤리적으로 공유할 수 있는 체계를 의미

전하고 공정하며 법적·윤리적으로 적합한 데이터의 공유를 지원하기 위해 데이터 트러스트[5]와 같은 메커니즘을 구축하는 것을 고려해야 한다.

셋째, 실현 가능한 AI 정책 환경의 조성이다. 정부는 연구개발 단계에서 신뢰가능한 AI 시스템으로의 신속한 전환을 지원하는 정책 환경을 장려해야 한다. 이를 위해 실험을 통해 AI 시스템을 테스트하고 적절하게 확장할 수 있는 제어된 환경을 제공하는 것을 고려해야 한다. 정부는 신뢰가능한 AI 시스템에 대한 혁신과 경쟁을 장려하기 위해 AI 시스템에 적용할 정책 및 규제 체계와 평가 메커니즘을 적절하게 검토하고 적용해야 한다.

넷째, 인적 역량 강화와 노동시장 변화에 대한 대비이다. 정부는 AI를 통한 노동시장과 사회의 변화에 대비하기 위해 이해관계자와 긴밀히 협력해야 한다. AI 역량 강화를 통해 광범위한 응용 프로그램에서 AI 시스템을 효과적으로 사용하고 상호 작용할 수 있도록 지원해야 한다. 정부는 AI로 대체된 근로자의 공정한 전환을 보장하기 위하여 다양한 조치를 취해야 한다. 직무 관련 역량강화 프로그램, AI로 대체된 실직자에 대한 지원, 새로운 기회에 대한 접근 등을 고려해야 한다. 또한 정부는 노동자의 안전과 직업의 질을 높이고, 기업가 정신 및 생산성을 촉진시키며, AI의 혜택을 광범위하고 공정하게 공유하도록 보장해야 한다.

마지막으로 신뢰가능한 AI 구현을 위한 국제협력을 활성화해야 한다. 개발도상국을 비롯한 다양한 이해관계자를 포함하여 정부는 신뢰가능한 AI의 원칙과 윤리적 함의를 발전시키기 위해 국제협력을 추진해야 한다. 이에 정부는 OECD와 세계 포럼 및 지역 포럼과 함께 AI 지식공유를 적절히 육성하기 위해 노력해야 한다. AI에 관한 전문지식을 장기간 축적하기 위해 국제적이고, 지역경계를 넘는, 개방된 다중이해관계자 이니셔티브를 장려해야 한다. 또한 신뢰가능한 AI에 대한 다중이해관계자 합의를 통하여 상호운용성이 뛰어난 글로벌 기술표준 개발을 장려해야 한다. AI 연구개발 및 구현에 관해 국제적으로 비교가능한 측정 지표를 개발하고 자체적으로 사용하도록 장려하며, 이러한 지표를 통해 신뢰가능한 AI를 구현하기 위한 제 원칙의 이행 과정을 평가하기 위한 증거 자료를 수집하도록 장려해야 한다.

2.4.4 신뢰가능한 AI 구현 진척도 지표 및 모니터링 체계

또한, 권고안에는 AI 권고안 발표에 따른 진척사항을 평가하기 위해 인공지능 연구·개발·적용을 측정하기 위한 지표 개발과 함께, 그 구현의 진척도를 평가할 수 있는 내용도 포함하고 있다. OECD AI 권고안은 인공지능 정책 마련을 위한 기준과 기반을 제공하는 최초의 정부 간 기준으로, 향후 관련 분석과 정부의 정책 구현을 위한 개발도구 등을 제공할 예정이다. 이와 관련하여, 위원회(Council)는 권고안의 이행을 지원하기 위해 CDEP에 실질적인 이행지침을 개발하고 AI 정책과 활동에 대한 정보를 교환하기 위한 포럼을 제공하며 다국 간 및 학제 간 교류를 확대할 것을 지시했다. CDEP는 해당 권고안에 기반해 신뢰가능한 인공지능과 관련된 후속작업을 계속적으로 추진하며, 유네스코, 유럽 의회(Council of Europe) 등과 함께 작업할 예정이다.

아울러 권고안 이행 등에 관한 모니터링은 AI에 대한 공공 정책의 포괄적 허브인 OECD AI 정책 관측소(OECD AI Observatory)를 통해 진행할 계획이다. AI 정책 관측소는 관련자들이 신뢰가능한 AI와 관련된 핵심 사항들을 비교하게 할 수 있도록 다양한 정보를 제공할 예정이다. 또한, AI 전략·정책·이니셔티브와 관련된 사항들을 공유하고 AI 지표(metrics) 측정, AI 정책과 모범사례를 지속해서 업데이트하여 신뢰가능한 AI를 위한 실무지침을 제공할 계획이다.

3. 맺음말

OECD AI 권고안은 신뢰가능한 인공지능에 대한 포괄적인 기본원칙을 제시한다. 권고안을 통해 신뢰 기반의 지능정보사회로의 발전에 필요한 핵심 기술 기반과 산업 생태계를 강화하기 위한 원칙, 그 원칙을 지키기 위한 정책 방향을 제시한다. 이러한 권고안의 실효성을 담보하기 위해서는 AI와 관련한 이해관계자들의 합의가 전제된 자율적 규칙의 마련이 병행되어야 할 것이다. 이에 정부는 국가 차원에서 신뢰가능한 AI 전문기술지원기관을 활용하고 각 산업 분야의 기업과 단체가 자율적 규칙을 만드는 것을 지원하고 기반이 되는 요소들에 대해 지원해야 한다. 또한 일자리의 전환을 준비하고 인공지능 도입이 본격화될 시기를 대비하여 사회 안전망을 구축해야 한다.

신뢰가능한 인공지능을 개발하기 위해서는 신뢰할 수 있는 데이터의 확보가 중요하다. 개인정보보호 이슈뿐만 아니라 데이터의 출처, 데이터 구매, 데이터 개방 등에 대한 활발한 논의가 필요한 시점이다. 인공지능 발전 단계는 현재 초기 단계로서 앞으로 가이드라인의 원칙과 위배되는 현안이 지속해서 등장할 것으로 예상하며, 발생하는 문제를 해결하기 위해 유연한 정책 수단이 필요하다.

향후 부작용과 위험을 줄이자는 논의로 탄생한 가이드라인이 암묵적인 규제나 억압이 될지 모른다는 우려의 목소리도 있다. 인공지능 서비스로 발생하는 문제 예방과 기업의 발전 사이의 균형점을 찾기 위해서는 다양한 이해관계자들의 합의점을 찾고 지속적인 보완과 유연성을 지닐 수 있도록 정책이 추진되어야 한다.

참고문헌

[1] OECD AI Policy Observatory, 2019, OECD
[2] OECD draft Recommendation on Artificail Intelligence 발표자료, 2019.3.22., OECD
[3] Recommendation of the Council on Artificial Intelligence, 2019, OECD
[4] 공공부문 AI의 핵심, '신뢰가능 AI', 2019, 한국정보화진흥원
[5] 신뢰가능 AI 구현을 위한 정책 방향, 2019, 한국정보화진흥원
[6] 지속가능한 인공지능(AI) 발전을 위한 OECD 논의 동향, 2019, 한국정보화진흥원